CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 122, No. 5, Pt. 1 of 2 November 2007

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Accurate vocal compensation for sound intensity loss with increasing distance in natural environments

Pavel Zahorik
Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292
pavel.zahorik@louisville.edu

Jonathan W. Kelly
Department of Psychology, Vanderbilt University, Nashville, Tennessee 37212
jonathan.kelly@vanderbilt.edu

Abstract: Human abilities to adjust vocal output to compensate for intensity losses due to sound propagation over distance were investigated. Ten normally hearing adult participants were able to compensate for propagation losses ranging from −1.8 to −6.4 dB/doubling source distance over a range of distances from 1 to 8 m. The compensation was performed to within 1.2 dB of accuracy on average across all participants, distances, and propagation loss conditions with no practice or explicit training. These results suggest that natural vocal communication processes of humans may incorporate tacit knowledge of physical sound propagation properties more sophisticated than previously supposed.

PACS numbers: 43.70.Mn, 43.66.Qp [QJF]
Date Received: July 12, 2007 **Date Accepted:** August 20, 2007

1. Introduction

Under ideal conditions sound intensity obeys an inverse square law with distance: Each doubling of sound source distance decreases sound intensity by 6 dB. Previous research has demonstrated that humans increase their vocal output in order to compensate for these sound propagation losses (Healey, Jones, and Berky, 1997; Johnson *et al.*, 1981; Markel, Prebor, and Brandt, 1972; Michael, Siegel, and Pick, 1995; Warren, 1968). Although the compensation appears to be performed naturally to facilitate effective communication over varying distances between talker and listener and is evident in children as young as 3 years of age (Johnson *et al.*, 1981), there is considerable variability in the amount of compensation reported in the literature. Warren reports increases in vocal output level of 6 dB per doubling distance, suggesting that talkers may perhaps have internalized the ideal inverse-square law relationship for sound propagation loss (Warren, 1968; Warren, 1981). Other studies have reported considerably less level compensation, ranging from 5 to less than 1 dB/doubling (Healey *et al.*, 1997; Johnson *et al.*, 1981; Markel *et al.*, 1972; Michael *et al.*, 1995).

One potential source for this variability is the extent to which the listening environments in which these past experiments were conducted approximated a free-field environment with an ideal 6 dB/doubling propagation loss. Departure from this ideal, such as in rooms with sound reflecting surfaces, in general results in propagation losses less than 6 dB/doubling. Accurate compensation in such environments would, therefore, require less than a 6 dB/doubling increase in vocal output level. This may explain why talkers did not increase their vocal output levels by a full 6 dB/doubling distance in a number of past studies (Michael *et al.*, 1995), although it is important to note that most past studies did not report physical sound propagation losses in the testing environments. For one study that did report propagation losses, there does seem to be a relationship (Michael *et al.*, 1995). A propagation loss of around 2 dB/doubling corresponded to a vocal compensation increase of around 2 dB/doubling for certain condi-

tions. Other results show a much less clear relationship: 4–6 dB/doubling propagation losses in the testing environment of another study corresponded to vocal compensation ranging from less than 2 dB/doubling for adults to more than 35 dB/doubling for children (Johnson *et al.*, 1981). Still unknown, however, is the extent to which talkers may be able to adjust their amount of vocal compensation to different environments with different propagation losses. An additional issue relates to the range of distances over which vocal compensation abilities have been evaluated. Valid tests of a relationship between propagation loss and vocal compensation amounts will require evaluation at multiple distances.

The current study seeks to determine the extent to which vocal compensation depends on the specific propagation losses present in the listening environment over a wide range of distances, and whether talkers can accurately modulate their amounts of compensation to match widely varying amounts of physical propagation loss across different natural listening situations. To the extent that these vocal compensation abilities are used in everyday vocal communication, one might expect normal adult talkers to have developed considerable skill and accuracy in these abilities.

2. Methods

2.1 Propagation loss and source directionality measurements

Estimates of the physical sound level decay with increasing source distance were made for each of four listening conditions used in the experiment: Two acoustic environments and two source orientations. The two acoustic environments, one outdoor and one indoor, had widely different reverberant properties. The outdoor environment was chosen to approximate a free-field listening situation. It was a grassy field approximately 80 m × 40 m, with closest non-ground sound-reflecting surfaces at least 20 m from the measurement locations. The indoor environment was a reverberant hallway, approximately 20 × 3.5 × 3 m (L × W × H) with hard walls, hard floor, and an absorptive ceiling. Source orientation was either directly facing the measurement location (0°), or else rotated 180° in the horizontal plane.

All decay measurements were made using a high-quality omni-directional microphone (Sennheiser KE4-211-2) mounted on a movable tripod. A small (12.7 cm full-range driver, 17.8 × 15.2 × 13.0 cm cabinet) high-output loudspeaker (MicroSpot Monitor, Galaxy Audio, Inc.) with high-quality amplification (D-75, Crown, Inc.) mounted on a tripod with a rotating head at a fixed location served as the sound source. Both microphone and loudspeaker were positioned 1.5 m above the ground surface. In the indoor environment, the source location was approximately 5 m from one end of the hallway, and approximately midway between the side walls. The measurement signal was spectrally-shaped noise (10 s duration), with flat spectrum between 0.1 and 1.5 kHz, decreasing at 60 dB/octave below 0.1 kHz and 20 dB/octave above 1.5 kHz. The spectral shape of this signal was chosen to roughly approximate the spectra of the speech signals used in subsequent portions of the experiment: Male and female talkers producing the vowel /a/. This signal was processed digitally using Matlab software (Mathworks, Inc.) and stored to a standard audio compact-disc for later presentation during measurement conditions. The output level of the measurement signal was fixed for all measurements and corresponded to 90 dBA at 1 m, 0°orientation, in the outdoor environment as measured via a calibrated sound-level meter (Realistic 33-2059, calibrated with a B&K piston-phone, model 4228). Background noise levels were approximately 48 and 37 dBA for the outdoor and indoor environments. Broadband (0.2–4 kHz) reverberation time, T_{60}, for the indoor reverberant environment was approximately 0.7 s, measured with the loudspeaker in the 180° orientation using an energy integration technique (Schroeder, 1965). Decay measurement results were represented in dB relative to the observed microphone output voltage (RMS) at 1 m for each of the two measurement environments and two source orientations.

Occasional nonstationary noise disturbances did occur during both measurement and later experimental sessions. During these occurrences the experimenters suspended the measurement and/or experimental session until the noise disturbance had subsided and discarded any potentially noise-contaminated data.

The extent to which the level decay measurements made using a loudspeaker sound source are valid for making inferences regarding level decay of vocal sound sources depends critically on the two sources having similar directional responses. This is a particularly important issue in acoustically reflective environments where source directivity can strongly affect sound propagation. Measurements were therefore made to estimate the directional responses characteristics of four representative talkers (two male, two female) producing the vowel /a/ and the measurement loudspeaker, using methods fundamentally similar to those described by Studebaker (1985). These measurements were conducted in a second quiet outdoor environment also chosen to approximate free-field conditions: A large grassy field free from sound reflecting surfaces other than the ground. Average noise level in this environment was approximately 40 dBA. All sound level measurements were conducted relative to the measurement location, which was at fixed distance of 1 m. Loudspeaker response measurements were made at 0° and 180° angular orientations in the horizontal plane at a distance of 1 m, using the same source material and microphone used in the level decay measurements. Loudspeaker output level was fixed for all measurements: 90 dBA at 1 m, 0° orientation. Vocal response measurements were also made at 0° and 180° orientations, although two matched measurement microphones (Sennheiser KE4-211-2) were used: One at distance of 1 m, and one fixed to the talker's head at a distance of approximately10 cm from the mouth. Talkers were instructed to use "conversational" vocal output levels and to produce the required vocalization for at least 2 s at each of the measurement orientations. Decibel levels were computed in 1/3-octave bands (0.25–5 kHz) for all measurements. For the voice measurements, the decibel difference in each frequency band between close and far microphone measurements was computed. This representation of relative output level allowed control for differences in absolute source output levels from measurement to measurement, and was used for all subsequent analyses.

2.2 Vocal output level compensation

Participants: Ten adult volunteers (seven females and three males, ages 20–29 years) were paid for their participation. All reported having normal hearing, normal or corrected-to-normal vision, and normal vocal abilities.

Design: In a completely within-participants design, there were two test environments (indoor or outdoor), two participant orientations (facing toward or away from the target location, referred to as 0° and 180° orientations, respectively), and four participant-target distances (1,2,4, and 8 m). Test environment order was counterbalanced. Within each test environment, participant orientation was blocked and order was fixed (0°, then 180°). Within each such block, the order of participant-target distances was also fixed, from nearest to farthest.

Stimuli and procedures: Participants attempted to compensate for the physical sound level loses associated with increasing distance by adjusting the output level of their own voice. The testing procedure was as follows: Participants were lead to the reference location (0 m) in the listening environment (either indoor or outdoor) where they remained for the duration of testing at a given source orientation. The experimenter instructed the participant to produce the vowel /a/ for approximately 3 s and adjust their vocal output level such that the level reaching the experimenter (distal level) remained constant at each of the four target distances. At each target location, starting from the 1 m location and successively increasing in distance for subsequent locations, the experimenter measured and recorded the sound level using a hand-held sound level meter (Realistic 33-2059), which served as the target. Participants were instructed to make initial /a/ productions at conversational levels for the 1 m target distance. For both orientations (0° and 180°), participants were instructed to look at the experimenter prior to vocal production in order to provide the participant with visual distance information. For the 180° orientation, this required participants to turn, look, and then return to the appropriate orientation prior to vocal production. All instructions were provided verbally to each participant at a fixed distance of approximately 1 m (0° orientation) between experimenter and participant prior to vocal compensation testing at different distances. Instructions were provided in detail in the initial environment and then repeated in capsule form in the second environment. No feed-

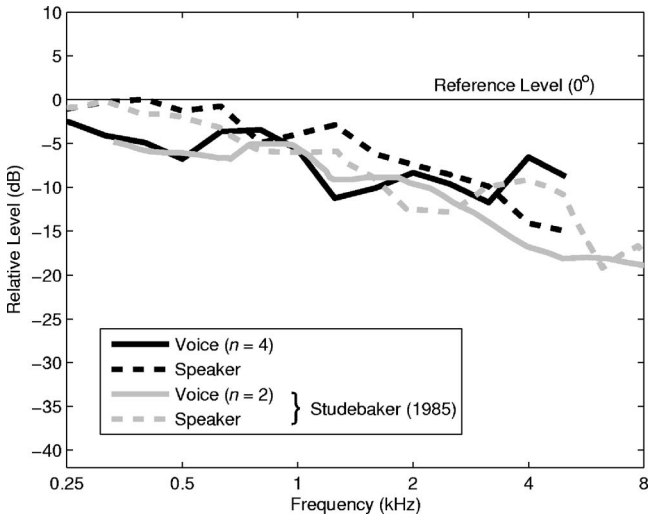


Fig. 1. Signal levels observed in 1/3-octave frequency bands at 180° in the horizontal plane relative to the levels at 0° azimuth for human voices and small loudspeakers measured in free–field environments. Voice levels are mean data, where *n* indicates the number of voices contributing to the mean values (see text for details).

back or explicit training related to sound propagation loss with distance was provided to the participants at any time during the experiment. Participant 103 was not tested at the 180° orientation in either listeningenvironment for unforeseen logistical reasons.

3. Results

3.1 Source directionality

Figure 1 displays the frequency-dependent source directionality results in which levels measured at 180° are compared relative to those measured at 0° (reference level). Small loudspeaker ($17.9 \times 11.1 \times 10.5$ cm cabinet) and vocal (continuous discourse averaged from one male and one female talker) directional response results from a previous study (Studebaker, 1985) are also displayed for comparison purposes. Small loudspeaker directionality is, in general, quite similar to that of the human voice. Both become more directional with increasing frequency at roughly the same rates. Verification of this similarity is particularly important when sound pressure levels averaged across frequency are considered, as in the case of the vocal output measurements made using a sound level meter. The mean level difference between loudspeaker and voice signals (speaker–voice) across 1/3 octave bands from 0.25–5 kHz was 1.5 dB for the current study, which is a slightly better match than the 2.8 dB difference that results from the comparison measurement data (Studebaker, 1985). Overall, this relatively close match in directional responses suggests that valid inferences regarding vocal source propagation may be made from loudspeaker-based sound propagation measurements when averaging across frequency.

3.2 Propagation loss

Sound propagation measurement results for all source orientations and measurement environments are displayed in Figs. 2(a)–2(d). Rates of sound level loss (dB) per doubling of source distance are also shown for each condition, determined via exponential fits to the data using a least-squares criterion. The fitted functions were adequate descriptions of the data in all cases. The RMS error between predicted and measured values was less than 0.3 dB in all conditions except indoors at 0°, where RMS error was still less than 1.5 dB. The outside 0° condition had a measured propagation loss quite similar to that predicted by the inverse-square law for free-field sources. The departure from this ideal 6 dB/doubling loss for the 180° orientation outside

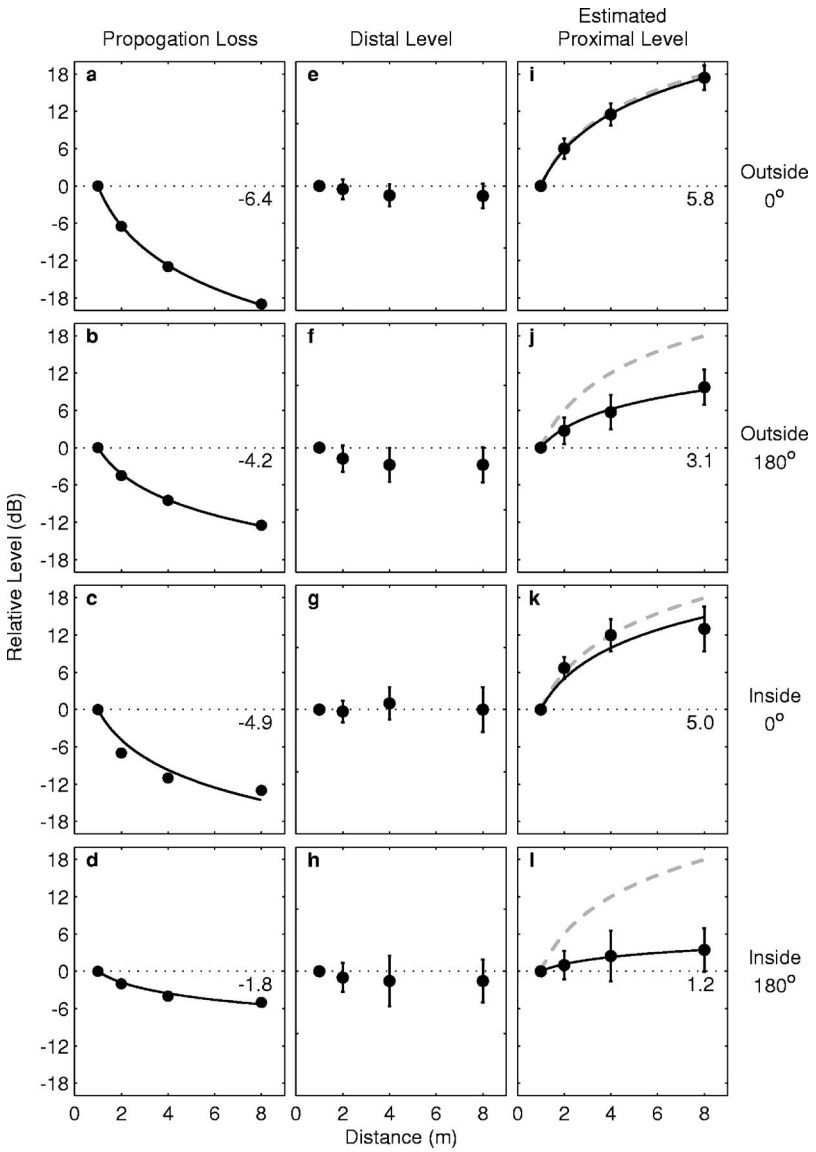


Fig. 2. Propagation loss and vocal compensation results for each of four acoustic conditions (inside and outside environments, 0° and 180° source orientations). All levels are expressed in dB relative to 1 m. Propagation losses are shown in the left column (a–d), and are well described by exponential fits to the data (solid curves). Slopes (dB/distance doubling) are displayed for each fit. Mean levels (dB) at each measurement distance for vocal sources when talkers are instructed to compensate for propagation losses are shown in the center column (e–h). Bars indicated 99% confidence intervals. Mean estimated levels (dB) at the talker's location based on the measured propagation losses are shown in the right column (i–l). Bars again indicate 99% confidence intervals. Solid curves show exponential fits to the data, with slopes indicated (dB/doubling) indicated for each fit. The gray dashed line displays a 6 dB/doubling increase for reference.

is not well understood, but likely resulted from acoustically reflective surfaces that did exist in the outdoor environment in the direction opposite of the measurement microphone, but at a distance of at least 20 m. As expected, sound reflecting surfaces in the indoor environment also resulted in propagation losses less than 6 dB/doubling. The loss for the 0° orientation was similar to that reported in a previous study in which the listening environment had a similar

reverberation time. The indoor 180° orientation had the least propagation loss, given that the energy reaching the measurement microphone in this case was mostly reverberant energy, which is relatively independent of source distance. Propagation loss in this case was also similar to the reverberant-only energy loss observed in a previous study in a room with similar reverberation time. Overall, these four conditions resulted in a considerable range of physical sound propagation losses with distance to the sound source.

3.3 Vocal output compensation

"Conversational" levels at 1 m ranged from 57 to 68 dBA across talkers, with a median level of 62 dBA. Level compensation results are shown in Figs. 2(e) and 2(f), where mean distal levels relative to the levels measured at 1 m for each participant in each condition are displayed. Given that the levels remain within 3 dB of the level at 1 m(0 dB) and lie within the confidence intervals at all distances, it may be concluded that talkers on average accurately increase their vocal output to remain constant at each measurement distance. The mean level relative to 1 m was −1.2 dB across all distances (not including 1 m) and conditions. Because the physical propagation losses are different in the different measurement conditions, the amount of increase in vocal output to compensate for increased distance also differs across condition. Figures 2(i)–2(l) displays estimates of the increases in vocal output level (proximal) based on the measured propagation losses to produce the measured distal levels. All data points are mean levels relative to 1 m. For comparison purposes, a 6 dB/distance doubling is also indicated in these plots (dashed gray line). These estimates further suggest that the amount of level compensation as a function of distance differed dramatically, on average, across the four measurement conditions. It is also clear from these data that talkers are, on average, not simply applying a 6 dB/doubling increase to their vocal outputs in all conditions, although they do apply this rule where it is appropriate to the observed physical propagation loss (outside, 0°).

Although the data in Fig. 2 suggest that talkers can, on average, accurately compensate for propagation losses with distance under different loss conditions, it is important to determine the extent to which individual talkers are also capable of accurate compensation. Individual talker data was, therefore, analyzed via separate exponential function fits, which were found to adequately describe the data in all cases. RMS error between predicted and measured levels ranged from 0.3 to 2.6 dB across all listeners and conditions, with half of all RMS errors below 1.2 dB. Figure 3 displays slope values from the fitted functions for all participants and conditions using vocal signals. Slope values based on functions fit to the mean level data across all participants [e.g., Figs. 2(e)–2(h)] are also displayed. Sound propagations losses (dB/doubling) are also shown for each corresponding measurement condition. Although the slope estimates based on individual data were more variable than those based on mean data, 0 dB/doubling fell within the 99% confidence regions all but two slope values [Participant 102, Fig. 3(b); Participant 103, Fig. 3(c)]. The median slope across all participants and conditions was −0.9 dB/doubling, with half of all slopes falling within +0.8, −1.8 of 0 dB/doubling. Overall, these results indicate that individual participants can compensate for the variable sound propagations losses present in the acoustic conditions with considerable accuracy.

4. Discussion and conculsions

This study has demonstrated that talkers can adjust their vocal output to compensate with considerable accuracy for sound propagation losses ranging from approximately −1.8 to −6.4 dB/doubling distance. This suggests that humans may have tacit knowledge of sound propagation properties more sophisticated than previously thought, and may explain at least some of the previously unexplained variance in past reports of vocal compensation abilities for changing distance (Healey *et al.*, 1997; Johnson *et al.*, 1981; Markel *et al.*, 1972; Michael *et al.*, 1995; Warren, 1968). From the standpoint of vocal communications, the ability to adjust vocal output for sound propagation losses to a listener's position is clearly advantageous, and this advantage is potentially extended with an accurate match between physical propagation loss and vocal output increase. Applying a 6 dB/doubling increase of vocal output for all situations

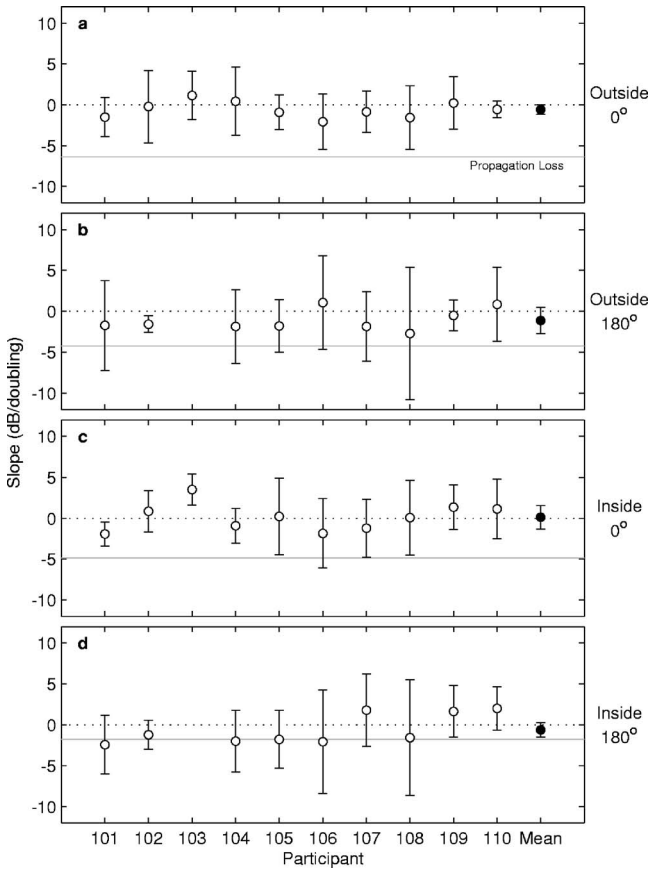


Fig. 3. Slopes of exponential fits to the distal level measurements of each participant and also to the mean levels across participants for four measurement conditions. Bars indicated 99% confidence intervals. Physical propagation losses (dB/doubling) are also shown for each condition.

would unnecessarily limit compensation distances in environments with less than a 6 dB/doubling loss. It is clear, however, that regardless of listening environment, vocal compensation abilities do have practical limits governed by various factors such as the effective dynamic range of the human voice and listening environment noise levels. Although neither of these factors was tested in this study, reasonably accurate vocal compensation was observed in quiet environments over a distance range of 1 to 8 m.

In certain respects, this vocal compensation ability is similar to another, more well-known, form of vocal compensation, known as the Lombard Reflex (Lombard, 1911), in which talkers increase their vocal output level when background noise level is increased. Both forms of compensation facilitate vocal communication by keeping signal-to-noise ratio constant at the listener's location, can be performed with considerable accuracy (i.e., signal-to-noise ratio held constant at the listeners location), and appear to be performed naturally as part of the vocal communication process. The Lombard Reflex has also been documented in other mammals (Scheifele *et al.*, 2005; Sinnott, Stebbins, and Moody, 1975) and songbirds (Cynx *et al.*, 1998), is remarkably robust to volitional control of human talkers (Pick *et al.*, 1989), and appears to depend critically on auditory feedback in its regulation of vocal output level (Siegel and Pick, 1974). Although informal observation suggests both that other species may at least roughly compensate for sound propagation losses, and that it may also be difficult for human talkers to

suppress distance compensation, the extent to which distance compensation depends on auditory feedback is unknown. Clearly additional scientific research is needed in all of these areas.

Results from this study may also have important implications for auditory distance perception, where systematic biases in distance estimates to sound sources have been documented in numerous studies using a wide range of stimulus conditions and psychophysical procedures (see Zahorik, Brungart, and Bronkhorst, 2005, for review). If the vocal compensation accuracy to changing distance observed here does represent tacit knowledge of sound propagation losses in the listening environment, then it is surprising that listeners are not able to use this knowledge to make accurate judgments of sound source distance. This seeming dissociation may perhaps be similar to the well-documented dissociation between accurate visually directed action and inaccurate conscious visual experience (Creem and Proffitt, 1998; Milner and Goodale, 1995), although at least one known deficit in auditorily directed action, dysarthria resulting from Parkinson's Disease, does not appear to affect talkers' compensation for sound propagation losses (Ho, Iansek, and Bradshaw, 1999). This suggests that vocal compensation abilities are not purely action-based. Accurate vocal compensation abilities may instead depend on perceptual processes at least partially distinct from those underlying (inaccurate) conscious experience of sound source distance. Further research will be needed to more fully evaluate these potential relationships.

Acknowledgments

The authors wish to thank Dr. Jack Loomis for his helpful comments and for the use of the facilities in which this study was conducted. Work supported in part by grants from ONR (N00014-01-1-0098) and NIH (F32EY007010, R03DC005709, R01DC008168).

References and links

Creem, S. H., and Proffitt, D. R. (**1998**). "Two memories for geographical slant: separation and interdependence of action and awareness," Psychon. Bull. Rev. **5**, 22–36.
Cynx, J., Lewis, R., Tavel, B., and Tse, H. (**1998**). "Amplitude regulation of vocalizations in noise by a songbird, Taeniopygia guttata," Anim. Behav. **56**, 107–13.
Healey, E. C., Jones, R., and Berky, R. (**1997**). "Effects of perceived listeners on speakers' vocal intensity," J. Voice **11**, 67–73.
Ho, A. K., Iansek, R., and Bradshaw, J. L. (**1999**). "Regulation of Parkinsonian speech volume: the effect of interlocuter distance," J. Neurol., Neurosurg. Psychiatry **67**, 199–202.
Johnson, C. J., Pick, H. L., Jr., Siegel, G. M., Cicciarelli, A. W., and Garber, S. R. (**1981**). "Effects of interpersonal distance on children's vocal intensity," Child Dev. **52**, 721–723.
Lombard, E. (**1911**). "Le Signe de l'Elévation de la Voix," Ann. Maladies Oreille, Larynx, Nez, Pharynx **37**, 101–119.
Markel, N. N., Prebor, L. D., and Brandt, J. F. (**1972**). "Biosocial factors in dyadic communication," J. Pers Soc. Psychol. **23**, 11–13.
Michael, D. D., Siegel, G. M., and Pick, H. L., Jr. (**1995**). "Effects of distance on vocal intensity," J. Speech Hear. Res. **38**, 1176–1183.
Milner, A. D., and Goodale, M. A. (**1995**). *The visual brain in action* (Oxford University Press, New York).
Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R., and Kearney, J. K. (**1989**). "Inhibiting the Lombard effect," J. Acoust. Soc. Am. **85**, 894–900.
Scheifele, P. M., Andrew, S., Cooper, R. A., Darre, M., Musiek, F. E., and Max, L. (**2005**). "Indication of a Lombard vocal response in the St. Lawrence River Beluga," J. Acoust. Soc. Am. **117**, 1486–1492.
Schroeder, M. R. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412.
Siegel, G. M., and Pick, H. L., Jr. (**1974**). "Auditory feedback in the regulation of voice," J. Acoust. Soc. Am. **56**, 1618–1624.
Sinnott, J. M., Stebbins, W. C., and Moody, D. B. (**1975**). "Regulation of voice amplitude by the monkey," J. Acoust. Soc. Am. **58**, 412–414.
Studebaker, G. A. (**1985**). "Directivity of the human vocal source in the horizontal plane," Ear Hear. **6**, 3l5–319.
Warren, R. M. (**1968**). "Vocal compensation for change in distance," *Proceedings of the 6th International Congress on Acoustics (Tokyo)*, pp. 61–64.
Warren, R. M. (**1981**). "Measurement of sensory intensity," Behav. Brain Sci. **4**, 175–223.
Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta Acust. **91**, 409–420.

Experimental study of surface wave propagation in strongly heterogeneous media

Dimitrios G. Aggelis and Tomoki Shiotani[a)]
Research Institute of Technology, Tobishima Corporation, 5472 Kimagase, Noda-shi, Chiba 270-0222, Japan
dimitris-tobishima@t-msweb.net, tomoki_shiotani@tobishima.co.jp

Abstract: In the present paper, the propagation of Rayleigh waves in a strongly heterogeneous medium is discussed. Scattering of stress waves is a difficult scientific problem. Specifically, the interaction of surface waves with distributed inhomogeneity seems highly complicated due to the existence of two displacement components. Rayleigh waves undergo significant attenuation and velocity change depending on the frequency and the inhomogeneity content. The aim of this study is to highlight the dispersive behavior of concrete, especially when damaged, and increase the experimental data in an area where the work is limited.

PACS numbers: 43.35.Cg, 43.20.Gp, 43.20.Jr [JM]
Date Received: May 14, 2007 **Date Accepted:** August 8, 2007

1. Introduction

Stress wave propagation has long been used for characterization of cement-based materials. Although correlations between wave parameters like velocity or attenuation and strength or distributed damage have been observed, the results are always qualitative, based mainly on empirical formulas. One reason is certainly wave scattering due to the extremely inhomogeneous nature of the material that contains pores, air bubbles, sand grains, and aggregates as well as distributed cracks. The length scale of the above phases varies from micrometers to several centimeters, influencing the propagation of practically any frequency and complicating the interpretation of the results. In order to understand the mechanics of propagation in cementitious materials, stress wave dispersion in mortar and concrete has been studied.[1–3]

Focusing on Rayleigh waves, propagation has been studied for multilayered media,[4] or material with surface opening cracks and slots.[5] However, the propagation of Rayleigh waves in a medium with distributed scatterers has not been treated adequately for concrete or other materials, while the body wave scattering has been extensively studied.[6–8]

This paper presents experimental results of surface wave propagation in cement-based materials with randomly distributed and oriented thin inclusions. The aims are: To highlight the dispersive nature of cementitious materials, to make suggestions for more accurate material characterization, and to increase the experimental data of dispersion and attenuation in a specific area where the work is limited so far. The results show that wave propagation velocity depends on inclusion content and frequency while the effects of dispersion are even visible in the raw waveforms. Rayleigh wave parameters are much more influenced by inhomogeneities than longitudinal ones.

2. Experimental details

The experimental setup consists of three wide band piezoelectric transducers with high sensitivity up to 1 MHz. They were placed on the same surface of the specimen (cube of 150 mm side), in a straight line, with a center-to-center distance of 20 mm. The excitation was conducted 20 mm away from the first receiver by pencil lead break, which introduces frequencies up to at least 300 kHz. The small distances were selected due to the severe attenuation, which

[a)]Currently with the Department of Urban Management, Graduate School of Engineering, Kyoto University, Japan.

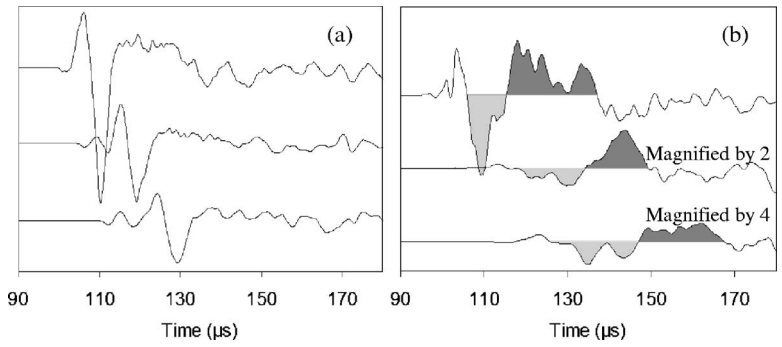


Fig. 1. Average of waveforms collected at 20 mm increments relative to the excitation source on the surface of mortar with inclusion content, (a) 0%, (b) 10%.

made signal acquisition troublesome for longer distances. The material was mortar with fine sand, and small vinyl plates were embedded during mixing at volume contents of 1%, 5%, and 10%. The presented results concern inclusions of dimensions $15 \times 15 \times 0.2$ mm. This shape resembles actual cracks more closely than the spherical inclusions used previously.[2] The acoustic impedance of mortar is approximately eight times higher than the impedance of the inclusions. Additionally, the major wavelength of approximately 20 mm (calculated using the major frequency peak and wave velocity) is comparable to the inclusion dimensions, potentially leading to strong scattering phenomena. The coupling between the sensors and the mortar surface was enhanced by roller bearing grease and no pressure was applied on the sensors in order to minimize any influence on the propagating wave.

3. Inhomogeneity effect on time domain waveforms

Before discussing propagation velocity and attenuation, it is interesting to note the remarkable degree of distortion visible in the time domain signals. Waveforms collected at different locations of a strongly heterogeneous medium are certain to exhibit discrepancies. However, averaging of individual waveforms reveals the common pattern (coherent field). In Fig. 1(a), the averages of waveforms collected by each receiver on a plain mortar specimen are depicted. A strong Rayleigh cycle follows after the initial longitudinal arrivals. Due to the inhomogeneous nature of mortar itself, the Rayleigh cycle becomes more rounded as the wave travels across the surface. Nevertheless, it is clearly visible in all three waveforms.

In the case of mortar with 10% inclusions, which is depicted in Fig. 1(b), the first receiver (20 mm from the source) reveals a highly distorted Rayleigh cycle, while for longer distances multiple peaks are observed. The distortion is likely the result of the distribution of energy in multiple paths. In a homogeneous medium the wave energy travels in a straight line [see Fig. 2(a)] and a reference Rayleigh point [e.g., the strong negative peak in Fig. 1(a)] is visible in all the waveforms. However, as the inhomogeneity increases, the presence of the scatterers distributes the energy to various paths with different length [see Fig. 2(b)]. Therefore, the waveform distorts and a single transit time cannot be easily identified. This can be discussed in terms of the wavelength to inclusion size ratio. Using the major frequency peak and wave velocity, the wavelength, λ, is calculated to be approximately 20 mm, and the sand grains of size less than 3 mm are not expected to strongly influence the wave. On the other hand, the plastic inclusions, which are 15 mm long in two dimensions, seem to cause strong interactions. Note that the central frequency (calculated as the centroid of the fast Fourier transform) is 108 kHz for plain mortar, while for mortar with 10% inclusions it is reduced to 88 kHz for the same propagation distance of 60 mm.

Although the inclusion content of each specimen was strictly controlled in total, it is reasonable that the small volume examined at each individual measurement exhibits variations in the number of inclusions and their orientation. The different arrangement of the scatterers

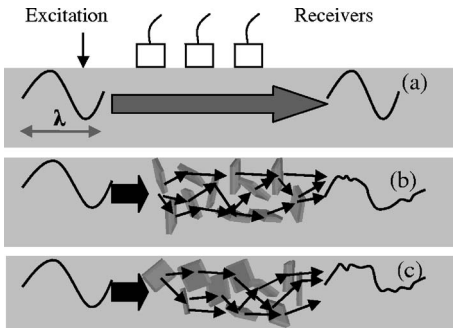


Fig. 2. Propagation in a medium with (a) no scatterers, (b) distributed scatterers, (c) different arrangement of distributed scatterers.

results in different waveforms [see Fig. 2(c)]. To highlight the effect of local variation, a number of individual waveforms for each material are shown in Figs. 3(a) and 3(b). Each waveform was recorded at a different point on the surface, but at the same distance from the excitation. In the case of Fig. 3(a), the waveforms come from plain mortar. It is seen that the change in measurement position causes no serious distortion. The Rayleigh cycle is always identified at the same time. This is a result of the limited local variations in the structure of mortar. In Fig. 3(b), where the case of material with 10% inclusions is concerned, the waveforms exhibit almost no similarities. This is due to strong local variations of the material structure that make any travel path unique. However, averaging the waveforms reveals a weak coherent field, as presented earlier in Fig. 1(b).

The above results are a manifestation of the dispersion induced by the inclusions. They imply that as damage is accumulated in concrete, the material behaves increasingly like an inhomogeneous medium. Therefore, interpretation of ultrasonic data based on homogeneous assumptions should be enhanced or modified in order to provide a wealth of new information in the field of concrete damage characterization.

4. Velocity results

Measuring the transit time of the leading edge of a waveform as it propagates between different receivers leads to the calculation of "pulse velocity". This parameter is widely used in nondestructive testing of concrete, being indicative of the velocity that the energy propagates, although it cannot be defined either as phase or group velocity. However, it is commonly accepted that it corresponds to longitudinal waves. Calculation of this parameter reveals that the increase of plastic inclusions has a decreasing effect on pulse velocity [see Fig. 4(a)]. The longitudinal

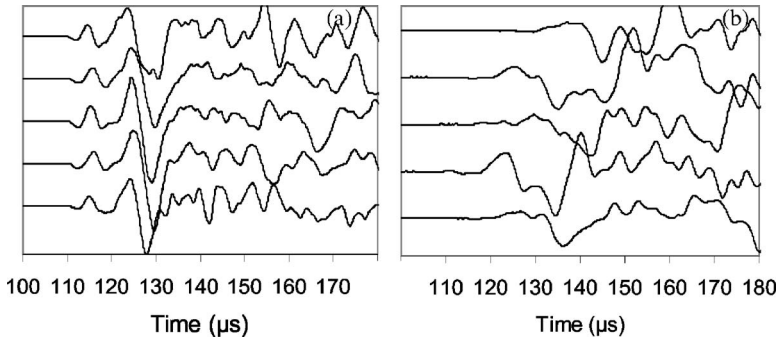


Fig. 3. Waveforms collected at different positions on the surface of mortar with inclusion content, (a) 0%, (b) 10%. The distance from the source is 60 mm.

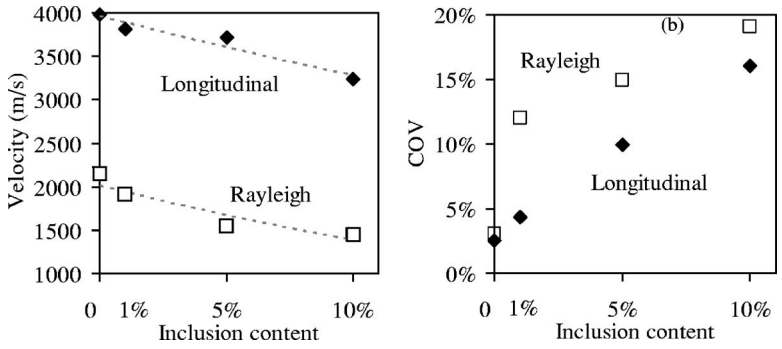


Fig. 4. Influence of vinyl inclusion content on (a) wave velocity and (b) experimental scatter of velocity measurements on mortar.

velocity decreases from almost 4000 m/s, in the case of plain mortar, to 3239 m/s for mortar with 10% inclusions, or suffers a decrease of approximately 18%. This is the average value calculated from 20 individual measurements on the surface of each specimen. It is worth noting that the experimental scatter clearly increases with the inclusion content. In Fig. 4(b) the coefficient of variation, COV (the standard deviation of the twenty measurements divided by the average value) is depicted vs the inclusion content. For the case of plain mortar the variation is quite low (approximately 2.5%), as a result of the uniformity of the material. The addition of inclusions increases COV up to 16%. This trend is another manifestation of the local variations in the composition and arrangement of the material.

Concerning Rayleigh waves, for no or low inclusion content a reference point can be identified in all cases, as mentioned earlier. However, there is no common reference in the signals from the 5% and 10% specimens, complicating velocity measurements. Therefore, cross-correlation between the signals collected from the first and third receiver was used. The resulting time lag can be considered as a measure of the transit time of the major part of the energy between the two transducer positions. This way, although a specific Rayleigh cycle is not always identified, a measure of energy velocity is available. The results again show a decreasing trend as the inclusion content increases. Inhomogeneity clearly diminishes the velocity from 2141 m/s for plain mortar to 1450 m/s for 10% of inclusions, as seen in Fig. 4(a). The decrease of 32% shows that Rayleigh wave propagation is more influenced by inhomogeneity than longitudinal waves.

The difference in the behaviors of longitudinal and Rayleigh wave velocities seems peculiar for an elastic material. According to well-established relations used for concrete,[4] this change of longitudinal to Rayleigh velocity ratio could be explained by a dramatic increase of Poisson's ratio from 0.24 for plain to 0.35 for "damaged" mortar, that is highly unlikely. The greater decrease of Rayleigh velocity is more likely the result of its propagation mechanics, including displacement components in two directions, which make it more sensitive to the existence of scatterers.

The experimental scatter of Rayleigh velocity values increases with inhomogeneity and is certainly higher than that of longitudinal waves. It is worth mentioning that, although for plain mortar the Rayleigh velocity COV is similar to the longitudinal waves' one, for material with just 1% of inclusions it becomes 12.3%, increasing up to almost 20% for high inclusion content.

5. Rayleigh phase velocity

The above-mentioned results concern the propagation velocity of the whole pulse. Since the medium is strongly inhomogeneous, each frequency component could be influenced in a different way, imposed by the content and size of the scatterers. The pulse used contains energy up to 300 kHz and it is interesting to examine how each frequency component behaves. Therefore,

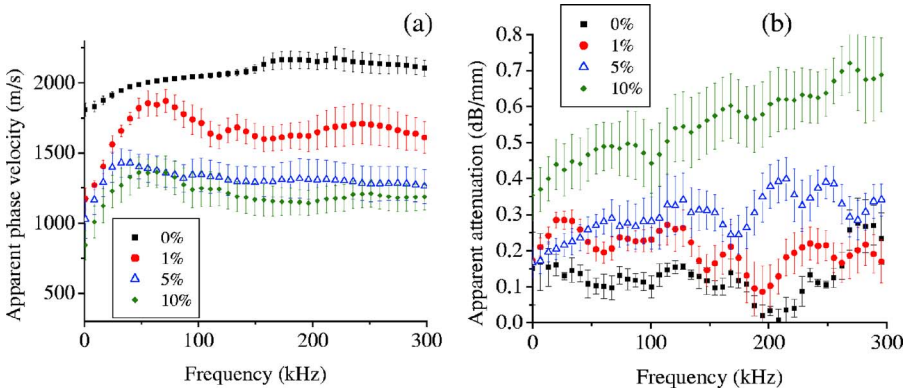


Fig. 5. (Color online) (a) Rayleigh dispersion curves and (b) attenuation curves for different inclusion content of mortar. The error bars stand for the standard deviation.

the frequency dependent phase velocity should be calculated. This is not trivial, since dispersion hinders the identification of a particular portion of the wave, such as the Rayleigh contribution.[9]

However, the energy of the Rayleigh waves is much higher than the other contributions and, therefore, isolating a part of the signal where the Rayleigh is expected, can yield information about this wave type with very little influence from other types. In this case, a window of 30 μs around the major Rayleigh arrivals was isolated and the rest of the waveform was zero-padded, as presented in other cases.[9] Using fast Fourier transform, the phase of the waveform was calculated and unwrapped. Therefore, the phase difference between waveforms collected at different distances from the excitation (i.e., the first and third receiver) leads to the calculation of phase velocity versus frequency curve.[10] The results are depicted in Fig. 5(a) for different materials. Each curve is the average of 20 individual curves, while the error bars stand for the standard deviation and increase with inclusion content, as expected. It is seen that plain mortar exhibits dispersive behavior, with velocity increasing up to about 200 kHz, as has been observed recently for the longitudinal velocity of cementitious materials.[1,2] The increased inhomogeneity induced by the plastic inclusions influences the velocity behavior more strongly. Material with 10% inclusions exhibits the lowest curve, but even 1% of vinyl lowers the curve considerably, compared to plain mortar.

Focusing on the shape of the curves, one detail is worth mentioning. At lower frequencies, the curves exhibit high gradients. However, above 100 kHz the changes are not as abrupt. This is a trend mentioned in many different multi-phase systems. Composites of solids in solids, or suspensions of solids in liquids exhibit similar trends.[7,8,11] At lower frequency bands the velocity exhibits changes, and possibly many resonance peaks. However, at higher frequencies phase velocity seems to converge to a value closer to the velocity of the matrix material. This trend is explained by scattering theory when body waves are concerned. However, in the present case a theoretical treatment is more complicated.

6. Attenuation

In order to obtain information concerning material damage, the attenuation behavior is studied complementary to the velocity. The presence of inhomogeneities increases the attenuation, mainly due to the redistribution of the energy to directions different from the receiver direction. Strongly inhomogeneous media are known to exhibit more dramatic variations in attenuation than in velocity. As seen earlier, inhomogeneity may result in velocity decrease of 30% compared to the velocity of the matrix [see Fig. 4(a)]. However, attenuation may increase even by orders of magnitude, as has been observed for many composite systems[8] and concrete.[12] The material attenuation is the combination of different mechanisms. In the present study the total attenuation $\alpha(f)$ was calculated by

$$\alpha(f) = -\frac{20}{x}\log\left(\frac{A_1(f)}{A_3(f)}\right), \quad (1)$$

where f is the frequency, $A_1(f)$ and $A_3(f)$ are the spectra from the first and third sensor respectively, with a separation distance $x=40$ mm.

As seen in Fig. 5(b), plain mortar exhibits the lowest attenuation while material with inclusion content 10% is more attenuative for the entire band.

It is noted that the attenuation of "damaged" is many times higher than the one of sound material. Even for material with only 1% inclusions the attenuation is certainly increased, making this feature suitable for characterization of slight damage in cases where the velocity exhibits small or no sensitivity.[12] This discrepancy increases with frequency. Therefore, higher frequencies are more powerful for characterization, but only in laboratory conditions so far. In large structures attenuation makes the acquisition and interpretation of stress wave data troublesome.

Another worth mentioning detail concerns the compatibility of dispersion and attenuation curves. Considering the Kramers–Kronig relations, a rapid increase of phase velocity [e.g., as in the lowest 50 kHz of Fig. 5(a)], should be accompanied by a clear attenuation peak within the same frequency bandwidth.[13,14] However, this is not observed for all the attenuation curves of Fig. 5(b). This can be partially attributed to the errors bars, which could mask an attenuation peak, especially for material with 10% inclusions. Another reasonable explanation though, is that the material does not behave linearly (the response is not proportional to its stimulus). Linearity is a condition for the validity of the Kramers–Kronig relations,[13,14] but cementitious materials have been reported to exhibit strongly non-linear behavior.[15] Of course, the Fourier spectroscopy approach assumes linearity, so if nonlinearity is a major issue, the attenuation coefficient and phase velocity determinations may not be perfectly valid. Hence, we have labeled these as the "effective velocity and attenuation spectra".

7. Conclusion

This paper presents experimental results of surface wave propagation through a strongly inhomogeneous medium with randomly distributed and oriented thin inclusions. The main observations concern the strong Rayleigh velocity dependence on frequency and inclusion content, the waveform distortion, and the frequency downshift which results in higher attenuation with increasing frequency. Additionally, the inclusions obstruct Rayleigh waves more strongly than longitudinal. This dispersive behavior is due to contributions from scattering, visco-elastic and microstructure mechanisms, making the theoretical investigation highly complicated. However, the experimental study of scattered waves' characteristics can provide additional correlations with damage, since the crack content leaves its signature on the propagating waveforms, dispersion, and attenuation curves. A next step of particular interest is to compare the influence of different shape and size of inclusions on the Rayleigh parameters.

Acknowledgements

The authors acknowledge Professor Demosthenes Polyzos from the Mechanical Engineering and Aeronautics Dept. of the University of Patras, Greece for useful discussions.

References and links

[1]T. P. Philippidis, and D. G. Aggelis, "Experimental study of wave dispersion and attenuation in concrete," Ultrasonics **43**, 584–595 (2005).
[2]J. F. Chaix, V. Garnier, and G. Corneloup, "Ultrasonic wave propagation in heterogeneous solid media: Theoretical analysis and experimental validation," Ultrasonics **44**, 200–210 (2006).
[3]W. Punurai, J. Jarzynski, J. Qu, K. E. Kurtis, and L. J. Jacobs, "Characterization of entrained air voids in cement paste with scattered ultrasound," NDT Int. **39**, 514–524 (2006).
[4]M. Sansalone and N. J. Carino, "Stress wave propagation methods," in *CRC Handbook on Nondestructive Testing of Concrete*, edited by V. M. Malhotra, and N. J. Carino Florida, CRC Press, Florida, 1991, pp. 275–304.
[5]K. van Wijk, D. Komatitsch, J. A. Scales, and J. Tromp, "Analysis of strong scattering at the micro-scale," J. Acoust. Soc. Am. **115**, 1006–1011 (2004).

[6]V. K. Kinra, and C. Rousseau, "Acoustical and optical branches of wave propagation," J. Wave-Mater. Interact. **2**, 141–152 (1987).
[7]J. Mobley, K. R. Waters, C. H. Hall, J. N. Marsh, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "Measurements and predictions of phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**, 652–659 (1999).
[8]D. G. Aggelis, S. V. Tsinopoulos, and D. Polyzos, "An iterative effective medium approximation (IEMA) for wave dispersion and attenuation predictions in particulate composites, suspensions and emulsions," J. Acoust. Soc. Am. **116**, 3443–3452 (2004).
[9]O. D. Dokun, L. J. Jacobs, and R. M. Haj-Ali, "Ultrasonic monitoring of material degradation in FRP composites," J. Eng. Mech. **126**, 704–710 (2000).
[10]W. Sachse and Y.-H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**, 4320–4327 (1978).
[11]J.-Y. Kim, J.-G. Ih, and B.-H. Lee, "Dispersion of elastic waves in random particulate composites," J. Acoust. Soc. Am. **97**, 1380–1388 (1995).
[12]S. F. Selleck, E. N. Landis, M. L. Peterson, S. P. Shah, and J. D. Achenbach, "Ultrasonic investigation of concrete with distributed damage," ACI Mater. J. **95**, 27–36 (1998).
[13]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "General relationships between ultrasonic attenuation and dispersion," J. Acoust. Soc. Am. **63**, 6, 1935–1937 (1978).
[14]K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "Differential forms of the Kramers-Kronig dispersion relations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 68–76 (2003).
[15]C. Payan, V. Garnier, J. Moysan, and P. A. Johnson, "Applying nonlinear resonant ultrasound spectroscopy to improving thermal damage assessment in concrete," J. Acoust. Soc. Am. **121**, EL 125–130 (2007).

Joint seismic-infrasonic processing of recordings from a repeating source of atmospheric explosions

Steven J. Gibbons,[a] Frode Ringdal, and Tormod Kværna
NORSAR, P.O. Box 53, N-2027 Kjeller, Norway
steven@norsar.no, frode@norsar.no, tormod@norsar.no

Abstract: A database has been established of seismic and infrasonic recordings from more than 100 well-constrained surface explosions, conducted by the Finnish military to destroy old ammunition. The recorded seismic signals are essentially identical and indicate that the variation in source location and magnitude is negligible. In contrast, the infrasonic arrivals on both seismic and infrasound sensors exhibit significant variation both with regard to the number of detected phases, phase travel times, and phase amplitudes, which would be attributable to atmospheric factors. This data set provides an excellent database for studies in sound propagation, infrasound array detection, and direction estimation.

PACS numbers: 43.28.Dm, 43.60.Fg [VO]
Date Received: June 21, 2007 **Date Accepted:** July 26, 2007

1. Introduction

A major component of the International Monitoring System (IMS) for the verification of compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT, www.ctbto.org) is a global network of infrasound sensor arrays deployed to detect atmospheric acoustic signals which could be generated by a nuclear explosion. The processes of detecting and locating events by association of infrasound phases present a very different set of challenges to those involved in the complementary seismic monitoring system. The most significant difference is probably that the propagation path and travel time of a given seismic phase from a given source to a given receiver will remain constant for all subsequent events at that source location over all timescales relevant to current monitoring requirements. In contrast, the travel time and propagation path of an atmospheric sound wave will depend strongly on atmospheric conditions (e.g., Georges and Beasley, 1977; Garcés *et al.*, 1998) which must be accounted for in any conclusions drawn from the detection of an infrasound phase. The modeling and understanding of atmospheric propagation effects need to be guided by well-constrained events with mining explosions typically used for this purpose (e.g., Sorrells *et al.*, 1997; Hagerty *et al.*, 2002; McKenna *et al.*, 2007). In this paper we draw attention to a source of repeating chemical explosions which generate infrasound signals and which are tightly constrained by seismic observations.

2. Seismic and acoustic observations of finnish explosions at regional distances

A series of seismic events detected by the ARCES array were estimated to have taken place at a distance of approximately 175 km in central Lapland (Fig. 1). They were readily identified as explosions since they occurred systematically in sequences and all were conducted at very characteristic times of day (for example, within 1 or 2 s of a full hour or half hour). Colleagues at the Kola Regional Seismological Center (KRSC) in Russia had installed three microbarograph sensors at sites in the Apatity seismic array and observed very coherent, high amplitude signals propagating across the infrasound mini-array with speeds characteristic of sound waves from

[a]URL: http://www.norsar.no/

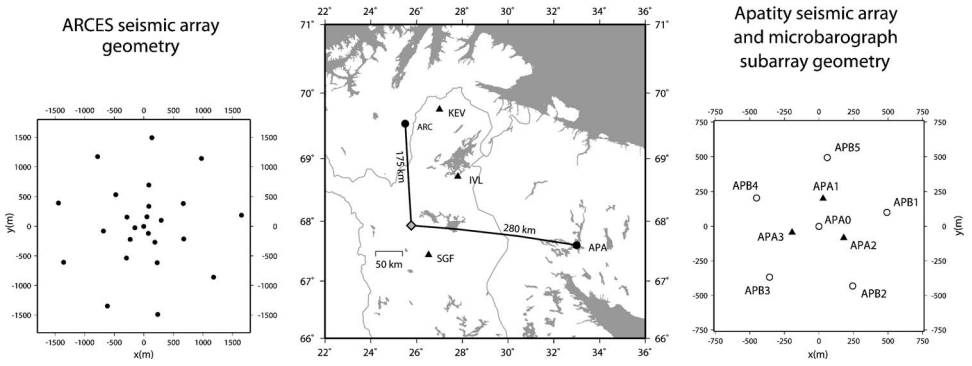


Fig. 1. Location of explosion site in relation to the arrays as indicated. Sites in the ARCES array contain only seismometers as do sites marked with white circles at Apatity. Black triangles at Apatity indicate both a seismometer and a microbarograph. Data from the SGF, KEV, and IVL three-component seismometers helped to constrain the absolute location of the events.

the appropriate direction. In addition, closer examination of the seismic waveforms from ARCES revealed additional high-amplitude signals arriving several minutes after the seismic arrivals which did not correspond to characteristic seismic wave velocities.

An example of such an event is displayed in Fig. 2. The first two pulses on the seismic traces (labelled A and B) correspond to the seismic P-phase (a compressional wave) and more slowly travelling S-phase (a shear wave). Since our station is an array, we can estimate the direction and velocity of these phases using broadband f-k analysis (based upon the method of Capon, 1969) from which, together with arrival times and velocity model, we can estimate the origin time and location. The ground motion some 10 min after the event is dominated by amplitudes comparable to those resulting from the direct seismic arrivals. Performing f-k analysis on a somewhat longer data segment (indicated on the seismogram by the letter C) reveals these waves to be very coherent across the array aperture and to fit very well the hypothesis of a plane wave propagating with air sound speed from a very similar direction. Note the higher resolution provided by the array for the slowly propagating sound wavefront (panel C) than for the seismic wavefronts (panels A and B). The lower-left panel of Fig. 2 displays the beams constructed from all sensors in the array using time-delays corresponding to the calculated slowness vectors. Whilst the observation of acoustic signals on seismogram traces is not uncommon (e.g., Lin and Langston, 2006; Cates and Sturtevant, 2002), it is a useful observation that, in this case, the seismic response to the pressure changes in the incoming infrasound wavefront is so uniform over the array that standard seismic array processing can be applied to infer accurately a direction of arrival of the atmospheric wave.

The repeatability of measurements for azimuth and velocity for the seismic phases together with the S-P travel time difference provided evidence for a similar source location for the different events. However, the similarity of each individual waveform was so great that a full-waveform multi-channel matched filter detector (Gibbons and Ringdal, 2006) could be applied, taking a single specimen waveform as a template and picking out subsequent events simply from the maxima of the correlation coefficient traces. This procedure fulfilled the multiple aims of (a) identifying automatically a large number of events, (b) calculating to subsample precision the relative origin times of events, and (c) confirming that events cannot be separated by more than a few hundred meters (Geller and Mueller, 1980).

The relative timing of events allows the waveforms from multiple events to be aligned and compared (Fig. 3). A large amplitude acoustic signal approximately 600 s following the origin time is observed for almost all of these events from 2002 but, unlike the seismic signals which are almost identical for each explosion, the temporal nature and amplitudes differ greatly from event to event. On some days no signal is visible and we sought to find all evidence present of atmospheric sound waves following these events. Estimation of marginal coherent signals

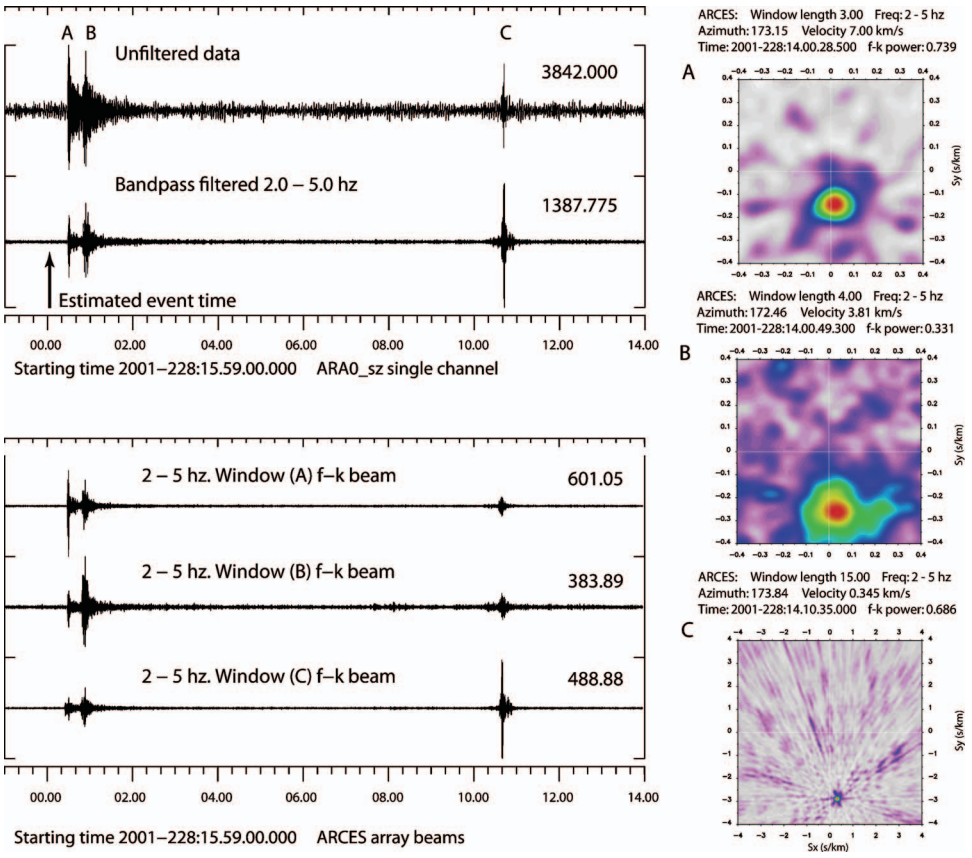


Fig. 2. Waveform data from the ARCES seismic array one minute prior to and 14 min after a surface level explosion in Northern Finland at a distance of approximately 175 km. The top panel shows waveforms on the central seismometer: The seismic P-phase (A), seismic S-phase (B), and an unidentified arrival approximately 640 s after the estimated origin time (C). The broadband f-k analysis plots indicate that all of these signals come from a backazimuth of approximately 173°. (Note the different scale for the slowness grid C.) The lower panel shows the beams from the full ARCES array for the slowness vectors indicated. All seismograms show velocity and the numbers above the traces indicate maximum amplitude in counts.

within a noise field is traditionally achieved using cross-correlation techniques (e.g. Jacobson, 1957) and the main technique used for infrasound processing on the IMS arrays is the progressive multichannel cross-correlation (P.M.C.C.) method (Cansi, 1995). A comprehensive summary of operational processing of infrasound data in the nuclear explosion monitoring context is provided by Brown *et al.* (2002) who define a detection statistic based upon the mean of all pair-wise channel correlations with time-delays corresponding to the theoretical plane wavefront models. The statistic Γ defined in Eq. (15) of Brown *et al.* (2002) was evaluated over successive 10 second time-windows of ARCES data following each of 141 events and Fig. 4 (left panel) displays the Γ value obtained whenever the maximum-gain slowness vector falls close to the expected value (see the slowness grid labelled with the letter C in Fig. 2). In practice, we required that Γ exceeded 0.01 with velocity in the interval [0.3 kms^{-1}, 0.4 kms^{-1}] and azimuth between 170° and 180°. A high value of Γ indicates a high correlation between the appropriately delayed channels and, whilst this value is highest for high signal-to-noise ratio (SNR) acoustic signals, significant values can be obtained even when the signal amplitude is smaller than the ambient noise level. Figure 4 confirms that evidence of a corresponding infrasound arrival was observed for almost every explosion, with only 5 out of 141 events showing no evidence of sound waves. The most common arrivals occur between 600 and 680 s after

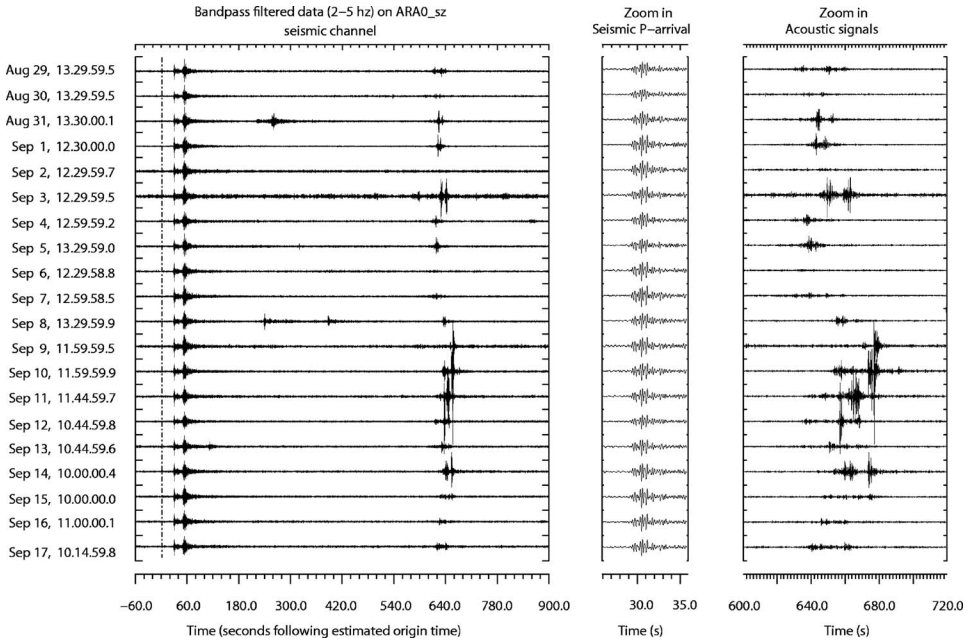


Fig. 3. Recordings on the ARCES seismic array (channel ARA0_sz) of 20 events at the Finnish explosion site in August and September 2002. The time provided to the left is the estimated event UTC origin time. All seismograms are aligned according to the maximum correlation coefficient and have identical vertical scaling such that each division represents ±1000 counts. Signal arrivals between 450 and 700 s after origin time are demonstrated by array analysis to propagate with sound velocity from an approximate 173° back azimuth. All arrivals between 200 and 450 s correspond to unrelated seismic events.

origin time with an apparently smooth variation over a several day time-scale. For a small number of events, an additional infrasonic phase (often with large amplitude) arrives between 500 and 570 s after the event origin time.

The same procedure was applied to the microbarograph sub-array at Apatity on data segments provided by colleagues at KRSC and, of the 91 events with available microbarograph data, only three provided no indication of signals from the anticipated direction. Many of these recordings are of very high quality and a high SNR frequently results in high correlation coefficients (as indicated by the colors in the right panel of Fig. 4). With the minimal configuration of three sites, the array gain is far poorer and there is no redundancy. If a single sensor is subjected to an outage or excessive noise, no direction estimate can be made regardless of how well the other sensors perform. This may partly explain the more speckled appearance of the APA panel. The noise levels at Apatity are high due to heavy industry and other local human activity which may hinder the observation of tele-infrasonic signals. Three distinct phases are frequently observed at Apatity approximately 900, 1000, and 1200 s following each event. Vinogradov and Ringdal (2003) analyzed waveforms from five such explosions in considerable detail, concluding that these well-observed phases had travel times consistent with the Iw, Is, and It phases as described by Brown *et al.* (2002).

3. Summary

We have identified a source of explosions which, in addition to generating seismic signals detected out to distances of several hundred kilometers, result in infrasound signals detected at the microbarograph array in Apatity at a distance of 280 km and on the ARCES seismic array at a distance of 175 km. The seismic signals provide excellent constraints on the source. Waveform similarity from event to event not only constrain the events to be almost co-located, but rule out the possibility of multiple events as is common for ripple-fired mining blasts (e.g., Gibbons *et*

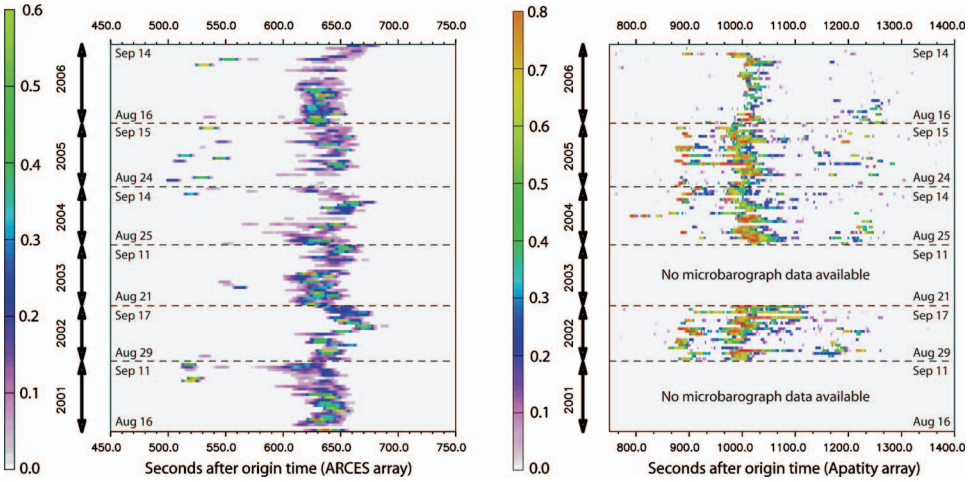


Fig. 4. Detection statistic over the full ARCES seismic array and the three-element microbarograph sub-array at Apatity within the time-windows as indicated following each of 141 identified explosions in northern Finland between 2001 and 2006. Events occur one per day between the dates indicated. A pixel is drawn every second, at time t, for each event provided that the preferred slowness and back azimuth evaluated over the 10.0 s long window beginning at time t fall within an acceptable range for acoustic waves from the given source. The color indicates the value of the detection statistic defined in Eq. (15) of Brown *et al.* (2002).

al., 2005). We conclude that differences in the occurrence and appearance of infrasonic arrivals from event to event are due to atmospheric conditions alone. The similar amplitude of the seismic signals from event to event imply similar explosion yields and, as observed by McKenna *et al.* (2007), this does not appear to influence the amplitude of the infrasound signals greatly. For many events, the sound waves observed at the ARCES seismic array did not exceed the ambient noise level. The presence of infrasound arrivals was, however, confirmed for almost all events by significant values of the Γ statistic defined by Brown *et al.* (2002).

It would clearly be of considerable interest to apply infrasonic propagation models to attempt to explain the variations in travel times and phase amplitudes documented here. However, such a study is well beyond the scope of the current paper. We should only like to note that the horizontal velocities or propagation times versus distance of the infrasonic waves could give an indication of the turning points associated with the various detected phases. For example, Brown *et al.* (2002) discuss generic travel time information for three main infrasonic phases (Iw, Is, and It, with turning points in the troposphere, stratosphere, and thermosphere, respectively) that might be detected at distances similar to those considered in this paper. The observation of infrasound signals at ARCES may provide useful data for subsequent studies of sound propagation at short distances (cf., Il-Young *et al.*, 2002).

Whilst the sensitivity to acoustic signals varies from sensor to sensor of the ARCES array, direction estimates for the sound waves are remarkably similar over many different subsets of sensors (e.g., Ringdal *et al.*, 2006). The uncertainty associated with direction measurements is of great importance for IMS arrays (Szuberla and Olson, 2004) with signal incoherence (Christie *et al.*, 2005) and strong sidelobes (Kennett *et al.*, 2003) presenting significant challenges for processing over large aperture arrays. Whilst the ARCES seismic array is still only a surrogate for the infrasound array IS37 to be built near Karasjok, the large number of sensors and corresponding wide range of sensor separations make this an ideal laboratory for coherence studies.

The recording of coherent infrasound wavefronts on seismic arrays may be more widespread than is presently assumed and an effort ought to be made to classify their occurrences on, for example, the IMS seismic arrays. The large amplitudes which can be generated (see Fig. 3) can be problematic in that they can potentially mask out important seismic arrivals. Indeed, one

of the few documented descriptions of infrasound on IMS seismic arrays is a description of beams deployed on the GERES array in southern Germany to identify and screen out sound waves generated by nearby military activity (Harjes *et al.*, 1993). However, rather than simply discarding such signals, these waveforms could be analyzed to address topical issues in infrasound array processing such as the discrimination of near- and far-field sound sources (Szuberla *et al.*, 2006).

The current status on the database described can be obtained by contacting the authors and updates are likely to be reported on in future NORSAR technical reports and elsewhere. Use of the waveform correlation detector on ARCES seismic data is being extended back in time and, so far, positive identifications have been made as far back as August 24, 1988. The picture is not yet complete as many years of data are still in magnetic tape archives and the conversion process is ongoing. We have no reason to believe that the events will not continue into the future and we would advocate passive field experiments to record and interpret both seismic and atmospheric signals from subsequent events.

Acknowledgments

We are grateful to colleagues at the Kola Regional Seismological Center in Apatity, Russia, for providing segments of seismic and infrasound data from the APA array for the requested time-windows. We acknowledge colleagues at the Institute of Seismology at the University of Helsinki who provided us with seismic event bulletins.

This work was sponsored by Army Space and Missile Defense Command (SMDC) under Contract No. W9113M-05-C-0224. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Space and Missile Defense Command or the U.S. Government.

References and links

Brown, D. J., Katz, C. N., Le Bras, R., Flanagan, M. P., Wang, J., and Gault, A. K. (**2002**). "Infrasonic signal detection and source location at the prototype international data centre," Pure Appl. Geophys. **159**, 1081–1125.

Cansi, Y. (**1995**). "An automatic seismic event processing for detection and location: The P.M.C.C. method," Geophys. Res. Lett. **22**, 1021–1024.

Capon, J. (**1969**). "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1418.

Cates, J. E., and Sturtevant, B. (**2002**). "Sesimic detection of sonic booms," J. Acoust. Soc. Am. **111**, 614–628.

Christie, D. R., Kennett, B. L. N., and Tarlowski, C. (**2005**). "Detection of regional and distant atmospheric explosions at IMS infrasound stations," in *Proceedings of the 27th Sesimic Research Review, Rancho Mirage, California, September 20–22, 2005. Ground-based Nuclear Explosion Monitoring Technologies. LA-UR-05-6407*, pp. 817–828.

Garcés, M. A., Hansen, R. A., and Lindquist, K. G. (**1998**). "Traveltimes for infrasonic waves propagating in a stratified atmosphere," Geophys. J. Int. **135**, 255–263.

Geller, R. J., and Mueller, C. S. (**1980**). "Four similar earthquakes in Central California," Geophys. Res. Lett. **7**, 821–824.

Georges, T. M., and Beasley, W. H. (**1977**). "Refraction of infrasound by upper-atmosphere winds," J. Acoust. Soc. Am. **61**, 28–34.

Gibbons, S. J., Kværna, T., and Ringdal, F. (**2005**). "Monitoring of seismic events from a specific source region using a single regional array: A case study," J. Seismol. **9**, 277–294.

Gibbons, S. J., and Ringdal, F. (**2006**). "The detection of low magnitude seismic events using array-based waveform correlation," Geophys. J. Int. **165**, 149–166.

Hagerty, M. T., Kim, W.-Y., and Martysevich, P. (**2002**). "Infrasound detection of large mining balsts in Kazahkstan," Pure Appl. Geophys. **159**, 1063–1079.

Harjes, H.-P., Jost, M. L., Schweitzer, J., and Gestermann, N. (**1993**). "Automatic seismogram analysis at GERESS," Comput. Geosci. **19**, 157–166.

Il-Young, C., Jun, M.-S., Joen, J.-S., and Min, K. D. (**2002**). "Analysis of local seismo-acoustic events in the Korean Peninsula," Geophys. Res. Lett. **29**, 29-1, 10.1029/2001GL014060.

Jacobson, M. J. (**1957**). "Analysis of a multiple receiver correlation system," J. Acoust. Soc. Am. **29**, 1342–1347.

Kennett, B. L. N., Brown, D. J., Sambridge, M., and Tarlowski, C. (**2003**). "Signal parameter estimation for space arrays," Bull. Seismol. Soc. Am. **93**, 1765–1772.

Lin, T.-L. and Langston, C. A. (**2006**). "Anomalous acoustic signals recorded by the CERI seismic network," Seismol. Res. Lett. **77**, 572–581.

McKenna, M. H., Stump, B. W., Hayek, S., McKenna, J. R., and Stanton, T. R. (**2007**). "Tele-infrasonic studies of hard-rock mining explosions," J. Acoust. Soc. Am. **122**, 97–106.

Ringdal, F., Kværna, T., Gibbons, S. J., Mykkeltveit, S., and Schweitzer, J. (**2006**). "Research On Seismic and Infrasonic Monitoring of the European Arctic," in *Proceedings of the 28th Seismic Research Review*, Orlando, Florida, September 19–21, 2006, Ground-based Nuclear Explosion Monitoring Technologies. LA-UR-06-5471, pp. 1017–1027.
Sorrells, G. G., Herrin, E. T., and Bonner, J. L. (**1997**). "Construction of regional ground truth databases using seismic and infrasound data," Seismol. Res. Lett. **68**, 743–752.
Szuberla, C. A. L., Arnoult, K. M., and Olson, J. V. (**2006**). "Discrimination of near-field infrasound sources based on time-difference of arrival information," J. Acoust. Soc. Am. **120**, EL23–EL28.
Szuberla, C. A. L., and Olson, J. V. (**2004**). "Uncertainties associated with parameter estimation in atmospheric infrasound arrays," J. Acoust. Soc. Am. **115**, 253–258.
Vinogradov, Y., and Ringdal, F. (**2003**). "Analysis of infrasound data recorded at the Apatity array," NORSAR Scientific Report: Semiannual Technical Summary No. 1-2003, NORSAR, Kjeller, Norway, pp. 68–77.

High frequency broad band scattering from water-saturated granular sediments: Scaling effects

Anatoliy N. Ivakin
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105
ivakin@apl.washington.edu

Jean-Pierre Sessarego
CNRS/LMA, 31 Chemin Joseph Aiguier, 13402, Marseille, Cedex 20, France
sessarego@lma.cnrs-mrs.fr

Abstract: Sound backscattering from water-saturated granular sediments at frequencies from 150 kHz to 8 MHz at oblique incidence was studied in controlled laboratory conditions. Two kinds of sediments, medium and coarse sands, were degassed, and their surface was flattened. In these conditions, the sediment granular structure can be considered as a controlling mechanism of backscattering. Comparison of frequency dependencies of backscatter for the two sediments with different mean grain size shows the existence of a persistent scaling effect that allows description of the backscattering strength as a function of one parameter, the mean grain size/wavelength ratio.

PACS numbers: 43.30.Ma [GD]
Date Received: June 3, 2007 **Date Accepted:** August 23, 2007

1. Introduction

Reported measurements of seabed scattering are largely confined to relatively low frequencies, up to 300 kHz (see, e.g., Jackson and Richardson[1] and references therein). Still higher frequencies, up to a few megahertz, are used in seafloor imagery (e.g., in pencil-beam sonars), particularly to observe dynamic processes at the seafloor. From the physics of sound-sediment interaction standpoint, however, interesting effects can be anticipated at these high frequencies as the wavelength can become comparable with the sediment typical grain sizes. Unfortunately, existing observations in this case are usually noncalibrated and, therefore, do not allow measurements of system independent scattering characteristics of the seabed, such as the bottom scattering strength. A recent exception is presented in Greenlaw *et al.*,[2] where the backscattering strength for sand sediments in shallow water was measured at frequencies from 265 kHz to 1.85 MHz. However, physical interpretation of results obtained in this work is quite ambiguous due to a significant uncertainty of the sediment parameters. As the authors noted,[2] the sediment surface was only relatively smooth and had, e.g., pockmarks created by fish foraging on the bottom fauna. A small but detectable sediment gas content was noticed as well. Therefore, acoustic observations could be affected by the small-scale roughness, the presence of gas bubbles in the sediment, and other mechanisms, e.g., those related to possible near-surface stratification of the sediment.[3,4]

Unfortunately such ambiguities are quite usual for experimental studies of sediment acoustics in shallow water where the dynamical complexity and unpredictability of environmental conditions can be so great that even very extensive time- and labor-consuming environmental measurements (such as those at the recent major sediment acoustics experiment, SAX04[5,6]) may not be enough to sufficiently reduce the uncertainty in interpretation of acoustic data. In this connection, conducting experiments in well-controlled laboratory conditions can become a valuable (but much less expensive) supplement to the experiments at sea. An impor-

tant advantage of laboratory studies of sediment scattering is that they make it possible to simplify the problem by reduction of the number of controlling parameters and allow observation of the effects of different scattering mechanisms separately. For example, the sediment can be degassed to eliminate effects of micro-bubbles, and its surface can be flattened to exclude contributions of the sediment roughness at scales greater than the grain size. This approach was used by Nolle *et al.*[7] in their study of acoustical properties of sand sediments at two frequencies, 500 kHz and 1050 kHz. Williams *et al.*[8] studied broadband backscattering from several types of the sediments in a wider frequency range, about 200 kHz to 2.5 MHz. Unfortunately, the bottom scattering strength was not obtained, as the measurements were still not calibrated.

In this paper (Sec. 2), we describe experiments conducted at the CNRS/LMA water tank facility on backscatter from sands over a wide frequency range, 150 kHz to 8 MHz. Note that existing methods for measuring the bottom scattering strength normally assume using narrow band signals (see, e.g., Jackson and Richardson,[1] Ch. G.2, pp. 497–502). To be applied to broadband transducers, significant modifications are required. One such modification is considered in Sec. 3 of this paper. It made possible, to our knowledge for the first time, measurement and analysis of the frequency dependence of the sediment backscattering strength over such a wide frequency range. The results are given in Sec. 4. The analysis assumes that, for the given conditions, only one mechanism of scattering can be dominating, which is due to the sediment granular structure, and that the controlling sediment parameter is the mean grain size. In this case a scaling effect is possible: Given grazing angle, the backscattering strength, which is system independent and depends only on frequency and sediment properties, can be presented as a unique function of only one parameter, the mean grain size/wavelength ratio. This effect is demonstrated and the scaling function is presented in Sec. 4 as well.

2. Experiments

Two kinds of water-saturated sediments with different grain sizes, moderately well sorted medium and coarse sands, were chosen for the study. Sediment properties such as the density, porosity, and grain size, were measured by nonacoustical methods. The medium sand had the mean grain diameter 0.245 mm, sediment/water density ratio 1.98, and porosity 36.5%. The coarse sand had the mean grain diameter 1.55 mm, sediment/water density ratio 2.02, and porosity 33%. Prior to the experiments, the two sediments were stored in containers filled with water treated with chlorine to exclude the presence of living organisms which might generate bubbles. The sediments were transferred in smaller containers (without exposure to air) to a large water tank (chlorine treated as well) where, in addition, they were sieved and agitated to eliminate remaining bubbles. After such preparations, the two sediments were placed in different plastic rectangular boxes of the same size with horizontal dimensions 17×23 cm and 9 cm in the vertical. The sediment surface was carefully flattened by scraping even with the box edges to eliminate roughness at scales larger than the sediment grains. Therefore, the necessary measures were taken to ensure that only the sediment granular structure be considered as a dominating mechanism (rather than large scale roughness and/or gas bubbles) controlling total scattering in and from the sediment.

Experiments on backscattering from the sediments were conducted using a typical monostatic geometry with the transducers acting as both source and receiver (see, e.g., Jackson and Richardson,[1] Fig. G3). Two broadband Parametrics transducers with nominal center frequencies 500 kHz and 5 MHz were used which, altogether, allowed covering continuously the wide frequency range of 150 kHz to 8 MHz. The transducers are circular pistons with radii 12.5 mm (500 kHz) and 6.5 mm (5 MHz). Their directivities are documented over all used frequencies and well described by the corresponding theoretical function $D=2J_1(ka \sin \vartheta)/(ka \sin \vartheta)$, where a is the piston radius, $k=2\pi/\lambda$ is the wave number in water, $J_1(\ldots)$ is the Bessel function of the first order, and ϑ is the angle between the direction of observation and the maximum response axis (MRA) of the transducer.

The position of the transducers was controlled by a system allowing their automatic vertical and horizontal translation with 0.1 mm accuracy and their rotation with 0.1° accuracy.

For calibration of each transducer, first the time series of the echo signal reflected from the water-air interface at normal incidence were measured and Fourier analyzed. Then, the transducer was rotated toward the sediment surface and set in a position with a fixed direction. Horizontal translations of the transducer (parallel to the sediment surface) were used to develop a statistically uniform ensemble of realizations for the echo signals. To provide the necessary statistics, a number of horizontal positions of the transducer was set automatically with a consecutive horizontal shift of 0.5 cm. At each position, time series of the echo signal scattered from the sediment (at a fixed incidence angle) were measured for a number of pings (up to 64) and coherently averaged to reduce possible effects of noise.

Fourier spectral analysis was performed on the echo signals with a flexible time-windowing. The windowing was used, first, to reduce the effects of reverberation from the sediment box walls and, second, to accommodate corresponding distances and the scattering areas of the sediment surface with the frequency dependent beam pattern of the transducer (as discussed below). The windowed echo frequency spectra, $F(f)$, were normalized by the spectra of the calibration signal, $F_0(f)$. These normalized spectra provided a data set for the statistical processing. Particularly, their second moments (the squared magnitudes averaged over realizations), were used to estimate the normalized scattered intensity

$$I_s(f)/I_0(f) = \langle |F(f)|^2 \rangle / |F_0(f)|^2, \tag{1}$$

and to obtain the frequency dependencies of the sediment backscattering coefficient, m_s, and its decibel equivalent, the scattering strength, $10 \log m_s$, as given in the next section.

3. Extracting scattering strength

The intensity of calibration and scattered signals in the frequency domain, I_o and I_s [see Eq. (1)], can be calculated from the corresponding signal time series taken in the observation time window. Specifically, the scattering intensity is defined as the squared magnitude of the Fourier transform of the scattered signal time series taken in the time window, $t \in [t_1, t_2]$, related to time-delays occurring in the process of scattering and defined by corresponding range-time inequality

$$ct_1 < r_i + r_s < ct_2, \tag{2}$$

with r_i and r_s being the distances from the source and receiver to the ensonified area (or, more accurately, the observation area) and c being the sound speed in water.

Consider the case of a monostatic geometry, where $r_i = r_s = r = ct/2$. The incidence angle (measured from the vertical), θ, is defined as follows: $\cos\theta = t_f/t$, $t \geqslant t_f = 2H/c$, where t_f is the arrival time in the vertical direction (fathometer return) and H is the vertical distance from the source to the sediment surface. For the angle window $\theta \in [\theta_1, \theta_2]$, taking into account that $\tan\theta d\theta = d \ln t$, we have $\theta_2 - \theta_1 \approx (t_2 - t_1)/t_w \tan\theta_w$ with $t_w = (t_1 + t_2)/2$ and $\theta_w = (\theta_2 + \theta_1)/2$ being centers of the corresponding time and angle windows.

The scattering cross section per unit area per unit solid angle (see, e.g., Ref. 1, p. 23) or, for brevity, the scattering coefficient, m_s, can be defined from the equation

$$I_s = I_o \frac{r_o^2}{r^4} A_{ef} m_s, \tag{3}$$

where I_o is the intensity at the distance $r_0 = ct_0$ in the direction of maximum directivity (on the transducer MRA) with t_0 being the arrival time of the calibration signal, and A_{ef} is an effective observed scattering surface obtained from the directivity D, defined with respect to the field amplitude, by integration over the observation area, defined by Eq. (2), as follows:

$$A_{ef} = \int_{A(t)} D^4 dA. \tag{4}$$

The integral (4) can be written using spherical variables (θ, φ), which are the incidence and azimuth angles of the incident wave. However, for directional sources not vertically oriented, it is more convenient to describe their directivity and analyze the integral (4) in angular coordinates measured not from vertical direction, but from the transducer MRA, i.e., from its depression incidence angle, θ_0. Consider new variables (not spherical), $\theta' = \theta - \theta_0$ and $\varphi' = \varphi \sin\theta$, which represent angular coordinates measured in two mutually orthogonal planes so that their intersection is the transducer MRA, where $\theta' = 0$ and $\varphi' = \varphi = 0$. One of these two planes is always vertical (the incidence plane). The second is an auxiliary one perpendicular to the incidence plane and referred to as the slant azimuth plane.

Using the new variables, the integral (4) can be written in the form

$$A_{\mathrm{ef}} = \frac{r^2}{\cos\theta_w}\Omega_{\mathrm{ef}}, \tag{5}$$

where Ω_{ef} is an effective solid angle of the observation area

$$\Omega_{\mathrm{ef}} = \int_{\theta_1}^{\theta_2} d\theta \int_{-\pi\sin\theta}^{\pi\sin\theta} D^4(\theta - \theta_0, \varphi')\, d\varphi'. \tag{6}$$

Note that the integral (6) at high enough frequencies (and correspondingly narrow directivities) has a maximal value at $\theta_w = \theta_0$, where the observation angle window matches with the maximum of the transducer directivity.

Assume that the transducer has directivity with an axial symmetry, so that for the directivity factor we have $D^4(\theta', \varphi') = Q(\vartheta/\vartheta_o)$, where ϑ is the angle between the direction of observation and the transducer MRA, and ϑ_0 is a parameter defining the directivity angular width. Note that at $\vartheta_0 \ll \theta$, limits in the φ'-integral in Eq. (6) can be replaced by infinities, and integration can be performed using an approximation $\vartheta \approx (\theta'^2 + \varphi'^2)^{1/2}$. Then, in the matching case, $\theta_w = \theta_0$, we obtain

$$\Omega_{\mathrm{ef}} \approx \vartheta_0^2 P\left(\frac{\theta_2 - \theta_1}{2\vartheta_0}\right), \quad P(\alpha) = \int_{-\alpha}^{\alpha} dx \int_{-\infty}^{\infty} Q(\sqrt{x^2 + y^2})\, dy. \tag{7}$$

The integral in Eq. (7) has two simple asymptotic approximations:

$$P \approx M_1 \alpha, \quad \alpha \ll 1, \quad M_1 = \int_{-\infty}^{\infty} Q(s)\, ds,$$

$$P \approx M_2, \quad \alpha \gg 1, \quad M_2 = 2\pi \int_0^{\infty} sQ(s)\, ds. \tag{8}$$

A uniform approximation (for arbitrary α) can be sought in the form of a scaling expression

$$P(\alpha) \approx M_2 P_\eta(\alpha M_1/M_2), \quad P_\eta(x) = (1 + x^{-\eta})^{-1/\eta}, \quad \eta > 0. \tag{9}$$

At any value of η, this approximation has two correct asymptotes [Eq. (8)] for small and large α. This allows use of η as a free parameter that can be chosen so as to provide the best fit of the scaling function [Eq. (9)] to the integral (7) at intermediate α, taking into account a specific transducer directivity function $Q(s)$.

In our case of the circular piston transducers, we can assume that at high frequencies $\vartheta_0 \equiv (ka)^{-1} \ll 1$ and perform the integration in Eqs. (7) and (8) numerically with $Q(s) = (2J_1(s)/s)^4$. For the integrals in Eq. (8) we obtain $M_1 \approx 2.4$ and $M_2 \approx 5.8 \approx M_1^2$. The result for the integration in Eq. (7) is shown in Fig. 1 by the solid curve. The uniform scaling approximation given in Eq. (9) is shown for different η by dashed curves. It is seen that the best agreement is provided by $\eta \approx 4$. Therefore, the set of Eqs. (1), (3), (5), (7), and (9) provides a practical

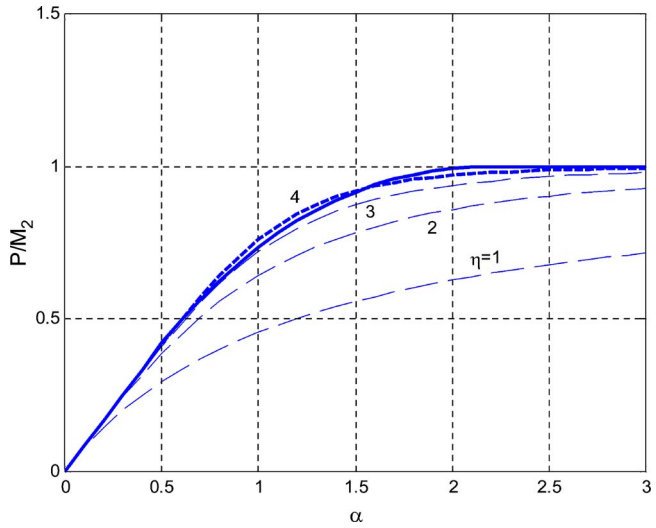


Fig. 1. (Color online) Choosing the parameter η for a scaling power law approximation (9): numerical integration for $P(\alpha)/M_2$ (solid curve) compared to the scaling function $P_\eta(\alpha M_1/M_2)$ for different η (dashed curves). The best fit is provided by $\eta \approx 4$.

algorithm for extracting the backscattering coefficient from analysis of the normalized scattered intensity in the frequency domain using time-windowing of the echo-signal.

4. Results

In Fig. 2, the frequency dependence of the backscattering strength is presented in a wide frequency range, 150 kHz to 8 MHz, for the two sediments, medium and coarse sand, at an incidence angle of 50° (grazing angle of 40°). It demonstrates, in particular, a noticeable difference in the scattering strength of the two sediments at frequencies below 2 MHz. For example, this

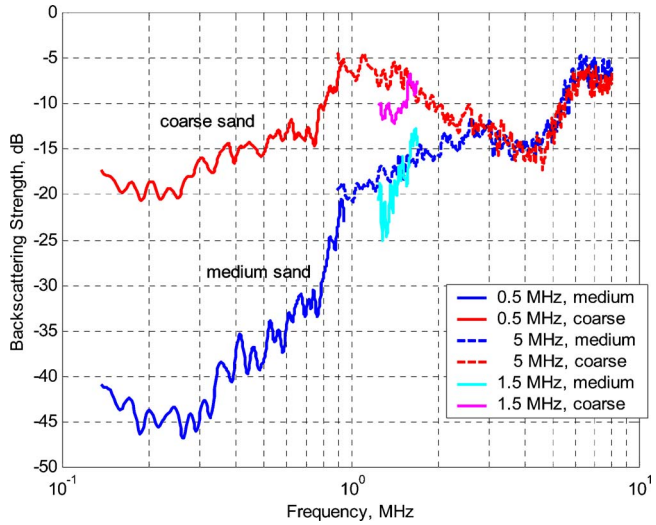


Fig. 2. (Color online) The backscattering strength at the incidence angle 50° obtained for medium and coarse sand using the broad band transducers with 500 kHz and 5 MHz nominal center frequencies. The center frequency of the 500 kHz transducer was nominal indeed, as the transducer had actually another, although weaker, maximum in its frequency spectrum, at 1.5 MHz, which was used to obtain additional scattering data.

difference at frequencies 400 kHz to 800 kHz is about 20 dB. It can mean practically important strong sensitivity of the scattering strength at these frequencies to the mean grain size of the sediment. Note also that at frequencies above 2.5 MHz the difference is small and, therefore, the scattering strength may have too little sensitivity to the grain size.

The ratio of the mean grain diameter to the acoustic wavelength in water, d/λ, varied, in our frequency range, from 0.024 to 1.3 for the medium sand, and from 0.15 to 8.3 for the coarse sand. Therefore, in the interval $0.15 < d/\lambda < 1.3$, we have two data sets for sediments with different mean grain size. This interval is especially interesting as it corresponds to a "transition frequency regime", where the acoustic properties of granular sediments can change dramatically. This change is due to transition from the "low frequency regime", where the ratio d/λ is very small and the continuum media assumptions are valid, to the "very high frequency regime", where this ratio is not small and the sediment must be considered as an essentially discrete granular medium. At these high frequencies, intrinsic (bulk) scattering due to the sediment granular structure becomes important. For example, the ratio d/λ becomes about 0.5 at frequencies around 3 MHz for the medium sand, and at 500 kHz for the coarse sand. Corresponding shift (for the sediments with different grain size) related to the "transition" effects and appearance of the new dominating scattering mechanism can be anticipated. In Fig. 2, such a scaling shift is clearly seen.

In Fig. 3, the scaling effect is demonstrated by plotting scattering data (same as in Fig. 2) versus the ratio d/λ. It is seen that the backscattering strength collapses to a function of only one parameter, at least over the interval $0.15 < d/\lambda < 1.3$, where both data sets are available. This confirms the assumption that, for the given conditions, only one mechanism of scattering is dominating, which is the intrinsic (bulk) scattering due to the sediment granular structure, and that the controlling sediment parameter is the mean grain size. In this case, the scattering strength must be a unique function of the ratio d/λ. This explains, e.g., that the two different grain-size cases match when shifted horizontally, but no shift in the vertical is needed (see Figs. 2 and 3). This is because the scattering strength is unique, system independent, and controlled only by the sediment parameters. This result is practically important, as it provides both prediction and inversion capabilities: if the scattering strength is measured at one frequency and one grain size, it can be predicted at others or inverted with respect to the grain sizes. Therefore, the scaling effects demonstrated here can potentially be used in acoustic sensing of marine sedi-

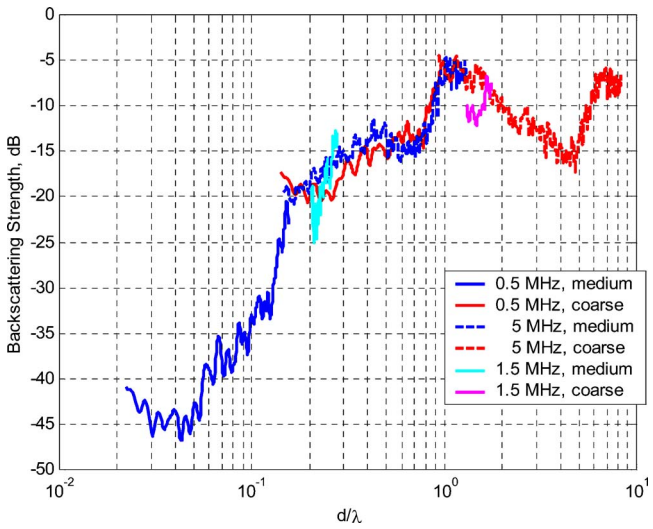


Fig. 3. (Color online) The same data as in Fig. 2, but plotted vs the sediment mean grain diameter ratio to the wavelength in water, d/λ.

ments to remotely estimate their mean grain size. Further analysis of these effects might be an interesting and promising subject of future theoretical and experimental studies.

Acknowledgments

This work was supported by the U.S. Office of Naval Research, and CNRS/LMA, France.

References and links

[1]D. R. Jackson and M. D. Richardson, *High Frequency Seafloor Acoustics* (Springer Science, New York, 2007).
[2]C. F. Greenlaw, D. V. Holliday, and D. E. McGehee, "High-frequency scattering from saturated sand sediments," J. Acoust. Soc. Am. **115**, 2818–2823 (2004).
[3]A. N. Ivakin, "High frequency scattering from sandy sediments: Roughness vs discrete inclusions," in *Boundary Influences in High Frequency Shallow Water Acoustics*, edited by N. G. Pace and P. Blondel, (University of Bath, UK, 2005), pp. 185–192.
[4]A. N. Ivakin, "Models of scattering for remote acoustic sensing of the seafloor," in *Proceedings of the Institute of Acoustics* (Bath University, Bath, UK, 2001), Vol. 23, Part 2, pp. 268–275.
[5]E. I. Thorsos, K. L. Williams, D. Tang, and S. G. Kargl, "SAX04 overview," in *Boundary Influences in High Frequency Shallow Water Acoustics*, edited by N. G. Pace and P. Blondel (University of Bath, UK, 2005), pp. 3–12.
[6]M. D. Richardson, K. B. Briggs, A. H. Reed, W. C. Vaughan, M. A. Zimmer, L. D. Bibee, and R. I. Ray, "Characterization of the environment during SAX04: Preliminary results," in *Proceedings of the International Conference on Underwater Acoustic Measurements: Technologies & Results*" (Heraklion, Crete, Greece, 28 June–1 July 2005).
[7]A. W. Nolle, W. A. Hoyer, J. F. Mifsud, W. R. Runyan, and M. B. Ward, "Acoustical properties of water-filled sands," J. Acoust. Soc. Am. **35**, 1394–1408 (1963).
[8]K. L. Williams, R. H. Hackman, and D. H. Trivett, "High-frequency scattering from liquid/porous sediment interfaces," J. Acoust. Soc. Am. **84**, 760–770 (1988).

 [DOI: 10.1121/1.2794864] Published Online 2 October 2007

Retrieving reflection responses by crosscorrelating transmission responses from deterministic transient sources: Application to ultrasonic data

Deyan Draganov, Kees Wapenaar, and Jan Thorbecke
Department of Geotechnology, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
d.s.draganov@tudelft.nl, c.p.a.wapenaar@tudelft.nl, j.w.thorbecke@tudelft.nl

Osamu Nishizawa
Institute for Advanced Industrial Science and Technology, Central 7, 1-1-1 Higashi, Tsukuba, Ibaraki, 305-8567, Japan
osamu-nishizawa@aist.go.jp

Abstract: By crosscorrelating transmission recordings of acoustic or elastic wave fields at two points, one can retrieve the reflection response between these two points. This technique has previously been applied to measured elastic data using diffuse wave-field recordings. These recordings should be relatively very long. The retrieval can also be achieved by using deterministic transient sources with the advantage of using short recordings, but with the necessity of using many *P*-wave and *S*-wave sources. Here, it is shown how reflections were retrieved from the cross correlation of transient ultrasonic transmission data measured on a heterogeneous granite sample.

PACS numbers: 43.20.Gp, 43.35.Cg, 43.60.Tj [AN]
Date Received: July 17, 2007 **Date Accepted:** August 27, 2007

1. Introduction

The cross correlation of observed wave fields (acoustic or elastic) at two points retrieves the Green's function at one of those points as if there were an impulsive source at the other. This was proven by researchers in different fields of science using different underlying theories.[1–7] Some of the theories require that the recorded wave fields are diffuse. These diffuse wave fields can be produced by uncorrelated noise sources and/or by multiple scattering inside or at the boundaries of a medium. Until now, most of the applications to measured data made use of diffuse fields. By crosscorrelating recorded coda waves and ambient noise, researches could retrieve surface waves,[8,9] diving *P*-waves,[10] and reflected body waves.[11] It was shown that for good retrieval results one requires recordings several orders longer than the normal recordings between the points of interest would be. This long recording time is required to reduce the amount of correlation between the different noise sources as much as possible or, in case of multiple scattering, to approximate the equipartitioned state as well as possible.

Other theories do not require diffuse wave fields, but instead make use of multitude of independent deterministic transient source responses. One of the advantages of using this method is that the recordings to be cross correlated do not need to be disproportionally long: The recording length is dictated by the expected arrival time of the latest event one wants to retrieve. The number of the recorded responses from transient sources multiplied by the length of each recording will still be much shorter than the required recording time in case of diffuse wavefields. To retrieve the Green's function between two receivers in an open system, these receivers should be completely surrounded by sources.[12,13] When this is not the case, spurious arrivals may appear.[14] To avoid this, a waveguide can be utilized.[15–17] Another solution is to retrieve parts of the Green's function from specific source-receiver geometries. One such possibility is to retrieve the reflection response at a free surface from the cross correlation of trans-

mission responses from sources illuminating the receivers from one side only. This can be very useful, for example, in seismology, where one can crosscorrelate transmission responses from subsurface sources (earthquakes) measured at the surface along receiver arrays to retrieve the earth's reflection response as if from sources at the surface. Having reflection responses would allow the application of advanced imaging techniques to obtain images of the subsurface with a higher resolution than those obtained by tomographic inversion of the transmission data. In the following, we test the method of retrieval of reflection responses through cross correlation of transmission data from deterministic transient sources. We apply the method to ultrasonic data measured on an inhomogeneous granite sample.

2. Theory

Let one have three-dimensional (3D) inhomogeneous, lossless, and source-free medium bounded by a free surface. In this medium, one chooses a domain D enclosed by a boundary $\partial D = \partial D_0 \cup \partial D_1$ with an outward-pointing normal vector. The boundary ∂D_0 coincides with (a part of) the free surface, while ∂D_1 is an arbitrary-shaped surface inside the medium. If there would be an impulsive force source f acting in the $\mathbf{x}_q$-direction ($q=1$, 2, or 3) at $\mathbf{x}_B = (\mathbf{x}_{B,1}, \mathbf{x}_{B,2}, \mathbf{x}_{B,3})$, then at $\mathbf{x}_A = (\mathbf{x}_{A,1}, \mathbf{x}_{A,2}, \mathbf{x}_{A,3})$ one could record the impulse response, i.e., Green's function of the medium $G^{v,f}_{p,q}(\mathbf{x}_A, \mathbf{x}_B, t)$, representing the particle velocity v observed in the $\mathbf{x}_p$-direction. If $\mathbf{x}_A$ and $\mathbf{x}_B$ are chosen in D or on ∂D_0, and the medium along and outside ∂D_1 is taken homogeneous with propagation velocity c^P for P-waves and c^S for S-waves and density ρ, then it can be shown that[13]

$$G^{v,f}_{p,q}(\mathbf{x}_A, \mathbf{x}_B, t) + G^{v,f}_{p,q}(\mathbf{x}_A, \mathbf{x}_B, -t) \approx \frac{2}{\rho c^P} \int_{\partial D_1} G^{v,P}_{p}(\mathbf{x}_A, \mathbf{x}, -t) * G^{v,P}_{q}(\mathbf{x}_B, \mathbf{x}, t) d^2\mathbf{x} + \frac{2}{\rho c^S} \int_{\partial D_1} G^{v,S}_{p,k}(\mathbf{x}_A, \mathbf{x}, -t) * G^{v,S}_{q,k}(\mathbf{x}_B, \mathbf{x}, t) d^2\mathbf{x}. \quad (1)$$

As the Green's function is a causal function of time, it does not interfere with its time-reversed version and one can simply take the causal part in the left-hand side of Eq. (1) to obtain the Green's function. In the right-hand side, the Green's functions denote observed particle velocities in the $\mathbf{x}_p$- and $\mathbf{x}_q$-directions at $\mathbf{x}_A$ and $\mathbf{x}_B$, respectively, due to impulsive sources at $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ along ∂D_1; the superscripts P and S denote sources of P-waves and sources of S-waves with different polarizations along the three axes; these polarizations are denoted by the subscript k and a summation is carried out over this subscript according to Einstein's summation convention for repeated subscripts. The asterisk denotes convolution, but as one convolves a signal with a time-reversed signal, one performs correlation. Hence, Eq. (1) shows that to retrieve the Green's function between the observation points $\mathbf{x}_A$ and $\mathbf{x}_B$, one needs at these points separate recordings from P- and S-wave sources. In actual implementations, this would require decomposition of the recorded wave fields into wave fields from P- and S-wave sources. When this is not done, spurious events will appear in the retrieved results.

3. Experimental setup and results

To test Eq. (1) we use a controlled laboratory experiment with an Oshima granite block with dimensions $300 \times 300 \times 100$ mm, in which there is a cylindrical hole of 150 mm in length and 15 mm in diameter filled with epoxy [see Fig. 1(a)]. The Oshima granite consists mainly of quartz, plagioclase, and biotite with grain sizes mostly from 1 to 5 mm. The estimated average P-wave and S-wave propagation velocities for the block are 4500 and 2700 m/s, respectively. The estimated P-wave and S-wave velocities for the epoxy are 2300 and 1400 m/s, respectively. For simplicity, we consider a two-dimensional (2D) experiment, using observation and source line arrays at opposite sides of the block lying in a plane perpendicular to the long axis of the cylinder. This means that in the second integral in Eq. (1), k takes on only the value of 2 meaning that we will need to record the response from P- and SV-sources. The observation array of 101 measurement points is placed along the free surface on the front wall of the granite

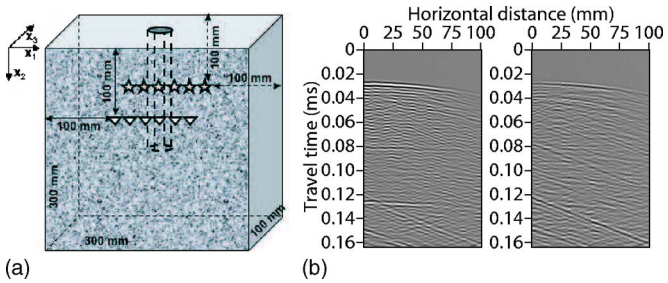


Fig. 1. (Color online) (a) Oshima granite block with dimensions $300 \times 300 \times 100$ mm. A cylindrical hole with a length of 150 mm and a diameter of 15 mm was drilled in the middle of one of the small walls and was filled with epoxy. An observation array of 101 measurement points lies on the front wall with a distance between the points of 1 mm. A source array of 21 source points lies on the back wall with a distance between the source points of 5 mm. The observation and source arrays are parallel to each other and are both perpendicular to the epoxy cylinder. (b) Observed transmission panels from a source transducer at 0 mm, i.e., at the left end of the array. The left-hand panel shows the observed normal particle velocity due to a compressional source transducer and the right-hand panel shows the observed normal particle velocity due to a shear source transducer. The recordings were 0.164 ms long.

sample; the distance between the observation points is 1 mm. At these points we measure only the normal component of the particle velocity. This means that in Eq. (1) the subscripts p and q take on the value of 3 [note the coordinate system in Fig. 1(a)]. The measurements at the observation points are performed using a laser Doppler vibrometer (LDV). Details about the LDV can be found in Nishizawa *et al.*[18] and in Sivaji *et al.*[19]

Compared to measurements in the field, in a controlled laboratory experiment one can use ultrasonic compressional and shear transducers to approximate separate *P*- and *S*-wave sources. The source array, consisting of 21 source points with 5 mm distance between the individual points, is placed along the back wall of the granite block. The sources are fed with a single-cycle 250 kHz sine wavelet. For this frequency, the wavelengths for *P*-waves and *S*-waves are 18 and 11 mm, respectively. At this scale, the granite sample is effectively heterogeneous and both compressional and shear waves are diffracted from the grains. Due to the finite diameter (5 mm) of the source transducers, their radiation pattern will not be like the desired point-source pattern.[18,20] This means that the compressional transducer will produce some amount of *S*-wave energy, especially at larger angles. Still, for the chosen source-receiver geometry the main energy will be radiated in the form of *P*-waves. A similar effect holds for the shear transducers, which will produce some *P*-wave energy.

Using stationary phase analysis[13] it can be shown that for the chosen 2D setup, sources placed along the walls perpendicular to the observation array would contribute strongly to the retrieval of surface waves, while sources placed along the wall opposite to the observation array (as in our experiment) will mainly contribute to the retrieval of reflection arrivals. We choose to concentrate on retrieving reflection arrivals as they can further be used to obtain a better image of the interior of the sample.

We perform the following measurements. A source transducer is placed and ignited at a certain source point and the resulting wave fields are measured at a point along the observation array. The electronic circuit of the LDV causes incoherent high-frequency noise to appear in the recorded wave forms. To eliminate this noise, we repeat the ignition of the source transducer at the same position 2000 times and at the observation point the 2000 recorded wave forms are summed to obtain a final wave form recording at that observation point. By placing a compressional transducer and recording the wave forms at all observation points in the way described earlier, we obtain a *P*-source transmission panel. Similarly, placing a shear transducer and repeating the measurements, we obtain a *S*-source transmission panel. This procedure is repeated for each of the 21 source points. Fig. 1(b) shows an example *P*-source transmission panel (left) with a clear *P*-wave direct arrival with an apex at about 0.024 ms and an example *S*-source transmission panel (right) with a clear *P*-wave direct arrival and a clear *S*-wave direct arrival

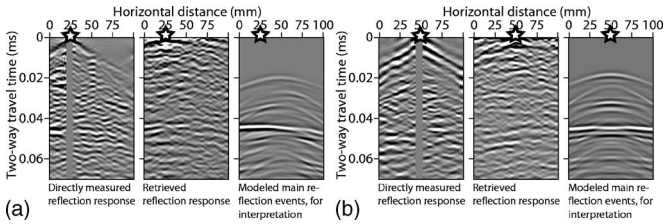


Fig. 2. (a) Comparison between the directly observed (left), retrieved through correlations (middle), and modeled with 2D elastic finite-difference scheme (right) reflection common-source panels for a source at 25 mm. (b) As in (a), but this time for a source at 50 mm. The modeled results were produced using a homogeneous background in which a circle was placed representing the epoxy cylinder. The numerically modeled results are only used for interpretation of expected reflection arrivals from the epoxy cylinder and the backwall, but are not representative for the clutter due to the inhomogeneous granite.

with an apex at around 0.039 ms. The transmission panels are 0.164 ms long aiming at recording the first few multiples of the direct *P*- and *S*-wave arrivals, i.e., the arrivals that have been reflected by both the front and the back wall in Fig. 1(a). Even though the later-arriving conversions and multiples are not readily interpretable due to scattering, parts of the first and the second multiple arrivals of the direct *P*-wave can be observed in the left panel in Fig. 1(b) (with estimated apices around 0.07 and 0.115 ms, respectively) and parts of the first multiple of the direct *S*-wave can be seen in the left-hand panel in Fig. 1(b) (with an estimated apex around 0.115 ms). On the panels we also see inclined linear arrivals caused by the presence of the left and right walls in Fig. 1(a). After cross correlation and integration [Eq. (1)], such events would contribute to retrieval of similar inclined events, which interfere with retrieved reflections. For this reason, we apply a frequency–wave number (f–k) filter to eliminate the inclined events. In the *P*-wave response, we also see nearly horizontal arrivals around 0.08 and 0.125 ms that arise from the presence of the to wall in Fig. 1(a) and representing 3D (out-of-plane) events. These arrivals cannot be filtered out as their behavior in the f–k domain is very similar to the direct transmission arrivals and their free-surface multiples. After cross correlation and integration, such out-of-plane events will cause nonphysical events to appear in the retrieved results.

By placing a source transducer at a position of an observation point and recording the resulting wave fields along the observation array, we obtain a reflection common-source panel. Such a panel will be used to verify possible retrieved reflections.

To implement Eq. (1), we do the following. We choose an observation point at a position where we would like to have a retrieved source, say $\mathbf{x}_B$. We take one *P*-source transmission panel [left-hand panel in Fig. 1(b)] and extract the trace $G_3^{v,P}(\mathbf{x}_B,\mathbf{x},t)$ at the position of $\mathbf{x}_B$ (the source is at $\mathbf{x}$). Then we correlate the trace with the whole panel, representing $G_3^{v,P}(\mathbf{x}_A,\mathbf{x},t)$ for a variable $\mathbf{x}_A$ and fixed $\mathbf{x}$, and the result is deconvolved with a wavelet extracted around zero time from the autocorrelation trace. The result, representing the first integrant in Eq. (1), is a correlation *P*-source panel. We repeat this for all *P*-source transmission panels, i.e., for all $\mathbf{x}$, by always correlating with the trace at the chosen position $\mathbf{x}_B$. Then, according to the first integral in Eq. (1), we sum the individual correlation *P*-source panels (i.e., integration along the source position $\mathbf{x}$). The previous procedure is repeated for all the *S*-source transmission panels as well. This is followed by summation of the outputs from the two integrals and extraction of the causal part – the result is a reflection common-source panel for a retrieved source position at $\mathbf{x}_B$.

After a reflection common-source panel is retrieved, it should be compared to a directly observed reflection common-source panel. The left-hand panels in Figs. 2(a) and 2(b) show the observed reflection common-source panel for a compressional transducer (a source is represented by a star) at 25 and 50 mm, respectively, after f–k filtering of the surface waves and deconvolution. The removal of the surface waves is not a trivial task as the inhomogeneities close to the surface cause dispersion. Due to the physical presence of the source transducer the near offsets could not be recorded. The middle pictures in Figs. 2(a) and 2(b) show the retrieved result using Eq. (1) for retrieved source positions at 25 and 50 mm, respectively. The presence

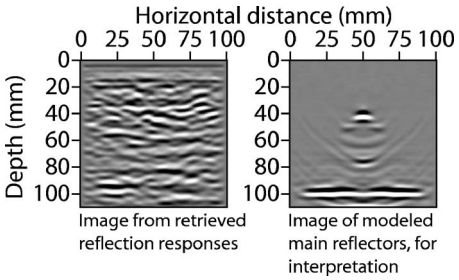


Fig. 3. Comparison between images of the interior of the sample obtained from (left) reflection events retrieved through cross correlation and (right) numerically modeled reflection events. The images were obtained using a migration procedure with homogeneous velocity of 4500 m/s.

of inhomogeneities in the Oshima granite makes both pictures look cluttered and makes the interpretation and comparison of reflection events difficult. To be able to interpret the main events and compare the panels, we performed 2D elastic numerical modeling with a finite-difference scheme using a homogeneous background velocity model with a circle in it representing the cross section of the epoxy cylinder. The dimensions and velocities of the model were the same as the dimensions and the estimated average velocities of the granite block. The right-hand pictures in Figs. 2(a) and 2(b) show the modeling results (without the surface waves) for the same source positions. Note that these modeling results should only be used for interpretation of expected reflection arrivals. The modeling results are not representative for the clutter due to the inhomogeneous background of the Oshima granite block. This means that a reflection arrival, visible in a modeled common-source panel, might be (partly) obscured by diffractions in the observed and the retrieved common-source panels.

In the numerically modeled results in Fig. 2, the most prominent events are the top-of-cylinder reflection arrivals with apices around 0.018 ms, the bottom-of-cylinder reflection arrivals with apexes around 0.035 ms, and the reflection arrivals from the granite's backwall, which arrive around 0.045 ms. The backwall reflection event is followed by a train of conversions. Due to the missing offsets and the remnants of the surface waves, it is hard to interpret the top-of-cylinder reflection in the directly observed reflection common-source panels. Parts of this reflection might be present for the source position at 50 mm. Another factor making the interpretation of the top-of-cylinder reflection difficult is the interference with linear events in the directly observed data coming from the left and right walls of the granite block and remaining after the $f-k$ filtering. Such events are easier to remove in the transmission panels and after cross correlation and summation their effect on the retrieved results is much smaller. On the retrieved reflection common-source panels, the apices of the top-of-cylinder reflection are readily visible and the wings of the reflection can be traced from 0 mm until the apex in Fig. 2(a) and from the apex until about 85 mm in Fig. 2(b). The backwall reflection in the retrieved common-source panels is at least as good as the reflection in the observed common-source panels. In Fig. 2(b) the retrieved backwall reflection is even more continuous than in the directly observed reflection panel. The bottom-of-cylinder reflections are not interpretable in both the directly observed and the retrieved common-source panels. In the retrieved reflection common-source pictures, we can see linear nonphysical events, like the ones at around 0.05 ms, which are a result from the presence of the out-of-plane events in the transmission data.

Having retrieved reflection common-source panels at the position of every receiver, we can apply advanced imaging techniques to look into the interior of the granite sample. The left-hand panel in Fig. 3 shows the imaging result from the application of prestack depth migration to the retrieved reflection events using a homogeneous model with a velocity of 4500 m/s, i.e., the estimated average propagation velocity of *P*-waves through the granite sample. The image is quite cluttered, as it can be expected for an image of an inhomogeneous sample, and because of this it is difficult to describe it in terms of imaged grains of the granite sample, imaged epoxy cylinder, imaged back wall of the sample, and artifacts due to the out-of-plane

events. To verify if we could image grains in the sample, we would need directly observed reflection measurements from a source at each of the receiver positions, which we do not have. To facilitate the interpretation of the left-hand image in Fig. 3, in the right-hand panel in Fig. 3 we show the obtained image from the migration of the numerically modeled results. This image can be used to identify only the main reflectors in the granite block—the epoxy cylinder and the backwall of the granite, but is not representative for imaged inhomogeneities (grains) inside the granite block. Comparing the two images in Fig. 3, we see that by using only a few sources at the backwall of the block, we have successfully imaged the top of the cylinder and the sample's backwall.

4. Conclusions

We showed how one can retrieve the Green's function between two receivers from the cross correlation of responses from deterministic transient sources. We showed results from the application of this method to ultrasonic data from an inhomogeneous granite block with an epoxy cylinder in it. We successfully retrieved reflection arrivals from the cross correlation of transmission measurements from separate compressional and shear sources. The retrieved reflection results allowed us to use an advanced imaging technique. We applied imaging using a homogeneous velocity model and successfully identified the main reflectors in the granite sample, i.e., the top of the epoxy cylinder and the back wall of the granite sample.

Acknowledgments

This research is supported by The Netherlands Research Centre for Integrated Solid Earth Sciences ISES; the Technology Foundation STW, Applied Science Division of The Netherlands Organization for Scientific Research NWO and the technology program of the Ministry of Economic Affairs (Grant No. DTN.4915). The measurement of the data was partly sponsored by NWO through a travel grant. The authors would like to thank Dr. J. Spetzler for helping with the measurements and Dr. X. Campman for the very helpful discussion.

References and links

[1]J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," Geophysics **33**, 264–269 (1968).
[2]G. T. Schuster, "Theory of daylight/interferometric imaging: tutorial," in *Proceedings of the 63rd EAGE Conference and Exhibition (European Association of Geoscientists and Engineers, Houten, 2001)*, 2001, Extended Abstracts A-32.
[3]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301-1–134301-4 (2001).
[4]K. Wapenaar, J. Thorbecke, D. Draganov, and J. Fokkema, "Theory of acoustic day-light imaging revisited," in *Proceedings of the 72nd Annual SEG Meeting (Society of Exploration Geohysicists, Tulsa, 2002)*, 2002, Expanded Abstracts ST 1.5.
[5]A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[6]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610-1–046610-8 (2004).
[7]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[8]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[9]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[10]P. Roux, K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, "P-waves from cross-correlation of seismic noise," Geophys. Res. Lett. **32**, L19303, 10.1029/2005GL023803 (2005).
[11]D. Draganov, K. Wapenaar, W. Mulder, J. Singer, and A. Verdel, "Retrieval of reflections from seismic background-noise measurements," Geophys. Res. Lett. **34**, L04305, 10.1029/2006GL028735 (2007).
[12]K. Wapenaar, J. Fokkema, and R. Snieder, "Retrieving the Green's function in an open system by cross-correlation: a comparison of approaches," J. Acoust. Soc. Am. **118**, 2783–2786 (2005).
[13]K. Wapenaar and J. Fokkema, "Green's functions representations for seismic interferometry," Geophysics **71**, SI33–SI46 (2006).
[14]R. Snieder, K. Wapenaar, and K. Larner, "Spurious multiples in seismic interferometry of primaries," Geophysics **71**, SI111–SI124 (2006).
[15]P. Roux and M. Fink, "Green's function estimation using secondary sources in a shallow water experiment," J.

Acoust. Soc. Am. **113**, 1406–1416 (2003).
[16]K. van Wijk, "On estimating the impulse response between receivers in a controlled ultrasonic experiment," Geophysics **71**, SI79–SI84 (2006).
[17]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[18]O. Nishizawa, T. Satoh, X. Lei, and Y. Kuwahara, "Laboratory studies of seismic wave propagation in inhomogeneous media using laser dopler vibrometer," Bull. Seismol. Soc. Am. **87**, 809–823 (1997).
[19]C. Sivaji, O. Nishizawa, and Y. Fukushima, "Relations between fluctuations of arrival time and energy of seismic waves and scale length of heterogeneity: An inference from experimental study," Bull. Seismol. Soc. Am. **91**, 292–303 (2001).
[20]X. M. Tang, Z. Zhu, and M. N. Toksoz, "Radiation patterns of compressional and shear transducers at the surface of an elastic half-space," J. Acoust. Soc. Am. **95**, 71–76 (1994).

Transfer matrix of conical waveguides with any geometric parameters for increased precision in computer modeling

Yakov Kulik
School of Physics, University of New South Wales, Sydney 2052, Australia
ykulik@phys.unsw.edu.au

Abstract: The existing formula for the transfer matrix of conical elements assumes constant wave number, which is only valid for sufficiently short conical elements. In acoustic waveguides, the phase velocity, attenuation constant, and hence complex wave number depend on frequency and cross-section radius. As for conical waveguides, the cross-section radius is position dependent, the transfer matrix must allow for a position-dependent wave number. Taking this into account, this letter presents an analytic derivation of the transfer matrix for conical waveguides with any geometric parameters, which can be utilized to improve the method of computer modeling of complex waveguides.

PACS numbers: 43.20.Mv, 43.58.Ta, 43.20.Bi [RW]
Date Received: August 1, 2007 **Date Accepted:** September 1, 2007

1. Introduction

The pressure and volume flow at one end of a waveguide are related to the pressure and volume flow at the other end by a transfer matrix.[1–3] This transfer matrix also determines the relation between input impedance and load impedance. Acoustic waveguides are modeled as a series of cylindrical and conical elements. Such analyses have applications in computer modeling of air-conditioning ducts,[4] exhaust mufflers,[5] and wind instruments in musical acoustics.[6,7] The phase velocity v and attenuation constant α, and hence the complex wave number k, depend on the cross-section radius R as well as on the frequency f. As the radius of the cross section is constant throughout a cylindrical element, k is constant, and the transfer matrix is given by Olson[2] and Fletcher and Rossing.[3] An expression for the transfer matrix of a conical element, neglecting the dependence on cross-section radius and assuming constant complex wave number, is given by Olson[2] and Fletcher and Rossing.[3] This letter presents a derivation of the transfer matrix of a conical element, of any length, any inlet and outlet radii, including the dependence of the complex wave number k on the cross-section radius R, where for a conical element, the cross-section radius is not constant but a function of position, in contrast to the case of a cylindrical element.

As a result of viscous and thermal wall losses, which increase as the radius is decreased, the complex wave number k in an acoustic waveguide is a function of cross-section radius R and for propagation of sound waves in air for $r_v > 10$ and to a good approximation for $r_v > 3$, this dependence is given[1,3,8] by the following relation:

$$k = \frac{2\pi}{\lambda} - i\alpha = \frac{2\pi f}{v(R)} - i\alpha(R) = \frac{2\pi f}{c}\left(1 - \frac{\delta}{Rf^{1/2}}\right)^{-1} - i\frac{\varepsilon f^{1/2}}{R}, \tag{1}$$

where $\delta = 1.65 \times 10^{-3}$ m Hz$^{1/2}$ and $\varepsilon = 3 \times 10^{-5}$ s$^{1/2}$. (The nondimensional parameter $r_v = R/(\eta/\omega\rho)^{1/2}$ is the ratio of the waveguide's cross-section radius to the viscous boundary layer thickness, which is equal to the square root of the dynamic viscosity η divided by the product of the angular frequency ω and density ρ.)

The existing computational method for modeling complex waveguides using transfer matrices consists of dividing them into cylindrical and conical elements, which make up the

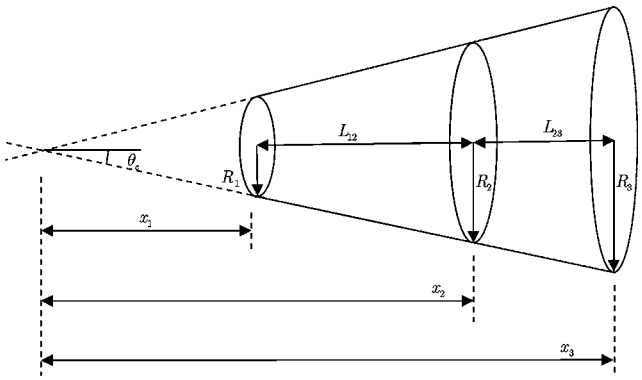


Fig. 1. Two successive conical elements of the same cone angle, from the same truncated cone with the geometric parameters of each conical element shown.

entire waveguide and for each frequency, numerically multiplying the transfer matrices to determine the transfer matrix of the whole complex waveguide. For computational methods of modeling waveguides and ducts, the reader is referred to.[9–11] For applications of computational methods for modeling wind instruments in musical acoustics, refer to.[3,6,7,12–14] Applications of numerical techniques in the design of mufflers in exhaust systems are discussed in Refs. 15 and 16.

The formula derived in this letter, for the transfer matrix of a conical waveguide has a convenient algebraic form and can be used to improve the method of computer modeling of acoustic systems, including air-conditioning ducts, exhaust mufflers and wind instruments in musical acoustics, increasing precision and reducing computation time. Using this formula, the transfer matrix for a long conical waveguide or a long conical component of a waveguide can be calculated accurately without the need to split it into short conical elements and multiply through a large number of transfer matrices numerically for each frequency, to achieve accurate results.

2. Analytic derivation

For a conical element of infinitesimal length, the complex wave number is constant. A finite conical waveguide may be analyzed as a set of infinitesimal conical elements. Fig. 1 shows two successive conical elements with the same cone angle.

Let the first conical element be between cross sections at distances x_1 and x_2 from the cone apex with length $L_{12}=x_2-x_1$ and effective wave number k_{12}. The transfer matrix for the first conical element may be written (in accordance with Ref. 3) as:

$$M_{12}=\frac{x_2}{x_1}\begin{pmatrix} -t_2\sin(k_{12}L_{12}-\theta_2) & i\sin(k_{12}L_{12}) \\ it_1t_2\sin(k_{12}L_{12}-\theta_2+\theta_1) & t_1\sin(k_{12}L_{12}+\theta_1)\end{pmatrix}$$
$$=\frac{x_2}{x_1}\begin{pmatrix} t_2\cos\left(k_{12}L_{12}-\left(\theta_2-\frac{\pi}{2}\right)\right) & i\sin(k_{12}L_{12}) \\ it_1t_2\sin(k_{12}L_{12}-\theta_2+\theta_1) & t_1\cos\left(k_{12}L_{12}+\left(\theta_1-\frac{\pi}{2}\right)\right)\end{pmatrix}. \quad (2)$$

In addition to those parameters already mentioned, the transfer matrix depends on several parameters characterizing the two ends of the conical element: $\theta_1=\arctan(k_1x_1)$, $\theta_2=\arctan(k_2x_2)$, $t_1=1/\sin\theta_1$ and $t_2=1/\sin\theta_2$.

The transfer matrix links the pressure P_1 and volume flow U_1 at one end of the conical element to the pressure P_2 and volume flow U_2 at the other end, via the following relation:

$$\begin{pmatrix} P_1 \\ Z_1 U_1 \end{pmatrix} = M_{12} \begin{pmatrix} P_2 \\ Z_2 U_2 \end{pmatrix}. \tag{3}$$

In this relation, $Z_1 = \rho c / S_1$ and $Z_2 = \rho c / S_2$ are the characteristic impedances at each of the two ends of the conical element, where S_1 and S_2 are the respective cross-sectional areas at each of the two ends, ρ is the density of the medium, c is the speed of sound in the medium. Analogous relations to Eqs. (2) and (3) define the transfer matrix of the second conical element.

The transfer matrix M_{13} for the combination of two conical elements is found by matrix multiplication of the two transfer matrices M_{12} and M_{23} of the two conical elements. By using trigonometric identities for all four entries of this resulting matrix, the transfer matrix for the combination of 2 conical elements simplifies to the following form:

$$\begin{aligned} M_{13} &= M_{12} M_{23} \\ &= \frac{x_3}{x_1} \begin{pmatrix} -t_3 \sin(k_{12}L_{12} + k_{23}L_{23} - \theta_3) & i \sin(k_{12}L_{12} + k_{23}L_{23}) \\ i t_1 t_3 \sin(k_{12}L_{12} + k_{23}L_{23} + \theta_1 - \theta_3) & t_1 \sin(k_{12}L_{12} + k_{23}L_{23} + \theta_1) \end{pmatrix} \\ &= \frac{x_3}{x_1} \begin{pmatrix} -t_3 \sin(\zeta_{12} + \zeta_{23} - \theta_3) & i \sin(\zeta_{12} + \zeta_{23}) \\ i t_1 t_3 \sin(\zeta_{12} + \zeta_{23} + \theta_1 - \theta_3) & t_1 \sin(\zeta_{12} + \zeta_{23} + \theta_1) \end{pmatrix} \\ &= \frac{x_3}{x_1} \begin{pmatrix} -t_3 \sin(\zeta_{13} - \theta_3) & i \sin(\zeta_{13}) \\ i t_1 t_3 \sin(\zeta_{13} + \theta_1 - \theta_3) & t_1 \sin(\zeta_{13} + \theta_1) \end{pmatrix}. \end{aligned} \tag{4}$$

This has the same form as the matrix for a single conical element with $\zeta = kL$ replaced by $\zeta_{13} = \zeta_{12} + \zeta_{23} = k_{12}L_{12} + k_{23}L_{23}$, where $\zeta_{ij} = k_{ij}L_{ij}$ may be considered the non-dimensional length of the conical element, since it is an additive parameter for successive conical elements.

For a conical waveguide, the cross-section radius R depends on the position x relative to the cone apex. In order to allow for the dependence of the complex wave number k on the cross-section radius R, and therefore on the position x relative to the cone apex, it is necessary to first divide the conical waveguide into infinitesimal elements.

To this end, Eq. (4) is extended to N conical elements, by mathematical induction:

$$\begin{aligned} M_{0N} &= M_{01} M_{12} M_{23} M_{34} \cdots M_{N-1,N} \\ &= \frac{x_N}{x_0} \begin{pmatrix} -t_N \sin(\zeta_{01} + \zeta_{12} + \cdots + \zeta_{N-1,N} - \theta_N) & i \sin(\zeta_{01} + \zeta_{12} + \cdots + \zeta_{N-1,N}) \\ i t_0 t_N \sin(\zeta_{01} + \zeta_{12} + \cdots + \zeta_{N-1,N} + \theta_0 - \theta_N) & t_0 \sin(\zeta_{01} + \zeta_{12} + \cdots + \zeta_{N-1,N} + \theta_0) \end{pmatrix} \\ &= \frac{x_N}{x_0} \begin{pmatrix} -t_N \sin(\zeta_{0N} - \theta_N) & i \sin(\zeta_{0N}) \\ i t_0 t_N \sin(\zeta_{0N} + \theta_0 - \theta_N) & t_0 \sin(\zeta_{0N} + \theta_0) \end{pmatrix}, \end{aligned} \tag{5}$$

where

$$\begin{aligned} \zeta_{0N} &= \zeta_{01} + \zeta_{12} + \zeta_{23} + \zeta_{34} + \cdots + \zeta_{(N-1),N} \\ &= k_{01}L_{01} + k_{12}L_{12} + k_{23}L_{23} + k_{34}L_{34} + \cdots + k_{(N-1),N}L_{(N-1),N} \\ &= \sum_{i=0}^{N-1} k_{i,i+1} \Delta x_i, \end{aligned} \tag{6}$$

and where

$$\Delta x_i = L_{i,i+1} = x_{i+1} - x_i. \tag{7}$$

In the limit as N goes to infinity and $\Delta x_i \to 0$, while keeping fixed the endpoints of the conical waveguide and its length $L = L_{0,N} = \Sigma_{i=0}^{N-1} \Delta x_i$, the Riemann sum in Eq. (6) above becomes the integral

$$\int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx = L\left(\frac{1}{L}\int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx\right) = \bar{k}L, \tag{8}$$

where $L=x_{\text{out}}-x_{\text{in}}$ is the length of the finite conical waveguide and $\bar{k}:=(1/L)\int_{x_{\text{in}}}^{x_{\text{out}}}k(x)\,dx$ is the effective wave number of the finite conical waveguide.

Therefore, the transfer matrix for a finite conical waveguide with any geometric parameters is

$$\begin{aligned} M &= \frac{x_{\text{out}}}{x_{\text{in}}}\begin{pmatrix} -t_{\text{out}}\sin\left(\int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx - \theta_{\text{out}}\right) & i\sin\left(\int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx\right) \\ it_{\text{in}}t_{\text{out}}\sin\left(\int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx + \theta_{\text{in}} - \theta_{\text{out}}\right) & t_{\text{in}}\sin\left(\int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx + \theta_{\text{in}}\right) \end{pmatrix} \\ &= \frac{x_{\text{out}}}{x_{\text{in}}}\begin{pmatrix} -t_{\text{out}}\sin(\bar{k}L - \theta_{\text{out}}) & i\sin(\bar{k}L) \\ it_{\text{in}}t_{\text{out}}\sin(\bar{k}L + \theta_{\text{in}} - \theta_{\text{out}}) & t_{\text{in}}\sin(\bar{k}L + \theta_{\text{in}}) \end{pmatrix}. \end{aligned} \tag{9}$$

This formula for the transfer matrix of a conical waveguide of any geometrical parameters is the main result of this letter.

An explicit expression can be written for the transfer matrix, including the spatial dependence of the complex wave number arising from the position-dependent cross-section radius in a conical waveguide, by expressing k in terms of x using Eq. (1) and the relation $R=x\tan(\theta_c)$, where θ_c is the half-angle of the cone. Thus

$$k(x) = \frac{2\pi f}{c}\left(1 - \frac{\delta}{\tan(\theta_c)xf^{1/2}}\right)^{-1} - i\frac{\varepsilon f^{1/2}}{\tan(\theta_c)x}. \tag{10}$$

After some rearrangement, this gives

$$k(x) = \frac{2\pi f}{c}\left(1 + \frac{b}{x-b}\right) - i\frac{\varepsilon f^{1/2}}{\tan(\theta_c)}\frac{1}{x},$$

where $b=\delta/\tan(\theta_c)f^{1/2}$. This expression may be integrated to give

$$\begin{aligned} \int_{x_{\text{in}}}^{x_{\text{out}}} k(x)dx &= \frac{2\pi f}{c}\left(\int_{x_{\text{in}}}^{x_{\text{out}}} 1\,dx + \int_{x_{\text{in}}}^{x_{\text{out}}} \frac{b}{x-b}dx\right) - i\frac{\varepsilon f^{1/2}}{\tan(\theta_c)}\int_{x_{\text{in}}}^{x_{\text{out}}}\frac{1}{x}dx \\ &= \frac{2\pi f}{c}\left\{L + b\ln\left(\frac{x_{\text{out}}-b}{x_{\text{in}}-b}\right)\right\} - i\frac{\varepsilon f^{1/2}}{\tan(\theta_c)}\ln\left(\frac{x_{\text{out}}}{x_{\text{in}}}\right) \\ &= \frac{2\pi f}{c}\left\{L + \frac{\delta}{\tan(\theta_c)f^{1/2}}\ln\left(\frac{x_{\text{out}} - \frac{\delta}{\tan(\theta_c)f^{1/2}}}{x_{\text{in}} - \frac{\delta}{\tan(\theta_c)f^{1/2}}}\right)\right\} - i\frac{\varepsilon f^{1/2}}{\tan(\theta_c)}\ln\left(\frac{x_{\text{out}}}{x_{\text{in}}}\right) \end{aligned} \tag{11}$$

where $\delta=1.65\times10^{-3}$ m Hz$^{1/2}$ and $\varepsilon=3\times10^{-5}$ s$^{1/2}$. Substitution of Eq. (11) into Eq. (9) gives an algebraic expression for the transfer matrix of a finite conical waveguide with any geometric parameters.

3. Special case of sufficiently small conical elements

For a short conical element ($|L|\ll x_{\text{in}}$), the best one-term Taylor expansion approximation for $\ln(x_{\text{out}}/x_{\text{in}})$ in terms of the small parameter L/x^* is at $x^*=\bar{x}=(x_{\text{in}}+x_{\text{out}})/2$, where the error for

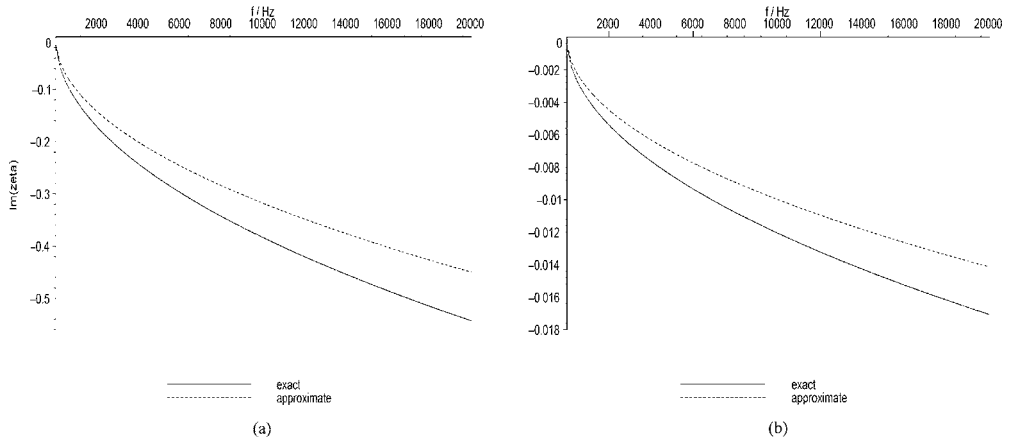


Fig. 2. Imaginary part of the approximate expression $\zeta_{\text{approx}}=k(\bar{x})L$ and the exact integral expression $\zeta_{\text{exact}}=\int_{x_{\text{in}}}^{x_{\text{out}}}k(x)dx$ for a conical waveguide with inlet radius 1 mm and outlet radius 5 mm having a length of (a) 318 mm and (b) 10 mm.

this approximation is third order and not second order as for an approximation at any other x^* between x_{in} and x_{out}. Consider the logarithm in the real part of the integral. The Taylor approximation requires the stronger condition that $|L|\ll|x^*-b|$, i.e. $|L|\ll|x^*-\delta/\tan(\theta_c)f^{1/2}|$. Even for a small conical element with $|L|\ll x^*$, the approximation may be invalid for sufficiently small frequencies or sufficiently small cone angles.

The approximation is a consistent over-approximation by $O((L/2(\bar{x}-b))^3)$ at $x^*=\bar{x}=(x_{\text{in}}+x_{\text{out}})/2$ and a consistent under-approximation or over-approximation by $O((L/(x^*-b))^2)$ for any other x^* between x_{in} and x_{out}. Thus, the best possible approximation retaining only the first term of the Taylor expansion is at $x^*=\bar{x}=(x_{\text{in}}+x_{\text{out}})/2$ with an error of order $O((L/2(\bar{x}-b))^3)$. Thus, for sufficiently small conical elements the best one-term Taylor approximation of the form $k(x^*)L$ for the integral expression is $k(\bar{x})L$ at the midpoint of the conical element.

Figure 2 shows for comparison, the imaginary part of the approximate expression $\zeta_{\text{approx}}=k(\bar{x})L$ and the exact integral expression $\zeta_{\text{exact}}=\int_{x_{\text{in}}}^{x_{\text{out}}}k(x)dx$ for particular values of the geometric parameters of a conical waveguide. Each expression is shown as a function of frequency and the difference between the two curves is clearly significant particularly for long and narrow waveguides.

4. Conclusion

The effective wave number of a conical waveguide cannot be determined only from the wave numbers at the two ends. Thus:

$$\bar{k}\neq\frac{k(x_{\text{in}})+k(x_{\text{out}})}{2},\quad \bar{k}\neq\sqrt{k(x_{\text{in}})k(x_{\text{out}})},\quad \bar{k}\neq\{[k(x_{\text{in}})]^{-1}+[k(x_{\text{out}})]^{-1}\}^{-1}.$$

The effective wave number depends not only on the wave number at the endpoints or at the midpoint of the waveguide, but in fact on the wave number at all points throughout the waveguide.

In this letter, the transfer matrix of a conical waveguide has been analytically derived, taking into account that the complex wave number depends on the cross-section radius, which is a function of position for a conical waveguide. As a result the complex wave number is a function of position in a conical waveguide, rather than a constant throughout its length as is the case for a cylindrical waveguide. Thus, the derived transfer matrix of a conical waveguide is valid for

any length, inlet and outlet radii, having taken into account the spatial dependence of the wave number. This is of particular interest in the computer modeling of acoustic systems, where high precision is important. This includes computer modeling of air-conditioning ducts,[4] exhaust mufflers,[5] and wind instruments in musical acoustics.[6,7]

Acknowledgment

The author would like to express his sincere gratitude to Professor Joe Wolfe and Associate Professor John Smith for discussion, valuable advice, and thorough editing of the manuscript. The author would also like to thank the reviewers and editor for valuable advice and comments.

References and links

[1]Lord J. W. Rayleigh, *The Theory of Sound* (Macmillan, London, 1887) (reprinted Dover, New York, 1945).
[2]H. F. Olson, *Acoustical Engineering* (Van Nostrand Company, Princeton, NJ, 1957), pp. 105 and 106.
[3]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1998), pp. 196, 198 and 210.
[4]Yu. V. Zakharov and L. M. Andreev, *Vessel Air Conditioning Systems Apparatus* (Shipbuilding, St. Petersburg, 1971) (in Russian).
[5]A. A. Skuridin and E. M. Mikheev, *Minimizing Noise and Vibration on a Vessel's Internal Combustion Engine* (Shipbuilding, St. Petersburg, 1970) (in Russian), p. 81.
[6]G. R. Plitnik and W. J. Strong, "Numerical method for calculating input impedances of the oboe," J. Acoust. Soc. Am. **65**, 816–825 (1979).
[7]A. Botros, J. Smith, and J. Wolfe, "The virtual Boehm flute—a web service that predicts multiphonics, microtones and alternative fingerings," Acoust. Aust. **30**, 61–65 (2002).
[8]A. H. Benade, "On the propagation of sound waves in a cylindrical conduit," J. Acoust. Soc. Am. **44**, 616–623 (1968).
[9]K. J. Baumeister, "Numerical techniques in linear duct acoustics—a status report," J. Eng. Ind. **103**, 270–281 (1981).
[10]A. Craggs, "The application of the transfer matrix and matrix condensation methods with finite elements to duct acoustics," J. Sound Vib. **132**, 393–402 (1989).
[11]R. J. Beckemeyer, "Transfer matrix for sound transmission in acoustically lined horns," J. Acoust. Soc. Am. **62**, S80–S81 (1977).
[12]R. Caussé, J. Kergomard, and X. Lurton, "Input impedance of brass musical instruments—comparison between experiment and numerical models," J. Acoust. Soc. Am. **75**, 241–254 (1984).
[13]J. D. Dudley and W. J. Strong, "A computer study of the effects of harmonicity in a brass wind instrument," Appl. Acoust. **30**, 116–132 (1990).
[14]N. Amir, G. Rosenhouse, and U. Shimony, "Input impedance of musical horns and the 'horn function'," Appl. Acoust. **38**, 15–35 (1993).
[15]B. S. Sridhara and M. J. Crocker, "Review of theoretical and experimental aspects of acoustical modeling of engine exhaust systems," J. Acoust. Soc. Am. **95**, 2363–2370 (1994).
[16]T. Kar and M. L. Munjal, "Generalized analysis of a muffler with any number of interacting ducts," J. Sound Vib. **285**, 585–596 (2005).

 [DOI: 10.1121/1.2784112] Published Online 2 October 2007

Urgency is a non-monotonic function of pulse rate

Frank A. Russo
Department of Psychology, Ryerson University, Toronto, Canada
russo@ryerson.ca

Jeffery A. Jones
Centre for Cognitive Neuroscience and Department of Psychology, Wilfrid Laurier University, Waterloo, Canada
jjones@wlu.ca

Abstract: Magnitude estimation was used to assess the experience of urgency in pulse-train stimuli (pulsed white noise) ranging from 3.13 to 200 Hz. At low pulse rates, pulses were easily resolved. At high pulse rates, pulses fused together leading to a tonal sensation with a clear pitch level. Urgency ratings followed a nonmonotonic (polynomial) function with local maxima at 17.68 and 200 Hz. The same stimuli were also used in response time and pitch scaling experiments. Response times were negatively correlated with urgency ratings. Pitch scaling results indicated that urgency of pulse trains is mediated by the perceptual constructs of speed and pitch.

PACS numbers: 43.66.Ba, 43.66.Hg [QJF]
Date Received: June 18, 2007 **Date Accepted:** August 5, 2007

1. Introduction

"*The ranges of perception are ranges of time, and the time is subdivided by us, by the construction of our bodies and by our organs of perception. And since these modern means have become available, to change the time of perception continuously, from one range to another, from a rhythm into a pitch, or a tone or noise into a formal structure, the composer can now work within a unified time domain*" (*p. 95*, *Stockhausen, 1990/1971*).

The effectiveness of an auditory warning is determined in part by its ability to convey urgency. Ideally, the warning will convey enough urgency so as to alert an observer and provoke some sort of action (Patterson, 1990). An auditory warning can be made more urgent in several ways, including increasing the pitch (frequency) and, in the case of warnings with multiple sound bursts, increasing the speed (i.e., decreasing the inter-onset time; Edworthy *et al.*, 1991; Guillaume *et al.*, 2003; Hellier *et al.*, 1993). Research has conventionally framed these two influences on the experience of urgency as independent and their effects have been compared directly using psychophysical methods. For example, applying Stevens' power law, Hellier and Edworthy (1999) found that larger changes to pitch than to speed were necessary in order to yield equivalent changes in experienced urgency (slopes of 0.37 and 1.35, respectively).

However, from a strictly temporal perspective, frequency and inter-onset time of auditory warnings may be interpreted as reciprocals of one another falling along different ends of the same rate-based continuum; that is, faster pulse rates (in terms of cycles per second or event onsets) tend to convey higher levels of urgency. Ecological psychoacoustics may provide a unifying framework for understanding rate-based predictors of urgency (Neuhoff, 2004). Specifically, it may be argued that it is not necessary to have mediation by mental representations of the physical world (such as conceptions about pitch or speed) in order to experience something as urgent or to act on it appropriately (Gibson, 1979). A related possibility is that the experience of urgency is based upon familiarity with probabilistic cues inherent in emotional speech. Listeners appear to perceive elevations in speech rate, pitch, or both not in terms of acoustics but in terms of a speaker's level of arousal (Scherer, 1986). In sum, it is possible that perceptual mediation is not necessary to experience urgency and that a single monotonic function can be used to characterize both rate-based predictors.

In this study, we addressed the role of perceptual mediation and whether psychometric functions concerning different rate-based predictors of urgency are truly independent. This question was assessed by developing a continuum of pulse-train stimuli that varied only in rate (Miller and Taylor, 1948). At the low end of the continuum, pulses were easily resolved, with changes in rate corresponding to speed (or tempo). At the high end of the continuum, pulses fused together perceptually leading to a tonal sensation with a clear pitch level. To our knowledge, this is the first study of urgency incorporating phenomenological changes in speed and pitch along a single rate-based continuum.

2. Urgency scaling experiment

A magnitude estimation task was conducted to assess the experienced urgency of pulse trains that varied in rate between 3.13 and 200 Hz.

2.1 Participants

Thirty participants (20 women and 10 men) with normal hearing were asked to provide magnitude estimations. The mean age of participants was 21 years (SD=1.9 years). The procedures were approved by the Wilfrid Laurier University Research Ethics Board and all participants gave informed consent.

2.2 Stimuli and apparatus

Pulses (bursts of white noise) were interleaved with silent intervals (of equal duration) at varying rates of repetition. The rate of the resulting pulse train was thus determined by the inter-onset-interval of the noise burst. Rates were equally spaced on a log scale ranging from 3.13 to 200 Hz. Sounds were presented in a double-walled sound-attenuated chamber over headphones (Sony Stereo Headphones MDR-XD200) at a presentation level of 70 dB SPL.

2.3 Procedure

Each sound was presented for a duration of 2.5 s with the order of sounds being independently randomized for each participant. Participants judged the experienced urgency of each sound using a free-modulus magnitude estimation task (Stevens, 1975) in which participants are asked to assign a numerical value to each sound. The absence of a fixed scale (e.g., 1 to 7) improves sensitivity of measurement by encouraging participants to actively compare each response to all other responses. Participants' responses initiated the following trial and each session lasted approximately 15 min.

2.4 Results and discussion

Responses of each participant were first standardized (converted to z-scores) in order to avoid biases arising from outliers (e.g., a participant that used an unusually wide range of numerical ratings). Although the absolute values of the reported standardized scores are not meaningful, the relative values should reflect real differences experienced by the sample of participants.

Standardized ratings of urgency across the sound frequencies tested may be seen in Fig. 1. Urgency ratings were not equivalent across pulse rates, $F(12,348)=5.26$, $p<0.0001$. Trend analysis revealed that the linear trend was not significant, $F(1,29)=1.16$, n.s. and the quadratic trend was only marginally significant, $F(1,29)=4.39$, $p<0.05$. The best-fitting polynomial trend was cubic, $F(1,29)=47.18$, $p<0.0001$, indicating that there are two changes in the directionality of the relationship between urgency and pulse rate (for the range of pulse rates tested). The strong cubic trend is a highly unusual scaling result and suggests the possibility of two underlying power functions. The first, peaking at 17.68 Hz, likely corresponds to the upper limit for which sequential pulses may be resolved (i.e., infrapitch). For higher pulse rates (e.g., beyond 35.36 Hz), sequential pulses appear to fuse and give rise to a clear pitch percept (see also Fitzgerald and Wright, 2005; Miller and Taylor, 1948; Pollack, 1969). Intermediate pulse rates would have had a repetition rate that was too high to resolve independent pulses but too

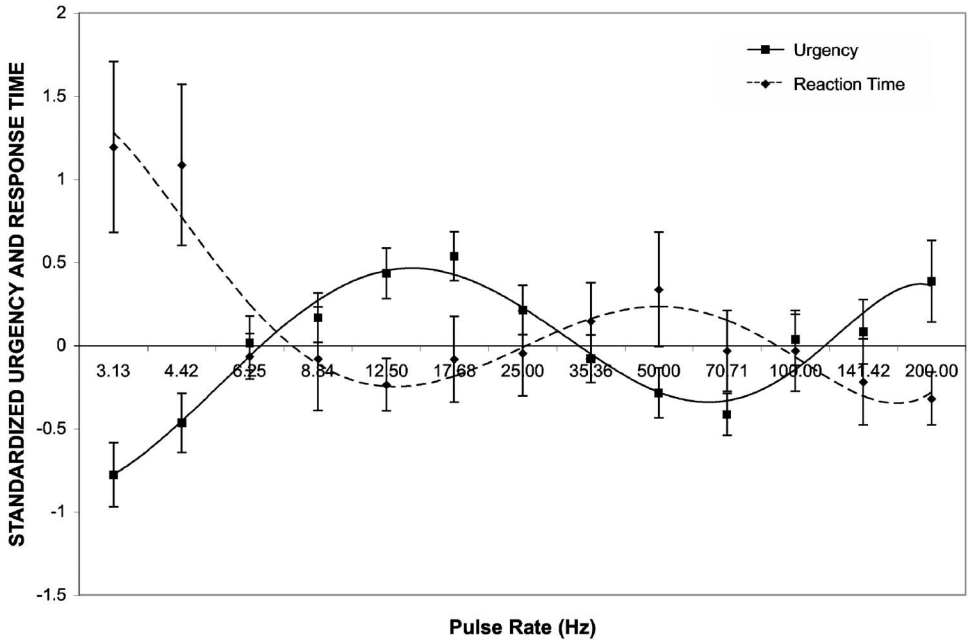


Fig. 1. Mean standardized urgency ratings and mean standardized response times for each pulse rate. Error bars represent standard error.

low to yield a clear pitch percept. These results strongly suggest that the experience of urgency in response to an auditory warning is mediated by perceptual constructs.

Following previous urgency research (Hellier *et al.* 1993; 1999), urgency ratings were mapped using an application of Stevens Power Law (Stevens, 1975), Eq. (1):

$$U = kf^m \tag{1}$$

where U is the judged urgency (in log units), f is the frequency or pulse rate (in log units), and k and m are the intercept and slope of the line (exponent) of best fit.

For the range of pulse rates from 3.13 to 17.68 Hz ("infrapitch region"), the power function exponent was 1.10. For the range of pulse rates between 35.36 and 200 Hz ("pitch region"), the exponent was 0.32. The relative and absolute value of these exponents are comparable to those reported by Hellier *et al.* (1999) for the warning parameters of speed and pitch.

3. Response time experiment

In order to verify whether the effects of the urgency mapping experiment reflected the experience of urgency, Experiment 2 asked participants to perform a tracking task (Makeig and Jolley, 1996) while occasionally responding to an auditory warning as quickly as possible. If the urgency ratings were based on experienced urgency, there should be an inverse pattern of response times for the same stimuli.

3.1 Participants

Nine participants (six women and three males) with normal hearing were asked to participate in the tracking experiment. The mean age of participants was 22.8 (SD=1.7 y). The procedures were approved by the Wilfrid Laurier University Research Ethics Board and all participants gave informed consent.

3.2 Stimuli and apparatus

The tracking task was presented on a PC computer with a standard video display and roller mouse using COMPTRACK software (Makeig and Jolley, 1995). The pulse-train stimuli and apparatus were identical to those described in Sec. 2.2.

3.3 Procedure

The tracking task required participants to use a computer mouse to keep a circular cursor within a target circle by compensating for simulated buffeting and gravitational forces that acted upon it. Participants were asked to simultaneously listen for the auditory warning and to respond to it as quickly as possible whenever it was heard by pressing a button on a button box (i.e., not the mouse button).

3.4 Results and discussion

Standardized response times may be observed in Fig. 1. All participants displayed a high degree of vigilance in the tracking task. An ANOVA confirmed that response times varied with regard to pulse rate, $F(12,96)=2.27$, $p<0.05$. Average response times to pulse trains were shortest for pulse rates of 17 Hz (385.36 msec) and 200 Hz (384.14 msec). By comparison, response times were much longer for the lowest pulse rates in the infrapitch region (3.13 Hz, 524.17 msec; 4.42 Hz, 546.04 msec) as well as for the lowest pulse rates in the pitch region (35.36 Hz, 435.46 msec; 50 Hz, 483.04 msec). It is important to note that these response time differences were obtained in a fully alert low-cognitive load condition. The response time differences would likely be further exaggerated under less optimal cognitive conditions. Nonetheless, the correlation between mean urgency ratings and mean response times was negative and highly significant, $r(11)=-0.84$, $p<0.0001$, suggesting that the polynomial function obtained for urgency reflected the *experience* of urgency, rather than a conscious integration of discrete perceptual mappings.

4. Pitch strength scaling experiment

To verify the phenomenological descriptions of pitch regions described in Sec. 2.4, pulse trains were subjected to a magnitude estimation task involving pitch strength; that is, the extent to which sounds give rise to a pitch percept. We expected perceived pitch strength to be at baseline (equal to pitch strength of white noise) between 17.68 and 35.36 Hz (transition region) and to rise sharply beyond 35.36 Hz (pitch region).

4.1 Participants

Twenty-seven participants (21 women and 6 men) with normal hearing were asked to provide magnitude estimations. The mean age of participants was 19.13 years (SD 1.5 years).

4.2 Stimuli and apparatus

The stimuli and apparatus were identical to that described in Sec. 2.2 with the exception that participants heard stimuli through Sennheiser HD 580 Precision headphones and the addition of a standard sound consisting of 2.5 s broadband white noise presented at 70 dB SPL.

4.3 Procedure

Participants judged pitch strength of each pulse train relative to the standard using a free-modulus magnitude estimation task. The following specific instructions were provided in order to clarify the objective of the task: "We have arbitrarily assigned the standard a pitch strength value of 10. You will be judging the pitch strength of the comparison sound in each presentation relative to the standard. If the comparison sound appears to be of the same pitch strength as the standard, then you would give a response of 10. If the comparison appears to have three times the pitch strength of the standard, then you would give a response of 30, and so on." In total, each participant judged all sounds four times each. Sessions lasted approximately 25 min.

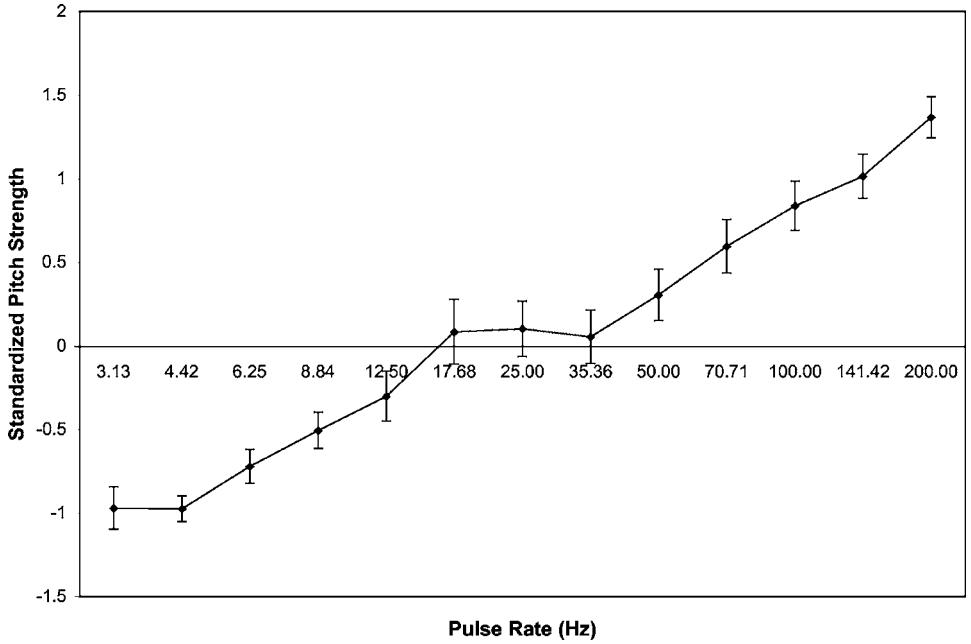


Fig. 2. Mean standardized pitch strength ratings for each pulse rate. Error bars represent standard error.

4.4 Results and discussion

The mean standardized ratings of pitch strength varied across pulse rates, $F(12,312)=33.68$, $p<0.0001$. Figure 2 shows that there was no change in pitch strength between 17.68 and 35.36 Hz (transition region). Pulse trains in this region had pitch strength that corresponded to that of an extended sample of white noise—i.e., pitch was ambiguous. There was a sharp rise beyond 35.36 Hz, which is suggestive of a fusing of pulses leading to a tonal sensation with an increasingly clear pitch level with increasing pulse rate. Examination of Fig. 2 also reveals a drop off in pitch strength below 17.68 Hz (infrapitch region). The drop-off is somewhat unexpected given that pulse trains in this region do not appear to fuse and may reflect metacognitive awareness of the transition from infrapitch to pitch with increasing pulse rate.

These scaling results are consistent with more objective tests of low-frequency pitch perception (Pressnitzer *et al.*, 2001; Russo *et al.*, 2007) and with pitch matching experiments involving pulse trains (Miller and Taylor, 1948). The latter study showed that the frequency of interruption of a white-noise pattern only corresponded to the matched pitch of a pure tone for frequencies above 40 Hz.[1]

5. Conclusions

The current study has identified two local maxima in a nonmonotonic (polynomial) function relating pulse rate to perceived urgency. The first maximum at 17.68 Hz corresponds to the positive correlation known to exist between speed and urgency (Guillaume *et al.*, 2003; Hellier *et al.*, 1993; 1995; 1999). Sounds in this region had very low pitch strength (i.e., infrapitch) consisting of a sequence of resolved pulses. The dip beyond 25 Hz likely reflects the transition from independent to fused pulses; that is, a transition from a sensation involving high-speed and no pitch to a sensation involving no speed and clear pitch. The second maximum, at 200 Hz reflects the positive correlation known to exist between pitch and urgency (Guillaume *et al.*, 2003; Hellier *et al.*, 1993; 1995; 1999). These findings suggest that the influence of rate-based cues on the experience of urgency in auditory warnings is mediated by perceptual constructs.

Acknowledgments

We wish to thank Farina Pinnock for research assistance. Research was supported by Discovery grants awarded to each author from the Natural Science and Research Council of Canada and a Research Fellowship from Wilfrid Laurier University awarded to the second author.

[1]We asked 23 participants with little to no music training do a pure-tone matching experiment with our pulse-train stimuli. Tuning accuracy improved as a function of pulse rate. Errors were consistently large for pulse rates below 35.34 Hz but fell to within about two semitones for pulse rates above 70 Hz.

Edworthy, J., Loxley, S., and Dennis, I. (**1991**). "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," Hum. Factors **33**, 205–231.
Fitzgerald, M. B., and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," J. Acoust. Soc. Am. **118**, 3794–3803.
Gibson, J. J. (**1979**). "The ecological approach to visual perception (Houghton Mifflin: Boston, MA).
Guillaume, A., Pellieux, K., Chastres, V., and Drake, C. (**2003**). "Judging the urgency of nonvocal auditory warning signals: Perceptual and cognitive processes," J. Exp. Psychol., Appl. **9**, 196–212.
Hellier, E. J., and Edworthy, J. (**1999**). "On using psychophysical techniques to achieve urgency mapping in auditory warnings," Appl. Ergon **30**, 167–171.
Hellier, E. J., Edworthy, J., and Dennis, I. (**1995**). "A comparison of different techniques for scaling perceived urgency," Ergonomics **38**, 659–670.
Hellier, E. J., Edworthy, J., and Dennis, I. (**1993**). "Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency," Hum. Factors **35**, 693–706.
Makeig, S., and Jolley, K. (**1995**). "COMPTRACK: A compensatory tracking task for monitoring alertness," In Technical Document No. 93-3C, Naval Health Research Center, San Diego, CA.
Miller, G. A., and Taylor, W. (**1948**). "The perception of repeated bursts of noise," J. Acoust. Soc. Am. **20**, 171–182.
Neuhoff, J. G. (**2004**). "Ecological psychoacoustics: Introduction and history," edited by J. G. Neuhoff, *Ecological Psychoacoustics* (Academic Press, New York), pp. 4–12.
Patterson, R. D. (**1990**). "Auditory warning sounds in the work environment," Philos. Trans. R. Soc. London, Ser. B **327**, 485–492.
Pollack, I. (**1969**). "Periodicity pitch for interrupted white noise–Fact or artifact?," J. Acoust. Soc. Am. **45**, 237–238.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
Russo, F. A., Cuddy, L. L., Galembo, A., and Thompson, W. F. (**2007**). "Sensitivity to tonality across the pitch range," Perception **36**, 781–790.
Scherer, K. R. (**1986**). "Vocal affect expression: A review and a model for future research," Psychol. Bull. **99**, 143–165.
Stevens, S. S. (**1975**). "Psychophysics: Introduction to its perceptual, neural, and social prospects," (Wiley, New York).
Stockhausen, Karlheinz. (**1990/1971**). "Stockhausen on Music: Public lectures in English from 1971 transcribed by Robin Maconie," (Marion Boyars, London, New York).

Chirp zeta transform beamforming for three-dimensional acoustic imaging

Maria Palmese and Andrea Trucco
Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, Via all'Opera Pia 11, I-16145 Genova (Italy)
palmese@dibe.unige.it, trucco@dibe.unige.it

Abstract: This letter considers the possibility of the generation of three-dimensional acoustic images with a limited computational load by considering the extension of Chirp zeta transform beam forming to the case of planar array and near-field conditions. This extension, with a few innovative solutions, allows for a dramatic reduction in the number of on-line operations over traditional time-domain beamforming.

PACS numbers: 43.60.Fg, 43.60.Lq [JC]
Date Received: July 23, 2007 **Date Accepted:** September 13, 2007

1. Introduction

Despite the introduction of commercial equipment, the building of systems able to generate a real-time three-dimensional (3D) acoustic video of an investigated environment[1,2] is still a challenge, mainly due to the cost of the needed planar array and the computational load of the associated signal processing. Wideband digital beamforming[1–4] is the most applied signal processing method, for both underwater and medical applications.

Although the cost of planar arrays can be reduced by the use of the sparse array concept,[1,2] some methods have been developed to implement the beamforming algorithm in the frequency domain[1,3] to mitigate the computational load. Zero-padded fast Fourier transform (FFT) beamforming[3] is the best known of these methods, being very efficient and producing exact results for narrowband signals. However, its extension to wideband signals is not straightforward and introduces some errors that reduce the method's appeal.

Although less recognized, Chirp zeta transform (CZT) beamforming[3] is a flexible frequency-domain method that can process wideband signals without any error, with a computational load equal to or lower than that of zero-padded FFT beamforming. However, its extension to planar arrays has not been attempted and its relative merits have not been assessed. Moreover, 3D imaging systems should be able to work in the near-field region, imposing a focused beamforming approach[1,2,4] and this causes additional difficulties, especially with frequency-domain methods.

In this letter, we present an extension of wideband CZT beamforming to planar arrays and near-field conditions, introducing some innovative ideas. The near-field difficulties have been solved by the adoption of the Fresnel approximation,[1,4,5] a useful definition of steering angles, and the setting of multiple focal regions. In addition, a technique to generate cubic resolution cells has also been introduced. The combination of such solutions can offer distinct computational advantages. Although the context assumed here refers to 3D acoustic imaging, mainly for underwater sonar systems, the proposed method can also be exploited in other contexts or environments.

2. CZT beamforming concept

Let us consider a linear array consisting of M sensors, with intersensor spacing d, and let us denote the steering angle with θ (measured from broadside) and the sound speed with c. If the time signals from the sensors are segmented into blocks of length K, then the discrete Fourier transform (DFT) coefficients, $B(k,\theta)$, of a corresponding segment of the beam signal steered in the direction θ, in the far-field region, are given by the following expression:

$$B(k,\theta) = \sum_{m=0}^{M-1} w_m S_m(k) \exp\left(-j2\pi f_k \left(\frac{md \sin\theta}{c}\right)\right) \tag{1}$$

where w_m is the apodizing weights applied to control the side-lobe level, $S_m(k)$ is the DFT coefficients of the signal $s_m(t)$ collected by the m-th sensor, and k is the frequency index, corresponding to the frequency $f_k = kf_s/K$ (f_s being the sampling frequency). The product md represents the position, x_m, of the m-th sensor along the array baseline.

According to Maranda,[3] if the qth beam of M_b beams is selected, the previous equation can be rewritten as follows:

$$B(k,\theta_q) = W^{-q^2/2} \sum_{m=0}^{M-1} w_m S_m(k) A^m W^{-m^2/2} W^{(q-m)^2/2} \tag{2}$$

where

$$A = \exp\left(-j2\pi f_k \frac{d}{c} \sin\theta_i\right) \tag{3a}$$

$$W = \exp\left(j2\pi f_k \frac{d}{c} \Delta s\right) \tag{3b}$$

$$\Delta s = \frac{\sin\theta_f - \sin\theta_i}{M_b - 1} \tag{3c}$$

where θ_i and θ_f are the initial and final steering angles, respectively, and the angle θ_q is one of a set of M_b predefined steering angles equispaced in the sine domain. This beam spacing perfectly agrees with the related angular resolution.[1]

It is worth noting that Eq. (2) expresses $B(k,\theta_q)$ as a discrete convolution of the sequences:

$$w_m S_m(k) A^m W^{-m^2/2}, \quad W^{m^2/2}, \quad m \in [0, M-1] \tag{4}$$

thereby allowing FFT methods for performing the so-called "fast convolution" to be used effectively. The computational advantage is duly discussed in Maranda.[3]

3. Extension to planar array and near-field

Let us now consider a planar array placed on the plane $z=0$, centered on the coordinate origin, composed of $M \times N$ sensors, with intersensor spacing d in both directions. The sensor identified by the indexes (m, n) is placed at position (x_m, y_n) and generates the signal $s_{m,n}(t)$. The steering direction is identified by two azimuth and elevation angles (θ_a, θ_e) that, different from conventional formalism,[2,5] are defined as shown in Fig. 1. The choice of such angles is useful to easily extend CZT beamforming to a planar array. Although conventional formalism also makes it possible to express the main computation as a discrete convolution it is definitely less direct and intuitive.

Further, if near-field conditions are considered, the plane wave assumption that has been exploited in Eq. (1), allowing the use of the simple phase term $x_m \sin\theta/c$, is no longer valid. To handle the curvature of the wave front, a focusing distance r_0 should be introduced in the phase term[1,4,5] which, in turn, becomes much more complex. To simplify such a phase term, the Fresnel approximation can be applied, obtaining the following term:

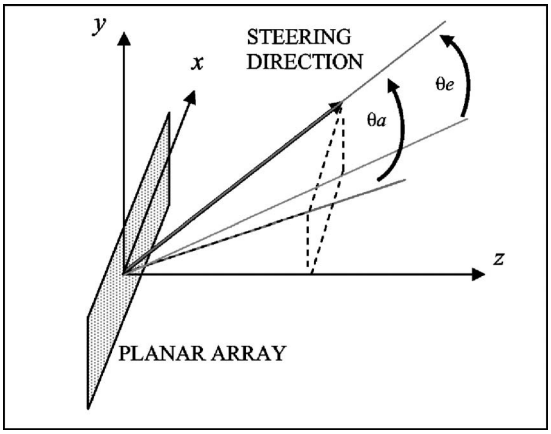


Fig. 1. Notation and geometry for a planar array.

$$\frac{x_m \sin\theta_a}{c} + \frac{y_n \sin\theta_e}{c} - \frac{x_m^2 + y_n^2}{2r_0 c} \tag{5}$$

The validity region of the Fresnel approximation (i.e., the region where the approximation errors are negligible) is duly discussed in Ziomek.[5] Here, we simply recall that, for a rectangular array centered on the coordinate origin, the focusing distance r_0 should satisfy the condition $r_0 > 0.678D$, D being the diagonal size of the array.

Altogether, if the Fresnel approximation is adopted, the computation of the beam signal in the (θ_a, θ_e) direction, focused at a distance r_0, requires the summing of the frequency bins of the received signals as in Eq. (1), replacing the phase term $md \sin\theta/c$ with that in Eq. (5). In this way it is possible to obtain the following equation:

$$B(k,\theta_a,\theta_e) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} w_{m,n} \exp\left(j2\pi f_k\left(\frac{x_m^2 + y_n^2}{2r_0 c}\right)\right) S_{m,n}(k) \times \exp\left(-j2\pi f_k\left(\frac{x_m \sin\theta_a}{c} + \frac{y_n \sin\theta_e}{c}\right)\right) \tag{6}$$

Let us now consider the computation of $M_b \times N_b$ beam signals identified by the indexes (p, q). For each of these the steering direction is $(\theta_{a,p}, \theta_{e,q})$, where $\theta_{a,p}$ and $\theta_{e,q}$ are two angles chosen from a set of M_b azimuth angles and a set of N_b elevation angles, respectively. The predefined angles are equispaced in the sine domain, ranging from an initial to a final angle. (Different initial and final angles are allowed for azimuth and elevation.) Paying attention to the beam signal (p, q), after some mathematics, the previous equation can be rewritten as follows:

$$B(k,\theta_{a,p},\theta_{e,q}) = W_a^{-p^2/2} W_e^{-q^2/2} \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} v_{m,n} S_{m,n}(k) A_a^m A_e^n W_a^{-m^2/2} W_e^{-n^2/2} W_a^{(p-m)^2/2} W_e^{(q-n)^2/2} \tag{7}$$

where W_a and A_a, as well as W_e and A_e, are defined as in Eq. (3), by adding the subscripts "$_a$" and "$_e$" respectively, where applicable. (Moreover, for W_e and A_e, N_b should replace M_b for the computation of Δs.) The term $v_{m,n}$ is defined as follows:

$$v_{m,n} = w_{m,n} \exp\left\{ j2\pi f_k \left[\left(m - \frac{M-1}{2} \right)^2 + \left(n - \frac{N-1}{2} \right)^2 \right] \frac{d^2}{2r_0 c} \right\} \tag{8}$$

Actually, moving from Eq. (6) to Eq. (7), a phase term, linear with f_k, which does not depend on the sensor indexes, has been neglected. This is reasonable as such a term simply introduces a known delay in the beam signal that can be easily fixed a posteriori.

In focused beamforming (independent of the time or frequency implementation), the depth of field (DOF)[1,6] is defined as the range interval around the focusing distance r_0 inside which the performance, evaluated in term of angular resolution and amplitude gain, only marginally degrades. The extension of the DOF depends on the specific value of r_0: the smaller the focusing distance r_0, the smaller the DOF extension.[6] Unfortunately, the range extension of the volume to be imaged typically exceeds that of the DOF. The problem is commonly solved by segmenting the received time signals into subsequent blocks that are processed using different focusing distances, which increase with time. In other words, the volume to be imaged is subdivided into multiple, adjacent focal regions. Each focal region is centered on a specific focusing distance and is contained inside the DOF corresponding to such a focusing distance. As in our method the time signals are already segmented into blocks of length K, so it is easy to apply different focusing distances to each block. However, in setting the value of K, it is essential that the spatial extension corresponding to the K time samples does not exceed the shortest of the extensions of the focal regions.

The general method and the specific solutions described above are used to extend the CZT beamforming to the planar array case and to allow correct focusing over an extended near-field volume. The accuracy of computing wideband beam signals by the proposed method is equal to that of the traditional delay-and-sum (D&S) beamforming,[1,4] except for the errors introduced by the Fresnel approximation. However, as discussed and quantified in Trucco,[7] such errors are really negligible inside the validity region[5] of the approximation. Therefore, the described method represents a computationally convenient way to compute beam signals that, inside the Fresnel validity region, negligibly differ from those computed by D&S beamforming.

Below, two original solutions devoted to further reducing the computational burden will be briefly introduced:

(1) It can generally be observed that it is computationally convenient to increase the length K of the signal blocks. This is in contrast to the need for short blocks close to the array, where the DOF extension is more limited. To overcome this problem the length of the signal blocks can vary, increasing in synchrony with the DOF extension.
(2) The angular resolution is fixed and so the lateral resolution worsens with the distance. Instead, the range resolution depends on the bandwidth and does not generally vary with the distance. However, the generation of a cubic resolution cell is often welcome,[1] and the worsening of the range resolution with the distance is perfectly acceptable. This makes it possible to save many operations by reducing the number of frequency bins considered (i.e., the bandwidth) with the distance. In other words, Eq. (7) can be computed for a number of indexes k that decreases block after block, according to the desired range resolution.

4. Results and conclusions

A comparison of the computational load of D&S beamforming[1,4] and that of CZT beamforming is carried out below. For each beamforming type, only operations that must be performed online are considered, thus excluding all the operations that can be performed off-line and whose results can be stored inside a memory. Let us analyze a sonar system with the following design parameters: a square array of $N \times N$ transducers (i.e., $M=N$), spaced $d=1.5$ mm from each other; a carrier frequency $f_c=500$ kHz, with a signal bandwidth of 250 kHz; a set of $N_b \times N_b$ beams (i.e., $M_b=N_b$); a volume of interest ranging from 1 to 100 m of distance; a sampling frequency $f_s=1.5$ MHz. The sound speed c is set at 1500 m/s. For the D&S beamforming, a

Table 1. On-line operations to create one 3D image for different choices of N and N_b.

	CZT beamforming	D&S beamforming
$N=64$, $N_b=100$	2.63×10^{10}	8.59×10^{12}
$N=100$, $N_b=100$	6.28×10^{10}	2.04×10^{13}
$N=100$, $N_b=128$	6.43×10^{10}	3.31×10^{13}

signal interpolation step has been considered,[4] performed by a FIR filter with 100 stages. For the CZT beamforming, the block length K ranges from 1024 to 4096, and the generation of cubic resolution cells is enabled.

Table 1 compares the number of on-line operations needed to create one image of the volume of interest. One can note that a computational advantage greater than two orders of magnitude is obtained for all the considered cases. As the proposed method applies the Fresnel approximation to the beamforming delays, the computed beam signals do not present any significant inaccuracy provided that they are pointed and focused inside the validity region[5] of such an approximation.

Acknowledgments

Special thanks to PARVIS Systems and Services, Milan, Italy, for its generous financial support of this work.

References and links

[1]V. Murino and A. Trucco, "Three-dimensional image generation and processing in underwater acoustic vision," Proc. IEEE **88**, 1903–1946 (2000).
[2]T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out* (Elsevier Academic, Amsterdam, 2004).
[3]B. Maranda, "Efficient digital beamforming in the frequency domain," J. Acoust. Soc. Am. **86**, 1813–1819 (1989).
[4]R. O. Nielsen, *Sonar Signal Processing* (Artech House, Boston, 1991).
[5]L. J. Ziomek, *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing* (CRC Press, Boca Raton, FL, 1995).
[6]R. C. Hansen, "Focal region characteristics of focused array antennas," IEEE Trans. Antennas Propag. **33**, 1328–1337 (1985).
[7]A. Trucco, "A least-squares approximation for the delays used in focused beamforming," J. Acoust. Soc. Am. **104**, 171–175 (1998).

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 and notices are 2 months prior to publication.*

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2006.

Independent Auditors' Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America as of 31 December 2006 and 2005 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

We conducted our audits in accordance with auditing standards generally accepted in the United States of America. Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes consideration of internal control over financial reporting as a basis for designing audit procedures that are appropriate in the circumstances, but not for the purpose of expressing an opinion on the effectiveness of the Society's internal control over financial reporting. Accordingly we express no such opinion. An audit includes examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation. We believe that our audits provide a reasonable basis for our opinion.

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of 31 December 2006 and 2005 and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

O'CONNOR DAVIES MUNNS & DOBBINS, LLP
May, 18, 2007
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
31 DECEMBER 2006 AND 2005

	2006	2005
Assets		
Cash and cash equivalents	$ 2,078,642	$1,178,921
Accounts receivable, net	314,044	412,950
Investments	8,098,936	7,351,777
Property and equipment, net	50,880	71,962
Other assets	361,294	403,516
	$10,903,796	$9,419,126
Liabilities and Net Assets		
Accounts payable and accrued expenses	$ 429,259	$ 332,808
Deferred revenue	1,442,037	1,513,580
Total liabilities	$ 1,871,296	$1,846,388
Net assets		
Unrestricted	$ 7,994,438	$6,578,903
Temporarily restricted	738,062	693,835
Permanently restricted	300,000	300,000
Total net assets	$ 9,032,500	$7,572,738
	$10,903,796	$9,419,126

 0001-4966/2007/122(5)/2487/5/$23.00

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2006 AND 2005

	2006	2005
Changes in Unrestricted Net Assets		
Revenue:		
Dues	$ 688,698	$ 664,866
Publishing—JASA	2,341,413	2,285,814
Standards	362,555	360,312
Spring Meeting	307,337	264,622
Fall Meeting	451,537	272,658
Other member services revenue	22,579	14,733
Other	181,269	121,436
Net assets released from restrictions	81,226	71,214
	$4,436,614	$4,055,655
Expenses:		
Publishing	$1,491,045	$1,805,982
Standards	386,729	388,934
Spring Meeting	359,404	300,916
Fall Meeting	297,535	302,801
Member Services	234,768	218,634
Other	440,327	438,754
Administration	531,828	493,632
	$3,741,636	$3,949,653
Net Income from Operations	$ 694,978	$ 106,002
Non-operating activities:		
Interest and dividends	$ 256,643	$ 177,702
Realized gain on investments	259,648	106,985
Unrealized gain on investments	204,266	151,862
	$ 720,557	$ 436,549
Increase in unrestricted net assets	$1,415,535	$ 542,551
Changes in Temporarily Restricted Net Assets		
Contributions	$ 15,426	$ 3,160
Interest and dividends	31,288	25,682
Realized gain	44,069	18,983
Unrealized gain	34,670	26,947
Net assets released from restrictions	(81,226)	(71,214)
Increase in temporarily restricted net assets	$ 44,227	$ 3,558
Change in net assets	$1,459,762	$ 546,109
Net assets at beginning of year	7,572,738	7,026,629
Net assets at end of year	$9,032,500	$7,572,738

Annual report of the Technical Committee on Physical Acoustics

(See September and October issues for additional reports)

It has been an honor for me to be your Chair and a pleasure to have the opportunity to work with a very talented, highly motivated, and friendly group of Physical Acoustics (PA) committee members (along with Acoustical Society of America (ASA) members across the board) in the long-range planning of ASA activities and special sessions for future meetings.

It has been a very productive and exciting year for the Physical Acoustics Technical Committee of the ASA as is evident from the success at both the Fourth Joint Meeting of the ASA with the Acoustical Society of Japan (ASJ) (held in Honolulu from 28 Nov.–2 Dec. 2006) and the 153rd Salt Lake City Meeting (4–8 June 2007).

The planning of the special sessions for the Joint ASA/ASJ Hawaii Meeting began during the tenure of our past Chair, Tom Matula (2003–2005). Tom received outstanding PA support with the following special sessions topics, and ASA organizer and co-chair (both in parenthesis): (a) Infrasound (Hank Bass, Milton Garces); (b) Sonic boom from supersonic aircraft and projectiles (Victor Sparrow); (c) Sound speeds, phonons, and the thermodynamics of condensed matter (Albert Migliori, Christian Pantea); (d) Cavitation (Claus-Dieter Ohl); (e) Sound propagation in inhomogeneous media - joint with Biomedical/Ultrasound/Bioresponse to Vibration (BB) (James G. Miller); (f) Celebration of Floyd Dunn—BB session joint with PA (Larry Crum, William D. O'Brien); and (g) Prediction and propagation of outdoor noise—Noise (NS) session joint with PA (D. Keith Wilson).

During the 151st Providence meeting, your current Chair met with Dr. Yo-iti Suzuki (Current President of ASJ and Chair of Technical Program Committee) and Dr. Hiroshi Sato (Current Secretary of ASJ Technical Program Committee) to develop a well thought out program involving all the technical committees. Subsequently, ASJ and ASA organizers collaborated to put together their special sessions. Here we mention that Dr. Jun-ichi Kushibiki (Tohoku University) was our ASJ PA representative.

Hank Bass volunteered to be the PA Technical Program Organizing Committee (TPOM) representative. Contributed papers were organized into two other sessions; (h) Thermoacoustics (co-chair Steven Garrett); and (i) Landmines and Ultrasonics (co-chair Roger Waxier). Thanks, Hank, for all your help.

The following sessions: (b) "Sonic boom," (d) "Cavitation," (e) "Inhomogeneous media," and (f) "Floyd Dunn," were held in two parts.

Physical acoustic leadership is present in the Society. Our President, Anthony Atchley (2005–2006), also served as ASA Technical Program Chair. Let's also mention that Giles Daigle (another former PATC Chair) is our ASA President Elect (2006–).

Larry Crum was invited to give a plenary lecture entitled "Therapeutic Ultrasound," at the opening ceremonies. Thanks, Larry, for an outstanding talk!

One of the highlights of the meeting was Hank Bass receiving the Silver Medal in Physical Acoustics. The citation reads "for leadership in physical acoustics and contributions to the understanding of atmospheric sound propagation." Congratulations Hank!

PA special sessions (a) "Infrasound" included continuous monitoring experiments, the eruption of the Tungurahua volcano, tornadoes and storms, high-altitude experiments, 3D modeling, the observation system in Isumi, Japan, and wind noise contributions; (b) "Sonic booms" covered propagation through turbulence, research in the Japanese Aerospace Exploration Agency, results from an entry vehicle, the supersonic biplane, ballistic range simulation, turbulence and atmospheric effects, discrete modeling, simulators, perception, human and structural response, effects of a Minuteman II missile, and spherically diverging N waves; (c) "Phonon/condensed matter" covered measuring elastic properties by inelastic neutron scattering, monocrystals and Blackman diagrams, propagation in ferromagnetics, thin films and ultrasonic spectroscopy, hydrogen absorbing crystals, elastic tensor of Osmium, capillary wave excitation, light and Brillouin scattering, hypersonic phononic and sonic crystals, and photoelastic Lamb wave visualization; (d) "Cavitation" included bubble fusion in deuterated acetone, ultrasonic frequency dependence in sonochemical reactions, comparing sonoluminescence and water vapor plasma spectra, nucleation and growth in cavitation bubbles, sonoluminescence in sulfuric acid, oscillation and bubble collapse, bubble shock interaction, plasma generation by E&M radiation, controlled cavitation, bubble and elastic particle interactions, and multibubble cavitation; (e) "Inhomogeneous media I" involved acoustic wave studies to characterize various elastic and porous material properties in cancellous, cortical, and trabecular bone using ultrasonic measurements, acoustic microscopy, and slow and fast wave studies; "Inhomogeneous media II" studied porous, granular, bubbly medium, sediments, nonlinear surface waves in soil, hysteretic nonlinearity of soil, and nonlinear acoustic landmine detection.

PA contributed sessions (h) "Thermoacoustics" topics included the amplification of acoustic intensity of pulse waves, spatial instabilities leading to spatial patterns, anharmonic resonators, loop-tube cooling systems, marginal instabilities, energy conversion studies, laser Doppler anemometry measurements of acoustic intensity in an arbitrary terminated pipe, and band structure in a multiside branched system; (i) "Landmines and Ultrasonics" covered an optimized seizmo-acoustic array, nonlinear cw and time-reversal landmine detection techniques, acousto-optic interactions from a surface wave perturbed by underwater sound, improvement of dispersion of nanometer diamond particulates by ultrasound, ultrasonic scattering by polycrystalline material, multiple sources for focusing ultrasonic waves, standing wave particle transport, and transfer function analysis for coherent and reverberant fields.

PA had planned four special sessions in Salt Lake City. Andi Petsculescu organized: (a) Acoustic probes of planetary environments, James Sabatier (who recently became the Chair of Education in Acoustics (ED)) planned (b) Physical acoustic demonstrations (joint with ED), Albert Migliori and Veerle M. Keppens co-organized (c) Ultrasound in condensed matter, neutrons, nanomaterials, magnetisms, and Steve Garrett took the Society to uncharted waters with his (d) Acoustical genealogy poster session (joint with ED and open to all).

Kent Gee and Richard Raspet (who is now our PA representative for standards) volunteered to be the PA TPOM representative. Contributed papers were organized into three other sessions; (e) Atmospheric acoustics (chair, Michael White); (f) Nonlinear and linear seismology and time reversal in solids (chair, Brian Anderson); and (g) Bubbles modes and scattering (chair, A Mark Wochner). Thanks, Kent and Richard, for all your help.

Congratulations to Edwin Carstensen, who received the 2007 Helmholtz–Rayleigh Interdisciplinary Silver Medal in Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics … "for contributions to the physics of biomedical ultrasound." Ed's work is inspiring.

PA special sessions (a) "Acoustic probes" topics included planetary exploration using acoustical instrumentation, microphone design for use on Mars' surface, propagation in planetary atmospheres (Venus, Earth, Mars, and Titan), and the Huygens probe on Titan and alien soundscapes; (b) "PA demonstrations" involved live demos showing Faraday waves (Holt), parametric array (Wilson), hysteresis in a soil-elastic plate oscillator (Korman Letcher), 2D resonant visualization using a bifringent solid (Gladden), ultrasonic signatures of human motion (Ekimov and Sabatier), angular dependence of target strength (Tucholski), a demo of human hearing using a cochlea analogue (Tomlinson), and a demo of a physical model of the active cochlea (Jovic); (c) "Ultrasound in condensed matter" topics included tribological properties of metal dichalcogenide nanostructures, monocrystal elastic constants, low-temperature magnetoacoustic measurements to study superconductive transitions in Sr_2RuO_2, neutron, x-ray, and nuclear resonance scattering applied to lattice dynamics in thermoelectric materials, charge ordering transition as probed by ultrasound, elasticity in metallic glasses, pulse-echo ultrasound in pulsed magnetic fields, thermal diffusivity in a superconductor using open-cell photoacoustic technique, third-order elastic constants of langasite single crystals, and diffuse waves in solids and spectral density predictions; (d) "Academic genealogy" covered the family trees of Rudnick–Putterman (Garrett), Brian E. Anderson (BEA), Harvard acoustics back to Helmholtz (Blackstock), Katherine Safford Harris (Bell-Berti), F. V. Hunt and Harvard (Wright), Jiri Tichy (Sommerfeldt and Leishman), R. T. Beyer (Korman and Letcher), Indiana University history (Yost), P. Marston (Matula), R. Bruce Lindsay (P. H Rogers), Apfel and Yale (Ketterling), National Research Council of Canada (Daigle and Stinson), and National Center for Physics and Acoustics (Bass and Sabatier).

PA contributed sessions (e) Atmospheric acoustics (chair, Michael White) covered surface waves in the nocturnal boundary layer, propagation at a forest edge, localization using an array of aerostats, refraction effects from wind and turbulent stability scale, noise reduction by windscreens, reverberation in an urban environment, time domain boundary conditions, impulse propagation in a boundary layer, maximum likelihood estimates, infrasound, and coherence through turbulence; (f) Nonlinear and linear seismology and time reversal in solids (chair, Brian Anderson) topics included nonlinear oscillations of a buried object, nonlinear cw and time reversal methods in landmine detection, seismic pulses from an explosive source,

acoustic/seismic transfer function, long term linear and nonlinear acoustic field tests of soils and source imaging using time-reversal experiments; (g) Bubbles, modes, and scattering (chair, Mark Wochner) covered sonoluminescence from transient cavitation, coupled radial and translational motion of a bubble, bubble shape oscillations, liquid capillary jet waveguide, Scholte–Stonely wave generation, nonlinear shear waves, leaky Rayleigh waves, thin plate and shell experiments involving zero-group velocity Lamb modes, mode shape predictions, Wigner space-time spectral energy density of a beam pattern at a liquid-solid interface, and multiple scattering of cylinders in a porous media; (h) Thermoacoustics (chair, Richard Raspet) topics included a high-frequency cooler, annular high-frequency prime mover, particle image velocimetry inside a stack-heat exchanger couple, synchronization of small thermoacoustic oscillators, acoustic conversion of heat to sound, computational methods, pressurized high-frequency acoustic heat engine, and miniature prime mover operating at 10 kHz.

All said, the Hawaii and Salt Lake City Meetings were a big success for PA.

For the past year Anthony Atchley has been our President. He has done an outstanding job for the ASA, and much success has come out of his hard work. He has also shown tremendous leadership, enthusiasm, and guidance in matters relating to the PATC. I am grateful and very fortunate to have had Anthony's guidance during my term as Chair. Although Anthony's term as President ended after Salt Lake City, he continues to serve as a member of the Executive Council joining Wayne Wright who also serves on the Executive Council.

Gilles Daigle has been elected our new President and Mark Hamilton has been elected our new President Elect. Congratulations and good luck to our new officers, Gilles and Mark! Victor Sparrow (who served on the Executive Council with Wayne) becomes our new Vice-President Elect and gains a seat on the Technical Council. Congratulations and good luck to Vic! PA representation, along with our Editor-in-Chief, Allan Pierce, appears to be very strong.

It is time to acknowledge members of PA who volunteer their services on various committees:

Tom Matula currently serves on the Medals and Awards Committee. Tom officially took over in Honolulu. Robert Keolian and James Sabatier, both predecessors of this position, have worked hard to do a great job for the Society and remain active in helping on this Committee. Steve Garrett serves on the Membership Committee; Phil Marston serves on the Books Committee, and Richard Raspet serves on the ASA Committee on Standards, replacing Sameer Madanshetty, who has done an excellent job. Todd Haye is our outstanding Student Council representative.

Other items: David Blackstock is extremely active in the Student Council and the program involving "Take a Student to Lunch." Keith Wilson is our editor of JASA-EL. Ron Roy who received the Eastman Fellowship returns from his 1 year sabbatical in Oxford.

Congratulations to Ken Gilbert, Andrea Prosperetti, Alexandra Tolstoy, Yves Berthelot, and Ron Roy, for 25 years of service to ASA. Supercongratulations to Gideon Maidanik and Richard Stern for 50 years of service.

Special thanks to Albert Migliori for bringing to the Society the innovative research that involves new approaches connecting ultrasonics and condensed matter physics.

The 2006 Physical Acoustics Summer School, PASS 2006, took place from 18–25 June 2006, at Sunrise Springs, La Cienega, New Mexico. The program is jointly run by NCPA (Hank Bass) and Perm State (Anthony Atchley), with strong ASA ties. The next meeting will take place in 2008.

I would like to take this time out to thank Larry Crum for his ongoing service, leadership, and quality of his research that he brings to both PA and BB. He is a role model and his enthusiasm is contagious.

The 154th Meeting of the ASA will take place in New Orleans 27 Nov.–1 Dec. 2007. There are two special sessions (see the most recent call for papers).

The 155th Meeting is a joint meeting called Acoustics'08 Paris which will be held in Paris 29 June–4 July 2008. The ASA, the European Acoustics Association (EAA), and Société Française d'Acoustique (SFA) are organizing a joint meeting integrating ecua: the 9th European Conference on Underwater Acoustics & euronoise: the 7th EUROpean Conference on NOISE Control, along with a celebration of the 60th anniversary of the SFA.

Your Chair has met with Manell Zakharia (SFA), (the Chair of Acoustics '08 in Paris), Technical Program Organizing Committee Co-chair Philippe Blanc-Benon (SFA), and communicated with my Physical Acoustics European special session co-organizer Walter Lauriks (EAA), along with other ASA organizers to help coordinate the 15 structured sessions that ASA members in Physical Acoustics have organized with their fellow European co-chairs.

Thank you, PATC members, for all your help.

MURRAY S. KORMAN
Chair

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.

2008

21–23 Feb	2008 National Hearing Conservation Assn. Conference, Portland OR [NHCA, Tel: 303-224-9022; Fax: 303-770-1614; Web: www.hearingconservation.org).
29 June–4 July	Acoustics08, Joint Meeting of the Acoustical Society of America (ASA), European Acoustics Association (EAA), and the Acoustical Society of France (SFA), Paris, France (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; Web: http://asa.aip.org/meetings.html).
2–6 July	International Clarinet Association Clarinetfest®2008, Univ. of Missouri—Kansas City, MO (Dr. John Cipolla, Department of Music, Western Kentucky University 1906 College Heights Blvd. #41029, Bowling Green, KY 42101-1029, Tel.: 270-745-7093; E-mail: john.cipolla@wku.edu).
27–30 July	NOISE-CON 2008, Dearborn, MI (Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: 515-294-6142; Fax: 515-294-3528; E-mail: ibo@inceusa.org).
21–25 July	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. Email icben2008@seglobal).

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 Meetings Calendar—National

At the 154th ASA Meeting in New Orleans, LA, at the Sheraton New Orleans Hotel, 27 November to 1 December 2007, the ASA Committee on Standards (ASACOS) and ASACOS Steering Committees will meet as below:

• ***Tuesday, 27 November 2007***

ASACOS Steering Committee

• ***Wednesday, 28 November 2007***

7:30 a.m. – ASA Committee on Standards (ASACOS) Meeting of the Committee that directs the Standards Program of the Acoustical Society.

9:15 a.m. – Standards Plenary Group (includes U.S. TAGs to ISO/TC 43, TC 43/SC 1, ISO/TC 108 and its subcommittees, and TC 29)

10:45 a.m. – Accredited Standards Committee S12, Noise

1:30 p.m. – Accredited Standards Committee S1, Acoustics

2:30 p.m. – Accredited Standards Committee, S3, Bioacoustics

4:00 p.m. – Accredited Standards Committee, S2, Mechanical Vibration and Shock

Standards Meetings Calendar—International

• ***19–23 May 2008***

Meeting of International Electrotechnical Commission **IEC/TC 29, Electroacoustics,** will be held in Stockholm, Sweden.

• ***26–30 May 2008***

Meetings of International Organization for Standardization **ISO/TC 43, Acoustics,** and **ISO/TC 43/SC 1, Noise,** will be held in Borås, Sweden.

• ***27–31 May 2008***

Meeting of **ISO/TC 108/SC 5, Condition Monitoring and Diagnostics of Machines**, will be held in Kyoto, Japan.

Recent International Standards Meetings

Meetings of **ISO/TC 108, TC/108 SC 2, TC 108/SC 3, and TC 108/SC 6**, joined by their working groups, were held in Berlin, Germany, 23–27 April 2007. Draft reports from some of these meetings are available upon request.

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 St. Julians, Malta during the week of 22–26 October 2007. 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 parentheses are for information only.

AMCA (Air Movement and Control Association)

NEW STANDARDS

BSR/AMCA 300-200x, Reverberant Room Method for Sound Testing of Fans (new standard)

Applies to fans of all types and sizes and is limited to the determination of airborne sound emission for the specified setups. (3 September 2007)

ATIS (Alliance for Telecommunications Industry Solutions)

REVISIONS

BSR ATIS 0600315-200x, Voltage Levels for DC Powered Equipment Used in the Telecommunications Environment [revision and redesignation of ANSI T1.315-2001 (R2006)]

Establishes requirements, and test procedures for voltage ranges and characteristics (i.e., transients and noise) associated with the input voltage of network communications equipment powered from dc power systems in the telecommunications environment. (3 September 2007)

SCTE (Society of Cable Telecommunications Engineers)

REVISIONS

BSR/SCTE 17-200x, Test Procedure for Carrier to Noise (C/N, CCN, CIN, CTN) (revision of ANSI/SCTE 17-2001)

Defines the measurement procedure for determining the ratio of carrier-to-thermal noise and "noise-like" interference for broadband telecommunications system components. The test involves measuring the noise levels, or the combined noise plus "noise-like" intermodulation product levels, relative to the carrier level of a CW signal. The noise contribution of the test equipment is also measured to allow for correction of readings near the test equipment noise floor.

BSR/SCTE 62-200x, Measurement Procedure (revision of ANSI/SCTE 62-2002)

Defines a method of measurement for Noise Figure of active Cable Telecommunications equipment. It is intended for measurement of 75-ohm devices having type "F" or 5/8-24 KS connectors, and for the measurement of true broadband noise as opposed to narrow-band disturbances. (27 August 2007)

Projects Withdrawn from Consideration

API (American Petroleum Institute)

BSR/API RP-531M-xx-200x, Measurement of Noise from Fired Process Heaters (new standard)

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.

ASA (ASC S12) (Acoustical Society of America)

BSR S12.60-200x/Part 2, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools - Part 2: Relocatable Classroom Factors (new standard)

Focuses on the challenges and unique circumstances surrounding the use and relocation of new relocatable classrooms and provides methods to quantify acoustical performance and site considerations.

Project Need: To amend ANSI S12.60-2002, which did not fully consider movement and reuse of relocatable classrooms, which have several unique characteristics that do not lend themselves to the existing acoustical guidelines.

Stakeholders: School administrators; local, state, and possibly federal agencies tasked with school building design and funding.

BSR/ASA S12.64-200x, Underwater Noise Measurement of Ships (new standard)

Specifies methods for making underwater measurements of noise radiated from manned or unmanned surface ships, using commercial-off-the-shelf technology for methodology, instrumentation, and data processing. It provides a grade matrix for measurements of varying degrees of precision. It does not establish noise limits for any purpose nor does it address diagnostics, noise reduction, environmental consequences of noise, effects of noise on marine life, or passive monitoring.

Project Need: To provide a commercial standard for performing underwater noise measurements of ships.

Stakeholders: Military and nonmilitary ship owners and designers, researchers, government agencies.

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.

ARI (Air-Conditioning and Refrigeration Institute)

NEW STANDARDS

ANSI/ARI 370-2001, Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment (new standard)

ASA (ASC S2) (Acoustical Society of America)

REAFFIRMATIONS

ANSI S2.21-1998 (R2007), Method for Preparation of a Standard Material for Dynamic Mechanical Measurements (reaffirmation of ANSI S2.21-1998 [R2002])

ANSI S2.22-1998 (R2007), Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials (reaffirmation of ANSI S2.22-1998 [R2002])

ANSI S2.23-1998 (R2007), Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials (reaffirmation of ANSI S2.23-1998 [R2002])

ANSI S2.27-2002 (R2007), Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery (reaffirmation of ANSI S2.27-2002)

IEEE (Institute of Electrical and Electronics Engineers)

NEW STANDARDS

ANSI/IEEE C57.127-2007, Guide for the Detection and Location of Acoustic Emissions from Partial Discharges in Oil-Immersed Power Transformers and Reactors (new standard)

Information Concerning

ASHRAE Standards

ANSI/ASHRAE Standard 87.2-2002, Method of Testing Fan Vibration – Blade Vibrations and Critical Speed, completed withdrawal public review on 3 May 2007.

As a result of the comments received during the public review, ASHRAE Standards Committee has approved the formation of a revision project committee and that the document be revised instead of withdrawn.

Standards News from Abroad

Proposed Foreign Government Regulations

Call for Comment

U.S. manufacturers, exporters, regulatory agencies and standards developing organizations may be interested in proposed foreign technical regulations issued by Member countries of the World Trade Organization (WTO). In accordance with the WTO Agreement on Technical Barriers to Trade (TBT Agreement), Members are required to report proposed technical regulations that may significantly affect trade to the WTO Secretariat in Geneva, Switzerland. In turn, the Secretariat disseminates the information to all WTO Members. The purpose of this requirement is to provide global trading partners with an opportunity to review and comment on the regulations before they become final.

The National Center for Standards and Certification Information (NCSCI) at the National Institute of Standards and Technology (NIST) distributes these proposed foreign technical regulations to U.S. stakeholders via an online service, Notify U.S. is an e-mail and Web service that allows interested U.S. parties to register, obtain notifications, and read full texts of regulations from countries and for industry sectors of interest to them. To register for Notify U.S., please go to Internet URL: http://www.nist.gov/notifyus/ and click on "Subscribe." NCSCI is the WTO TBT Inquiry Point for the U.S. and receives all notifications and full texts of regulations to disseminate to U.S. Industry. For further information, please contact: NCSCI, NIST, 100 Bureau Drive, Gaithersburg, MD 20899-2160; Telephone: (301) 975-4040; Fax: (301) 926-1559; E-mail: ncsci@nist.gov or notifyus@nist.gov.

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 389-6:2007, Acoustics – Reference zero for the calibration of audiometric equipment – Part 6: Reference threshold of hearing for test signals of short duration

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 13374-2:2007, Condition monitoring and diagnostics of machines – Data processing, communication and presentation – Part 2: Data processing

ISO 15230:2007, Mechanical vibration and shock – Coupling forces at the man-machine interface for hand-transmitted vibration

ISO 19499:2007, Mechanical vibration – Balancing – Guidance on the use and application of balancing standards

ISO 22096:2007, Condition monitoring and diagnostics of machines – Acoustic emission

EQUIPMENT FOR FIRE PROTECTION AND FIRE FIGHTING (TC 21)

ISO 7240-19:2007, Fire detection and alarm systems – Part 19: Design, installation, commissioning and service of sound systems for emergency purposes

IEC Standards

ROTATING MACHINERY (TC 2)

IEC 60034-9 Ed. 4.1 b:2007, Rotating electrical machines – Part 9: Noise limits

ELECTRICAL EQUIPMENT IN MEDICAL PRACTICE (TC 62)

IEC 60601-2-37 Ed. 2.0 b:2007, Medical electrical equipment – Part 2-37: Particular requirements for the basic safety and essential performance of ultrasonic medical diagnostic and monitoring equipment

ULTRASONICS (TC 87)

IEC 61157 Ed. 2.0 en:2007, Standard means for the reporting of the acoustic output of medical diagnostic ultrasonic equipment

IEC 61689 Ed. 2.0 en:2007, Ultrasonics – Physiotherapy systems – Field specifications and methods of measurement in the frequency range 0.5 to 5 MHz

AUDIO, VIDEO AND MULTIMEDIA SYSTEMS AND EQUIPMENT (TC 100)

IEC 60728-4 Ed. 3.0 en:2007, Cable networks for television signals, sound signals and interactive services-Part 4: Passive wideband equipment for coaxial cable networks

IEC 60728-5 Ed. 2.0 en:2007, Cable networks for television signals, sound signals and interactive services-Part 5: Headend equipment

ISO Draft Standards

ACOUSTICS (TC 43)

ISO/DIS 3747, Acoustics-Determination of sound power levels of noise sources using sound pressure-Comparison method in situ (17 November 2007)

ISO/DIS 17201-5, Acoustics-Noise from shooting ranges-Part 5: Noise management (27 October 2007)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 2041, Mechanical vibration, shock and condition monitoring-Vocabulary (8 November 2007)

GLASS IN BUILDING (TC 160)

ISO/DIS 16940, Glass in building-Glazing and airborne sound insulation-Measurement of the mechanical impedance of laminated glass (7 October 2007)

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
TAG TC 29	**IEC/CDV 60318-1 (29/633/CDV)** Electroacoustics–Simulators of human head and ear–Part 1: Ear simulator for the calibration of supra-aural and circumaural earphones (21 December 2007)

TAG ISO/TC 43	**ISO/DIS 389-9** Acoustics–Reference zero for the calibration of audiometric equipment–Part 9: Preferred test conditions for the determination of reference hearing threshold levels (17 October 2007)
TAG ISO/TC 43/SC 1	**ISO/DIS 17201-5** Acoustics–Noise from shooting ranges–Part 5: Noise management (26 December 2007)
TAG ISO/TC 43/SC 1	**Second ISO/CD 10844** Acoustics–Specification of test tracks for the purpose of measuring noise emitted by road vehicles and their tyres (Revision of ISO 10844:1994) (2 November 2007)
TAG ISO/TC 108	**Systematic Review ISO 16587:2004** Mechanical vibration and shock–Performance parameters for condition monitoring of structures (17 December 2007)
TAG ISO/TC 108	**ISO/DIS 2041** Mechanical vibration, shock and condition monitoring–Vocabulary (7 January 2008)
TAG ISO/TC 108/SC 2	**Systematic Review ISO 14839-2:2004** Mechanical vibration- Vibration of rotating machinery equipped with active magnetic bearings–Part 2: Evaluation of vibration (17 December 2007)
TAG ISO/TC 108/SC 4	**ISO/CD 28802 N536** Ergonomics of the physical environment–
and **ISO TC 159/SC 5**	The assessment of environments by means of an environmental survey involving physical measurements of the environment and subjective responses of people (12 October 2007)
TAG ISO/TC 108/SC 6	**Systematic Review ISO 10813-1:2004** Vibration generating machines–Guidance for selection–Part 1: Equipment for environmental testing (17 December 2007)
TAG ISO/TC 108/SC 6	**Systematic Review ISO 5344:2004** Electrodynamic vibration generating systems–Performance characteristics (17 December 2007)

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.

High-Frequency Seafloor Acoustics

Darrell R. Jackson and Michael D. Richardson

Springer, 2007. 616 pp. $139.00 (hardcover). ISBN 978-0-387-34154-5

High-Frequency Seafloor Acoustics, by Darrell Jackson and Michael Richardson, is part of the Underwater Acoustic Series sponsored by the Office of Naval Research, and was written for sonar engineers and researchers working in underwater acoustics. It is intended to be a compendium of the recent accomplishments in the field as well as a perspective on continuing research. These goals are accomplished admirably through precise text and clear and relevant figures and graphs. The book is a comprehensive guide to all aspects of seafloor acoustics and a valuable resource to any underwater acoustician or acoustical oceanographer whose work includes interaction with the ocean bottom. The authors state that the frequency range covered is "very roughly" 10 kHz to 1 MHz, and hence consideration of geological structure on scales larger than a few meters is excluded. This restriction coincides with the underwater acoustic community's post-cold war focus on the acoustics of littoral regions.

The book assumes a natural division into two parts. The first half (Chapters 1–7) is on sediment characterization and the second half (Chapters 8–16) covers theories and models describing acoustic interaction with the sediment. While Chapters 1 and 2 are introductory, Chapters 3 through 7 discuss environmental characterization including physical and geoacoustic properties. Specifically, Chapter 3 covers the nature of marine sediments including classification, structure, bubbles, ripples and surface roughness structure from hydrodynamic and biological processes. Chapter 4 details physical properties including grain size distribution, grain morphology, sediment conductance and permeability, grain properties such as density and bulk modulus and interstitial fluid properties. Geoacoustic properties are discussed in Chapter 5. These are properties that are of primary interest to acousticians and of less interest to geologists and geotechnical engineers. These properties are compressional and shear wave sound speed and attenuation as a function of sediment properties and frequency, Biot "slow wave" properties and properties of rock. Seafloor roughness is considered in Chapter 6. The standard method of characterizing rough interfaces is introduced as well as common measurement methods. Lastly, Chapter 7 introduces sediment heterogeneity including measurement techniques, statistical characterization and the relationship between sediment heterogeneity and sediment bulk properties. A section detailing the bias due to the finite resolution of volume measurements is also included in this chapter.

For each physical measurement discussed, the origin of the measured phenomenon is reviewed in detail, standard metrics are introduced, multiple measurement techniques are presented and the results are given in the framework of the relevance to geoacoustic behavior. Consider for example Chapter 6 on interface roughness. The origin of interface ripples and isotropic roughness (including wave generation and bioturbation) is first discussed. Next, the power law method of characterizing the seafloor roughness is presented. Four methods of measurement including stereo-photography, laser line scanning, manual tracing and electrical conductivity probing are then discussed in detail. Lastly, the interface roughness is related to other physical properties and open research questions related to seafloor roughness are explored. The authors also discuss the limitations of different measurement procedures in reference to the prediction of acoustic behavior. This general approach is repeated throughout many of the chapters.

One unique contribution of this monograph is the unified presentation in Chapter 5 of empirical regressions between sediment physical properties (porosity, bulk density, mean grain size) and geoacoustic properties (compressional and shear sound speed and attenuation) representing hundreds of laboratory core and in-situ measurements from more than 150 sites around the world. This is one of the largest collections of such data available in one place in the open literature.

The second part of the monograph details wave propagation theories and the resulting predictive models of acoustic interaction with sediments. Chapters 8 through 10 cover fluid theory, elastic theory, and poro-elastic theory, as well as some other theories that do not fit into these three main categories, such as Buckingham's theory. In these chapters, each theory and model is presented in detail and experimental verification is given where available. Spherical wave effects are detailed and discussed. Common approximations such as Gassman's equations, the effective density fluid model and Wood's equation are derived. Limitations of the models and approximations are discussed. Higher order theories that accommodate squirt flow, intraparticulate porosity and mobile solid material are mentioned and referenced but not described in detail.

Chapters 11 through 16 cover the acoustic observables of reflection, interface and volume scattering and backscatter. Specifically, Chapter 11 covers reflection including comparison of both in-situ and laboratory measurements with theory and issues involved with reflection coefficient measurement. Chapter 12 details seafloor scattering measurements. Both monostatic and bistatic experiments are discussed. Results of the angle and frequency dependence of scattering are shown. Chapter 13 discusses four roughness scattering models: the small-roughness perturbation approximation, the Kirchhoff approximation, the small slope approximation and the composite-roughness approximation. Scattering by discrete features is also included. A detailed section is provided to introduce the reader to common properties of roughness scattering models. Experimental tests of the models are given when available and the mathematical accuracy of the approximations is explored. Chapter 14 covers sediment volume scattering from both a fluid and elastic model framework. Some seabed heterogeneity measurements are presented as well as a comparison between backscattering strength predictions and measurements. Chapter 15 discusses acoustic penetration of the seafloor. Specifically, sub-critical penetration is explored using the perturbation theory of seafloor roughness scattering. Backscatter statistics are detailed in Chapter 16. Gaussian statistics are discussed as well as non-Gaussian statistics such as the Rayleigh, lognormal, Weibull, K and Rayleigh mixture distributions. In addition Crowther's and McDaneil's backscattering fluctuation models are presented. In these chapters, the authors detail the limitations of both the measurements and models for each acoustic observable and identify open research areas.

Finally, the monograph has 13 appendices that occupy 113 pages. Details, derivations and supporting material that would unnecessarily burden the general conversational tone of the text are included here, providing additional degrees of rigor and completeness. Examples are the calculation of seawater properties, conversion between different types of roughness and volume spectra, discussions of transducers, complex signals, and causality, and thorough derivations of some of the scattering models.

In conclusion, this work is an excellent reference for those researching any area that is associated with high frequency sediment acoustics. Each chapter is self-contained so that readers can refer to the topic in which they are interested. The chapters are well referenced especially with respect to overview papers. Therefore, the monograph is a great starting point into a much larger literature. For those already in the field, the author's explanation of sediment property metrics will hopefully help standardize the report-

ing of future measurements and the explanation of measurement techniques will aid theoreticians in understanding the limitations of environmental characterization. Lastly, at the close of each chapter, the authors summarize the state of research in the area and suggest future advances. Although the community should not rely solely on these suggestions for future work, numerous gaps in knowledge that beg further exploration are clearly identified. The reader is left hoping that the world's various funding agencies are as keen to tackle these unsolved problems as Jackson and Richardson have been in describing our current knowledge of high frequency sediment acoustics.

MARCIA J. ISAKSON
Applied Research Laboratories
The University of Texas at Austin
P.O. Box 8029, Austin TX 78713-8029

PRESTON S. WILSON
Mechanical Engineering Department,
The University of Texas at Austin
P.O. Box 8029, Austin TX 78713-8029

Ultrasonic Nondestructive Evaluation Systems

Lester W. Schmerr, Jr. and Sung-Jin Song

Springer, New York, 2007. 550 pp. price: $149 (hardcover). ISBN-10: 0387490612, ISBN-13: 978-0387490618

Assessing the mechanical integrity of engineering materials is crucial to the safe and cost-effective operation of any mechanical system. Over the past 40 years, ultrasonic nondestructive evaluation has been established as the prime inspection modality for a number of industrial applications and is now a mature discipline taught at both undergraduate and graduate levels.

A number of textbooks have been written on different aspects of the subject, ranging from the fundamental physics of elastic wave propagation and scattering, such as the contributions by Graff,[1] Auld,[2] and Achenbach,[3] to more practical textbooks which focus on engineering applications like Krautkramer and Krautkramer,[4] Kino,[5] and Rose.[6]

This book is a welcome addition which distinguishes itself from the available literature because it is dedicated to the modeling of all the components of an ultrasonic measurement system. Although these models are already available, they are commonly described in self-contained frameworks. Instead, this book is the first effort to present them under a unified structure which shows how they combine together to generate the signal received from a flaw. The publication of this work is timely since the NDE community recognizes that future progress in the field will be brought by a more systematic use of numerical models in conjunction with advanced hardware technology.

The book is divided into 13 chapters, each of them followed by a series of excellent questions which aid the learning process. The book is very well presented and makes clever use of seven appendices which complement its main body by supplying background material without affecting the smooth presentation of the key topics. The last appendix provides useful MATLAB routines which are used in the questions at the end of each chapter and which invite students to have a *hands-on* approach to learning.

The book begins with an overview of ultrasonic system modeling and a description of a typical ultrasonic system, which is divided into five main parts: the pulser, cabling, transmitting transducer, specimen, and receiver. The pulser, cabling, and transmitter are treated separately in Chapters 2, 3, and 4, respectively. The working principles of each component are clearly explained and modeled in the form of equivalent electrical circuits. Chapter 5 examines the propagation of ultrasonic waves through the sample from the transmitter to the receiver and provides an equivalent electrical circuit model for both the propagation and reception of the mechanical signal.

The electromechanical properties of a transducer, which appear in the electrical models introduced in Chapters 4 and 5, are difficult to characterize in practice. Therefore, Chapter 6 describes a new pulse-echo method for measuring the transfer matrix of an ultrasonic transducer and introduces the concept of calibration. Chapter 7 combines the models developed for each component into a complete model for the measurement system.

The next three chapters are devoted to the analysis of ultrasound radiation and its interaction with flaws. Chapter 8 provides a clear overview of the theory of radiation of sound in both fluids and solids. The authors apply a number of theoretical tools originally developed in optics, such as the angular spectrum method, to ultrasonic waves and present many of the results from optics with a terminology accessible to readers unfamiliar with electromagnetic wave theory. In this context, Chapter 9 is a complete description of Gaussian beam theory. The chapter begins with an introduction to conventional Gaussian beams, which are used in optics to model a laser beam. Subsequently, the theory is generalized to the more complex scenarios encountered in NDE applications. Therefore, after introducing the paraxial wave equation in solids, the propagation and transmission/reflection phenomena are simulated and a multi-Gaussain beam model is introduced to represent the propagation of a beam through multiple curved interfaces. Chapter 10 is devoted to the analysis of wave scattering by a flaw and is aimed at defining simplified models to be used in the complete model of the measurement system. The general scattering problem is stated in terms of scattering amplitude, which is first introduced for scalar waves and then generalized to the case of elastic waves. The Kirchhoff approximation is introduced and applied to canonical scatterers such as spherical cavities, side drilled holes, and elliptical cracks.

Chapter 11 presents several ultrasonic measurement models which are derived from the reciprocity principle and are expressed in terms of characteristic parameters of the flaw to be inspected, such as its far field scattering amplitude.

The last two chapters conclude the book by applying the principles discussed in the previous chapters to build complete models of ultrasonic measurement systems which are then compared with experimental results. To this end, Chapter 12 introduces a number of MATLAB routines which simulate different inspection scenarios. Subsequently, in Chapter 13 these are used to simulate calibration measurements and model assisted flaw identification and sizing.

The contents and high quality of the book make it especially valuable for graduate students and educators. Beside its immediate relevance to modelers, the book is also appropriate for experimentalists, who can gain a better insight into the interactions between the different components of an ultrasonic system.

I have only a minor criticism. NDE is rapidly moving toward more extensive use of array technology; I believe that discussing how the principles illustrated in this book apply to arrays of sensors would have been a valuable addition to the content of the book. Overall, I have enjoyed reading this book, which clearly is a testimony to the many years of experience of the authors, and will become a reference textbook in NDE.

FRANCESCO SIMONETTI
Department of Mechanical Engineering
Imperial College London
London SW7 2AZ, United Kingdom

[1]K. F. Graff, Wave Motion in Elastic Solids (Dover, New York, 1991).
[2]B. A. Auld, Acoustic Fields and Waves in Solids (Krieger, Malabar, 1990).
[3]J. D. Achenbach, Wave Propagation in Elastic Solids (Elsevier Science, Amsterdam, 1984).
[4]J. Krautkramer and H. Krautkramer, Ultrasonic Testing of Materials (Springer, Berlin, 1990).
[5]G. S. Kino, "Acoustic waves-Devices," Imaging and Analog Signal Processing (Prentice-Hall, Englewood Cliffs, NJ, 1987).
[6]J. L. Rose, Ultrasonic Waves in Solid Media (Cambridge University Press, Cambridge, UK, 1999).

FORUM

Forum is intended for communications that raise acoustical concerns, express acoustical viewpoints, or stimulate acoustical research and applications without necessarily including new findings. Publication will occur on a selective basis when such communications have particular relevance, importance, or interest to the acoustical community or the Society. Submit such items to an appropriate associate editor or to the Editor-in-Chief, labeled FORUM. Condensation or other editorial changes may be requested of the author.

Opinions expressed are those of the individual authors and are not necessarily endorsed by the Acoustical Society of America.

Automobile entertainment responsive to interior environment

David Lloyd (ben Yaacov Yehuda) Klepper[a)]
Student, Yeshivat Beit Orot, Shmuel ben Adiya 1, Mt. of Olives, Jerusalem 97400, Israel

(Received 24 May 2007; revised 5 August 2007; accepted 7 August 2007)
[DOI: 10.1121/1.2782798]

I. INTRODUCTION

I was standing at a bus stop ("autobus stop" in Israel) at a recent evening, and a car approached without headlights shining. I pointed my finger at the car and yelled "Or" (light in Hebrew), and turning, I saw the driver responding positively. I thought "Isn't highway safety *everyone's responsibility!"* Imagine a driver who is an active shares/stock investor driving on a multilane highway and about to pass a large trailer-truck/lorry rig. The noise level is increasing, and the desire to be sure he hears the latest quotes prompts him to attend to the stereo radio volume control, as the car noise level increases due to the noise of the vehicle ahead. But he does not see that ahead of the vehicle he is about to attempt to pass, there is another, similar but slower moving, and his eyes dart to the radio volume control just as the truck/lorry driver belatedly switches on his signal indicator. Or imagine a similar traffic situation with a symphony orchestra violinist intently listening to a cassette tape of a work conducted by the very guest conductor who will be in charge of the rehearsal which is his destination. Sound systems that vary their volume levels to match changes in noise have been with us since 1962, and modern technology makes such equipment far simpler, economical, and sophisticated for unobtrusive operation.

II. BACKGROUND

The first commercially successful car radios were introduced around 1930.[1] The author assumes that they were a step forward in highway safety, because he can recall long solo automobile trips in the U.S.A., where drowsing at the wheel was actively prevented by interesting radio music and speech programs. This good far outweighed any safety problem of momentary distraction to change stations or volume or tone. The reader who drives (and most do so) can evaluate this hypothesis by his/her own experiences. Noise levels in automobiles can vary with changes in speed, engine power because of grades, road surface conditions, opening or closing of windows, tunnels, bridges, walls, and passing of other vehicles, particularly trucks and motorcycles. My subjective impression from 42 years of driving automobiles is that such changes in noise can distract a driver from observing the road ahead as the driver adjusts volume, and each reader can judge if this judgment corresponds to his/her experience.

Automatic level or "gain" control of audio amplification systems was introduced simultaneously by several companies, using vacuum tube technology, and some car radio manufacturers quickly introduced it to prevent overloading and loudspeaker destruction, with wide changes in input program level.[1–3] But a sound system automatically changing its level to suit changes in the environment, in the first case simply noise level, waited until California's Altec Lansing Co. introduced the "NOALA" for "Noise Operated Automatic Level Adjustment" in 1961 and first installed at Dulles Airport near Washington DC, opened in 1962. The Altec product was refined and made far more sophisticated over the years, but the first model controlled levels in steps by relays, and was intended for insertion between the power amplifiers and the loudspeakers line, with the unused power being wasted in level-dropping resistors. Resistor bypass relays were controlled by preamplified signals from sensing microphones, but test installations were tried with the loudspeakers used as the sensing microphones during periods of zero program and thus zero change of NOALA level during program.[4] Widespread use of background noise in systems using the NOALA product meant microphone sensing was the preferred option. At Dulles, a 30 W power amplifier was used as a signal distribution amplifier, and the relay contacts and resisters inserted between its output and the 200 W power amplifiers driving the distributed loudspeakers. All-electronic analogue equipment by several manufacturers quickly supplanted relays, and the art progressed to digital equipment, where such control is usually combined with additional signal control appropriate to the installation. Nearly all modern airports, large modern railroad stations, and bus terminals use this technology today, and the equipment has been steadily improved with respect to reduced costs and space, high reliability, and smoothness in operation. Digital technology is often employed, today, in a majority of new systems.

[a)]Electronic mail: daveklepper@yahoo.com

 0001-4966/2007/122(5)/2499/2/$23.00

The sophistication of adjustment of fixed theater, concert hall, and auditorium sound systems allowed such systems to approach a high level of perfection by 1970, but the operators of the million dollar sound systems that toured with rock stars lacked the time before shows for sophisticated adjustments. However, the demands for perfection were great, and a decade later, John Meyer and his colleagues at Meyer Sound Laboratories (again, California) developed "Source Independent Measurement." The electronic input to the amplification portion of the system (power amplifiers, loudspeakers) is compared with the signals at microphones in the auditorium, and including the effects of the auditorium, presenting the sound system operator with pictures of the adjustments desirable while the show was in progress. Soon the technology was licensed to competing sound companies, and a high degree of automation was included. Also, for fixed installations with widely varying requirements, seating patterns, stage setups, etc., this type of adjustment capability has proved a blessing.[5] Like the NOALA concept discussed above, this "SIM" equipment has also been improved over the years, with the major steps being its licensing by Meyer Sound Laboratories to competing firms, transition to digital technology, and application to fixed (installed) systems.

III. RECOMMENDATIONS FOR TODAY

The equipment developed by Meyer Sound Laboratories was mainly intended for adjustment of frequency response, not overall volume, but can very easily do both. Advances in digital technology have brought down its cost and size and can easily be incorporated into automobile radios. Alternative technologies exist that sense only during quiet, but the quiet time can be extremely short. The interior sensing microphones can be very small and inconspicuous, and one may be sufficient in most automobiles. Almost all signal processing today for such systems is by digital microprocessors.

The widespread implementation of this concept should be one step in bringing private automobile safety closer to that available in bus, tram, metro, and railroad public transportation. However, possibly an on-off switch might be required for those who would rather tolerate some noise masking instead of level changes not in the score for some types of music.

[1]R. D. Read, "From batteries to the space age," Sound Commun. **46**(2), 56–60 (2000).
[2]F. W. Roberts and R. C. Curtis, "Audio automatic volume control systems," J. Audio Eng. Soc., **1**(4), 310–316 (1953).
[3]E. L. Torick and R. Allen, "A VOX system for operation at high and variable ambient levels," J. Audio Eng. Soc. **15**(3), 303–307 (1967).
[4]Altec Lansing Catalogues of 1962 and after. These are best available from Electro-Voice Voice/Mark IV.
[5]J. Meyer, "The right prescription," Sound and Video Contractor, **18**(8), 60–67 (2000).

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.

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
GEOFFREY EDELMANN, *Naval Research Laboratory, Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,221,621

43.30.Vh SYSTEM AND METHOD FOR IDENTIFICATION AND QUANTIFICATION OF SONAR TARGETS IN A LIQUID MEDIUM

Mark Robert Patterson *et al.*, assignors to College of William & Mary
22 May 2007 (Class 367/88); filed 6 April 2005

A method to identify targets within a sidescan sonar image is described. It begins with several image processing steps before feeding an area of interest into a radial basis function neural network.—GFE

7,233,545

43.30.Wi SYSTEM AND METHOD FOR DETERMINING THE LOCATION OF AN ACOUSTIC EVENT

Edward P. Harvey, Jr. and Jack McGinn, assignors to McGinn-Harvey Holdings, LLC
19 June 2007 (Class 367/127); filed 7 September 2005

This preposterous document begins by declaring that sonobuoys can only detect targets on the same plane (e.g., ships, not submarines) and proceeds to patent the most basic and manifest concept of triangulation via time delay.—GFE

7,228,236

43.30.Yj SUBARRAY MATCHING BEAMFORMER APPARATUS AND METHOD

Thomas J. Barnard and Manuel Francisco Fernandez, assignors to Lockheed Martin Corporation
5 June 2007 (Class 702/14); filed 8 January 2001

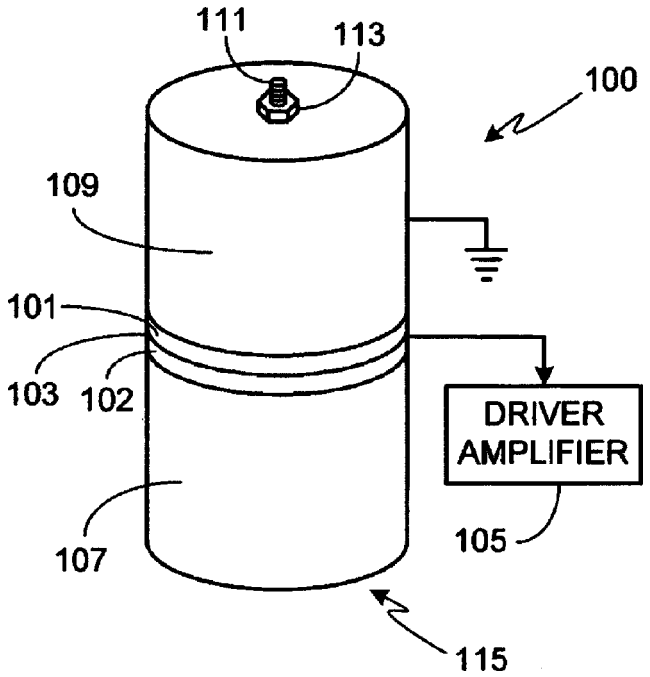
Array element location is often a difficult parameter to estimate, especially for towed arrays. A technique is proposed that utilizes a short portion of the array, whose position is presumed known (closest to the ship), to estimate weights for the rest of the array.—GFE

7,224,103

43.35.HI ACOUSTIC DRIVER ASSEMBLY WITH RECESSED HEAD MASS CONTACT SURFACE

Ross Alan Tessien and David G. Beck, assignors to Impulse Devices, Incorporated
29 May 2007 (Class 310/325); filed 5 May 2005

A design for an acoustic driver assembly is put forth; specifically, one



with the intent of producing sonoluminescence.—GFE

 0001-4966/2007/122(5)/2501/16/$23.00

7,224,809

43.35.Yb METHOD FOR THE ACOUSTIC LOCALIZATION OF PERSONS IN AN AREA OF DETECTION

Juergen Hoetzel, assignor to Robert Bosch GmbH
29 May 2007 (Class 381/92); filed in Germany 20 July 2000

The intent of this patent is to detect the presence, absence, and type of passenger in a car in order to safely deploy air bags. The method proposed is active ultrasonic and passive human speech detection.—GFE

7,230,509

43.38.Ar ACOUSTIC MIRROR

Ralph Stoemmer, assignor to Epcos AG
12 June 2007 (Class 333/133); filed in Germany 11 December 2001

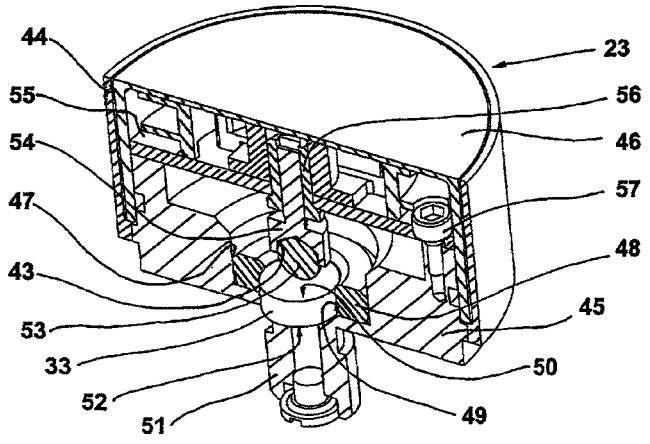
An acoustic mirror composed of alternating layers of differing impedance is proposed for use in a bulk acoustic wave resonator.—GFE

7,066,028

43.38.Fx ACOUSTIC SENSOR FOR MONITORING MACHINING PROCESSES IN MACHINING TOOLS

Valerio Dondi, assignor to Marposs Societa' per Azioni
27 June 2006 (Class 73/649); filed in Italy 28 February 2002

The subject at issue here is an acoustic pickup mounted near a grinding wheel, or similar machining operation, for the purpose of monitoring the machining operation. According to the patent, prior technology would typically involve a pickup located in the motor housing near the grinding wheel shaft. Such a pickup is said to be too sensitive to motor noises and not able



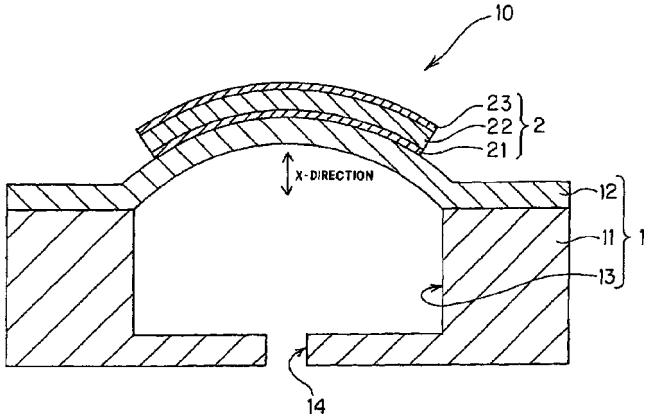
to pick up the acoustic signal from the grinding action. The proposed solution is a piezoelectric (PZE) element **50** located within the grinding wheel itself, in a small housing **23** positioned near the center of rotation of the wheel. An electronic circuit on circuit board **55** amplifies the PZE signal immediately, reducing noise pickup from electrical interference or other sources. The amplified signal is transmitted magnetically through plastic faceplate **46** to a second external amplifier circuit.—DLR

7,221,075

43.38.Fx PIEZOELECTRIC/ELECTROSTRICTIVE DEVICE

Tsutomu Nanataki *et al.*, assignors to NGK Insulators, Limited
22 May 2007 (Class 310/324); filed in Japan 19 November 2004

The authors disclose an interesting device which appears to be an ultrasonic transducer. It is formed by depositing piezoelectric or electrostrictive material layer **2** on top of ceramic hemispherical boss **12**. This curved unimorph then radiates efficiently in what would appear to be the 300 kHz



frequency range, judging by the dimensions. Back cavity **13** and support **11** also play a role in establishing the resonant behavior. The authors go into detail on optimizing the sensitivity vs the piezoelectric layer thickness, dome height, and cavity depths.—JAH

7,221,077

43.38.Fx ULTRASONIC TRANSDUCER AND MANUFACTURING METHOD THEREOF

Yukihiko Sawada, assignor to Olympus Corporation
22 May 2007 (Class 310/369); filed in Japan 1 April 2003

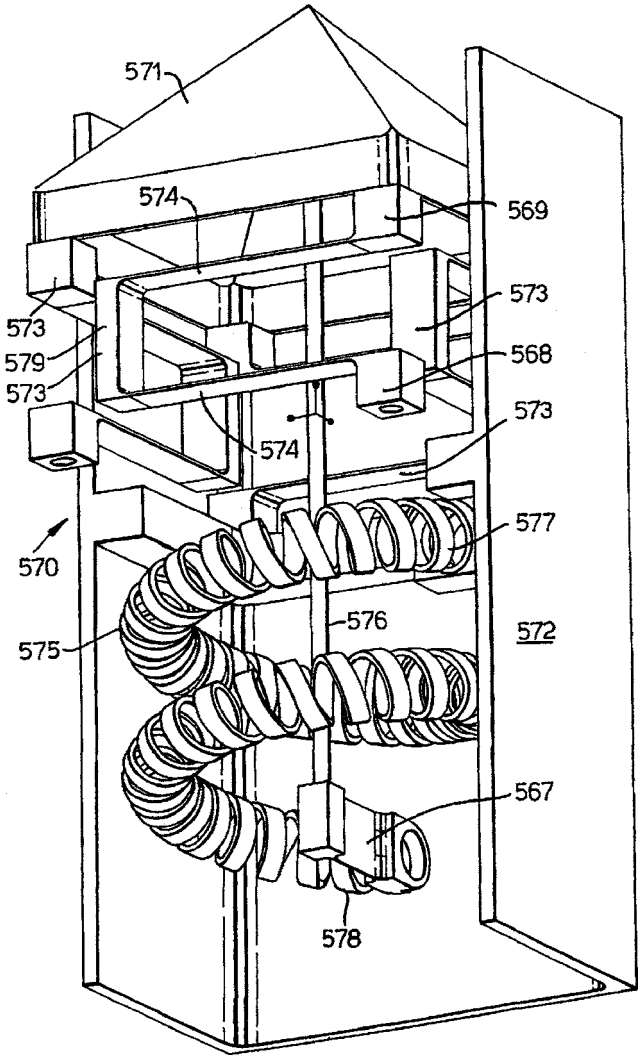
This patent discloses a method for making a cylindrical ultrasonic transducer out of piezoceramic elements cut into wedges and retained on a backing layer. There is some novelty to the way that the electrical connections get made, but it looks very difficult to fabricate. Little else is of interest in this patent since the basic geometry and fabrication concepts were used in sonar about 40 years ago.—JAH

7,224,813

43.38.Fx LOUDSPEAKER USING AN ELECTRO-ACTIVE DEVICE

Anthony Hooley *et al.*, assignors to 1... Limited
29 May 2007 (Class 381/190); filed in United Kingdom 21 December 1999

This patent sets forth an interesting concept: The use of a piezoelectric helical spring as a loudspeaker motor. Although the diagram appears to be some kind of Jack in the Box toy, it serves to illustrate the principals involved if one simply ignores everything in the top half of the picture. As explained in the patent, if a long, bending-type piezo wafer is constructed in the form of a coil spring, then the spring will tend to coil or uncoil with the application of a positive or negative voltage. Carried one step further, such a spring **575** is formed into a larger helix with one end **577** fixed and the



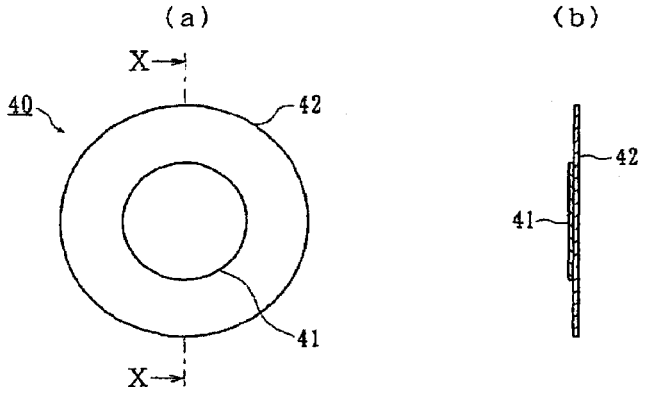
other end **578** attached to rod **576**. In this configuration, the tendency of the spring to coil and uncoil results in a net upward or downward displacement which can be used to drive a conventional loudspeaker cone. Moreover, if the piezo strip is divided into discrete elements, it can be driven directly by a digital audio signal. The patent makes no mention of electrical impedance or conversion efficiency, nor is there any indication that a working model has actually been constructed.—GLA

7,227,966

43.38.Fx PIEZOELECTRIC SPEAKER

Fujihiko Kobayashi, Fuji-shi, Shizuoka-ken, Japan
5 June 2007 (Class 381/190); filed 10 September 2002

A common form of piezoelectric loudspeaker consists of a piezoelectric ceramic disk **41** laminated to a larger ''vibration plate'' **42**. This patent teaches that improved wide-band performance can be realized by incorporating radial slits in the vibration plate, or dividing it into a number of



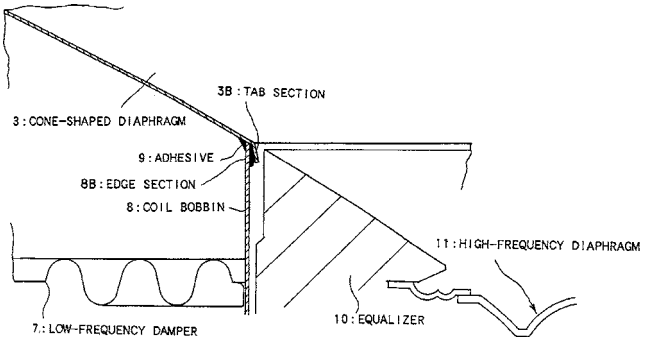
independent parts ''of any configuration,'' or mounting it off-center, or making a composite plate of different materials, or adding a second stiffening plate, etc., etc. The possibilities are limitless.—GLA

7,190,804

43.38.Ja SPEAKER APPARATUS

Tohru Nakamura and Tomohiro Suenaga, assignors to Pioneer Corporation
13 March 2007 (Class 381/423); filed in Japan 27 June 2003

''Cone-shaped diaphragm'' **3** is attached to voice ''coil bobbin'' **8** using ''tab section'' **3B** instead of attaching somewhere below the top of the bobbin former. A ''high-frequency diaphragm'' **11**, sometimes called a



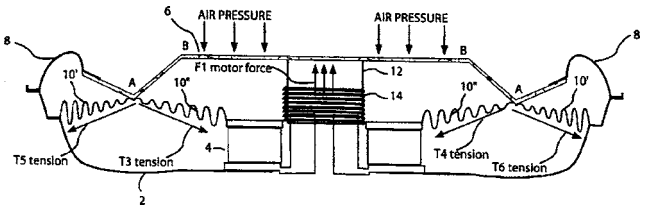
whizzer cone, is fitted to the interior of the bobbin using a means of attachment **10**, here called an ''equalizer'' (why, is not fully clear to this reviewer), which is said to offer advantages compared to applications where the whizzer cone suspension is said to be below the top of the bobbin.—NAS

7,197,154

43.38.Ja METHOD AND AUDIO SPEAKER WITH MINIMIZATION OF WOBBLE OF THE VOICE COIL

Joseph Y. Sahyoun, Redwood City, California
27 March 2007 (Class 381/404); filed 17 May 2004

This is another of a number of patents relating to subwoofers. In the present patent, an improvement in the spider assembly **10′**, **10″** is described. The patent also describes some variations on the shape of the inner perimeter of surround **8** and how it attaches to diaphragm **6** and the molds to form



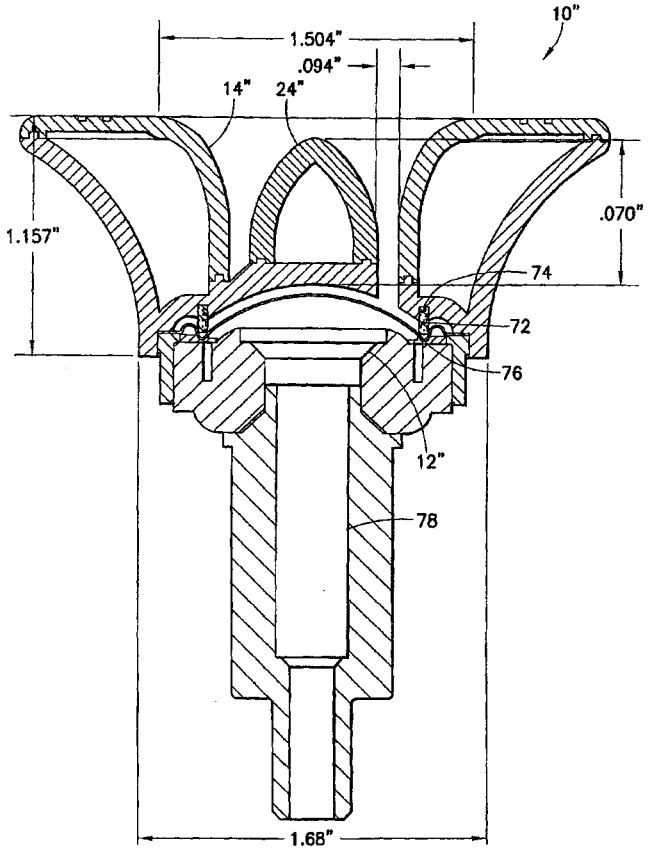
same. These are said to better center the assembly and also reduce the voice coil wobble, especially at the extremes of the cone's motion.—NAS

7,203,329

43.38.Ja AUDIO SPEAKER SYSTEM EMPLOYING AN AXI-SYMMETRICAL HORN WITH WIDE DISPERSION ANGLE CHARACTERISTICS OVER AN EXTENDED FREQUENCY RANGE

Eric J. Alexander *et al.*, assignors to SoundTube Entertainment, Incorporated
10 April 2007 (Class 381/340); filed 11 February 2004

A horn **14**″ is affixed to a dome tweeter **12**″ with annular foam gasket **72** in opposing grooves **74**, **76**. The assembly **10**″ is part of a two way



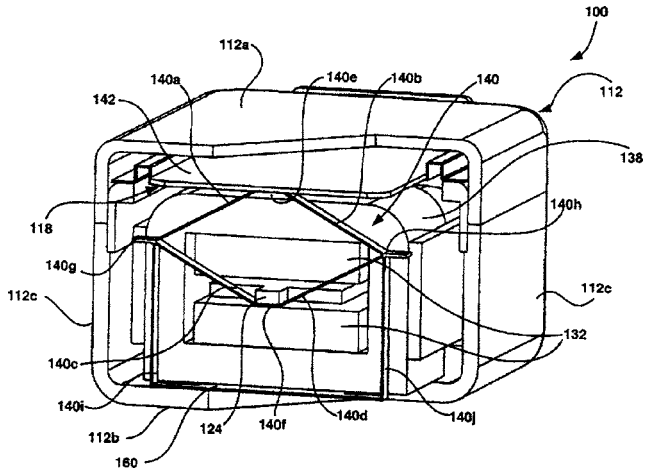
speaker system—hollow stem **78** fits into the cone type low-frequency driver. Wide dispersion over a broad frequency range is claimed.—NAS

7,203,334

43.38.Ja APPARATUS FOR CREATING ACOUSTIC ENERGY IN A BALANCED RECEIVER ASSEMBLY AND MANUFACTURING METHOD THEREOF

David Earl Schafer and Mekell Jiles, assignors to Knowles Electronics, LLC
10 April 2007 (Class 381/418); filed 21 November 2003

"A paddle **142** of a diaphragm **118** of a receiver **100** is manufactured using one or more layers of a material selected for their inertial mass and rigidity. The paddle may have a layered structure with stiff outer layers such as aluminum and a less dense inner layer, such as thermoplastic adhesive.



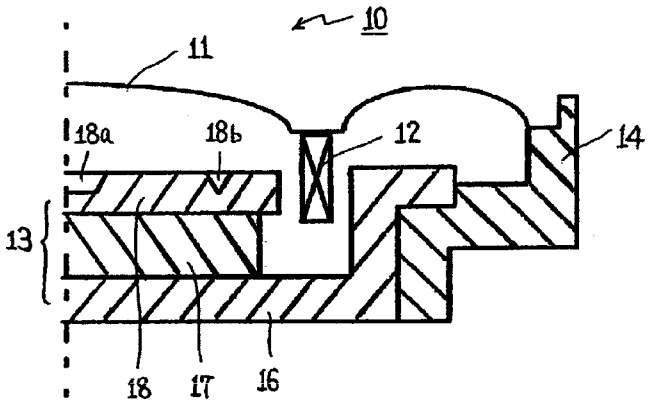
The inner and outer layers are selected to give an inertial mass matching that of an armature **124** of the receiver **100** and to give lowest frequency bending resonances above a desired range, for example, 16 kHz."—NAS

7,224,817

43.38.Ja LOUDSPEAKER

Kazuyuki Kosuda, assignor to ONKYO Corporation
29 May 2007 (Class 381/412); filed in Japan 16 May 2002

This is a puzzling patent. The figure shows a partial cross-section of a miniature edge-driven loudspeaker having a usable frequency range of roughly 500–5000 Hz. Diaphragm **11** is driven by voice coil **12**, with a sealed magnetic structure **13** at the rear. According to the patent, sound waves from the rear of the diaphragm reflect from top plate **18**, resulting in response peaks and dips. The solution is to add irregularities **18a**, **18b** to



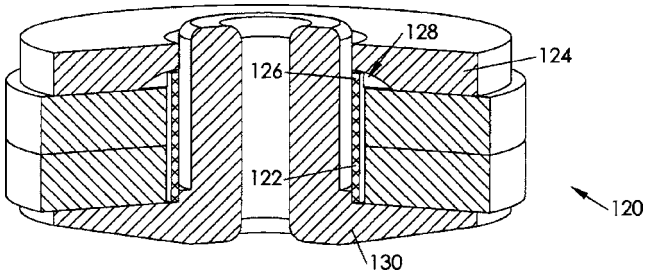
"diffuse" the reflections and smooth out response. Well, such grooves and pits would enlarge the air cavity slightly but, considering the dimensions involved, any acoustic diffusion must occur well above the frequency range of the loudspeaker. The patent document includes "before" and "after" response curves which, contrary to the text, seem to indicate that the effects are indeed inconsequential. Onkyo's motives for patenting this configuration remain inscrutable.—GLA

7,227,970

43.38.Ja SHORTING RING FIXTURE FOR ELECTROMAGNETIC TRANSDUCER

Richard C. Calderwood and Enrique M. Stiles, assignors to STEP Technologies Incorporated
5 June 2007 (Class 381/412); filed 26 February 2004

A high-quality moving-coil loudspeaker may include a shorting ring in its magnetic gap to flatten impedance and reduce distortion. In this variant,



shorting ring **126** also serves as a jig to establish concentric alignment of the magnetic circuit components.—GLA

7,224,640

43.38.Pf UNDERWATER ACOUSTIC BEARING SENSOR

J. Clay Shipps and Hikmet Andic, assignors to Wilcoxon Research, Incorporated
29 May 2007 (Class 367/13); filed 24 August 2005

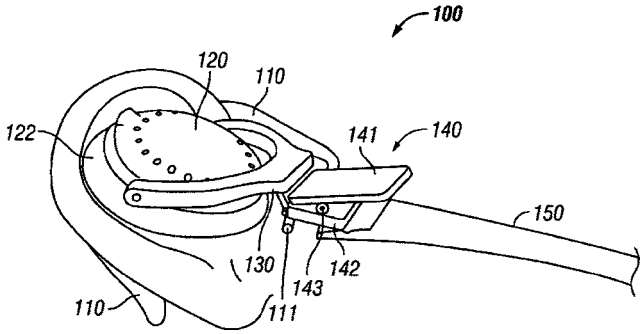
In order to overcome the ambiguity in absolute position associated with acoustic vector sensors, the patent proposes a combination that includes heading, pitch, and roll sensors.—GFE

7,221,771

43.38.Si OVER-THE-EAR HEADSET

Fred DeKalb and Ric Baughman, assignors to Plantronics, Incorporated
22 May 2007 (Class 381/381); filed 27 February 2003

Over-the-ear headsets are small, lightweight, and easy to use. However, human ears come in a wide variety of shapes and sizes, making it difficult to ensure good acoustic coupling between the receiver and the ear



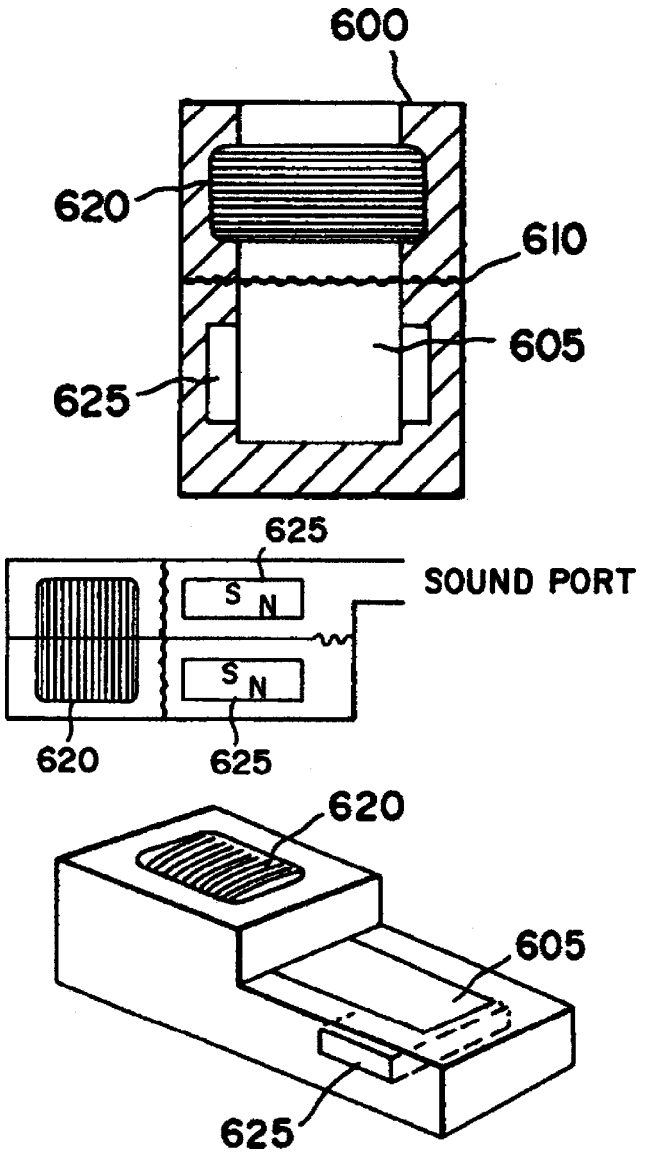
canal. According to this patent, the design shown provides ''improved comfort, sound quality, and stability.'' Earhook **110** fits behind the ear and pivoting receiver **120** is held snugly against the ear by spring **143**. Lever **141** allows the receiver to be held in its fully open position while donning the headset.—GLA

7,206,425

43.38.Si ACTUATOR FOR AN ACTIVE NOISE CONTROL SYSTEM

Michael A. Vaudrey *et al.*, assignors to Adaptive Technologies, Incorporated
17 April 2007 (Class 381/328); filed 21 January 2004

A completely-in-the-ear-canal (CIC) active noise control (ANC) speaker **600** is claimed having a diameter of 4 mm and length of 11 mm with magnet **625** of intermediate length. Coil **620**, energized with ac current,



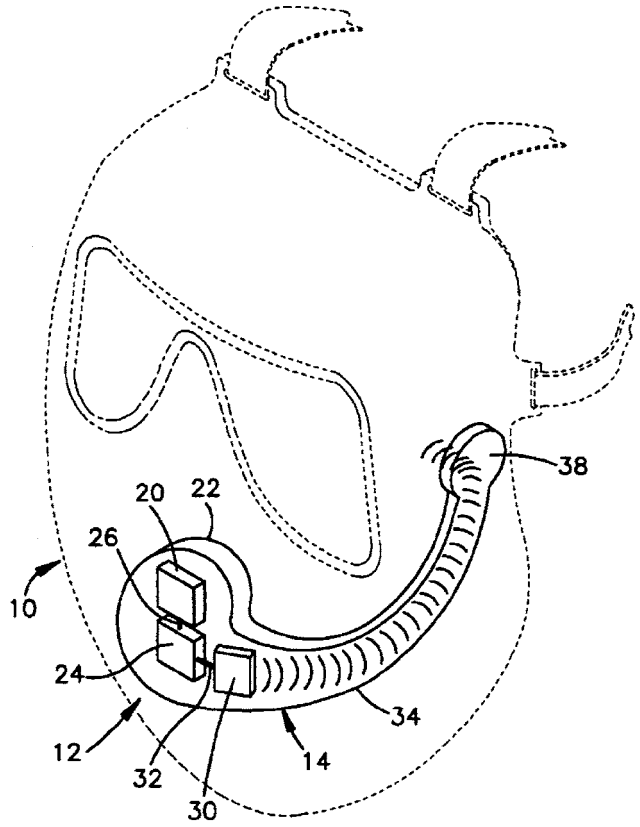
drives diaphragm **605** to produce sound through the sound port to a level as much as 110 dB SPL in a volume of 0.5 cc.—AJC

7,221,966

43.38.Si WIRELESS COMMUNICATION SYSTEMS FOR MASKS OR HELMETS

Joseph Birli *et al.*, assignors to Ultra Electronics Audio Pack, Incorporated
22 May 2007 (Class 455/569.1); filed 4 November 2003

It is difficult to pin down exactly what has been patented here. Safety masks, motorcycle helmets, and the like are readily available with built-in communications systems. The patent argues that the components of such a system should be externally mounted, or that sometimes some components should be internally mounted, or that various mounting locations should be provided. Well, we will certainly require a microphone, which ''may be of any type suitable for use in this application.'' And a receiver, which ''may be



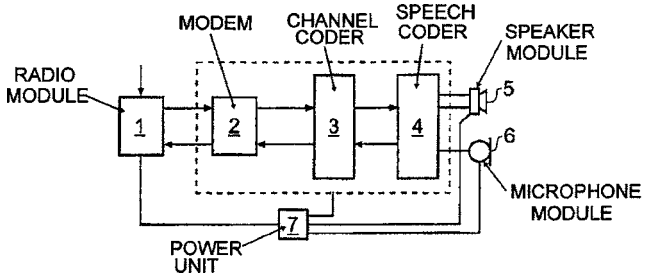
of any type suitable for use in this application.'' For two-way communications, a transceiver is required, and it ''may be any suitable small transceiver, including or associated with a battery.'' All of which ''may be mounted on mask **10** in any suitable manner.'' There may be a spark of ingenuity hidden in one of the 52 patent claims, but this reviewer frankly, lacked the patience to wade through them.—GLA

7,231,235

43.38.Si WIRELESS COMMUNICATION TERMINALS

William Basil Harrold, assignor to TTPCom Limited
12 June 2007 (Class 455/574); filed in United Kingdom 3 April 2002

Some cellular telephones come complete with speech recognition software so that, in the hands-free mode, the user can verbally dial a number and then carry on a conversation. However, the process must be initiated manually because audio circuitry is turned off to conserve power when the



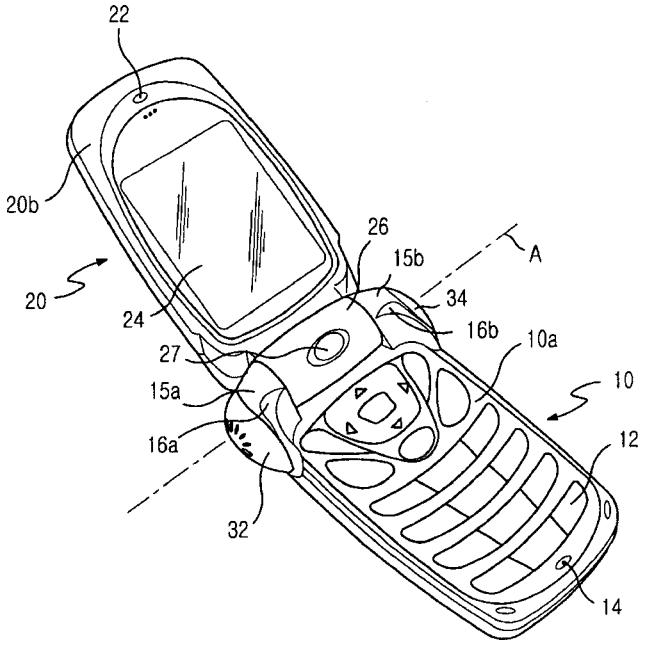
phone is in standby operation. This patent points out that even while powered down the phone periodically checks for paging signals, and that this feature can be augmented to also check for a distinctive audio signal such as a whistle. Thus, the user whistles for a half-second or so (the frequency is unimportant) and the phone responds with a chirp to acknowledge that hands-free operation has been initiated.—GLA

7,228,158

43.38.Si PORTABLE TERMINAL CAPABLE OF PROVIDING STEREO SOUND

Jong-Seong Lee *et al.*, assignors to Samsung Electronics Company, Limited
5 June 2007 (Class 455/575.3); filed in Republic of Korea 12 November 2002

This stereo cellular telephone includes a camera lens, a center speaker, and two tiny additional loudspeakers mounted in side projections **32**, **34** which are incorporated into the cover hinge assembly. The patent attorneys



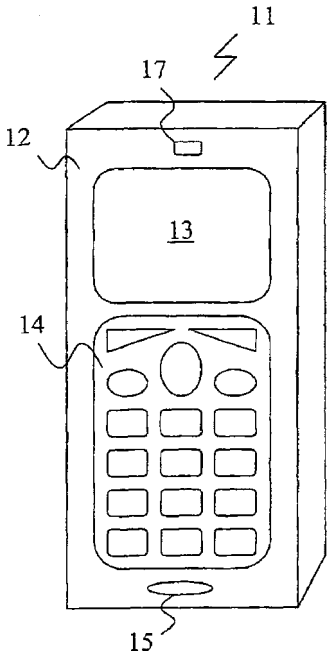
have managed to describe the overall configuration in a version of English that is almost impenetrable, however a curious recurring theme is the precise location and orientation of the screws used to attach the side loudspeaker covers.—GLA

7,233,678

43.38.Si SPEAKER ARRANGEMENT FOR COMMUNICATION TERMINAL

Mats Erixon *et al.*, assignors to Sony Ericsson Mobile Communications AB
19 June 2007 (Class 381/345); filed in the European Patent Office 1 July 2002

To conserve space in a cellular telephone, the loudspeaker is mounted inside the electromagnetic resonance chamber behind display **13**. Sound from the front of the speaker is coupled to the outside of the housing



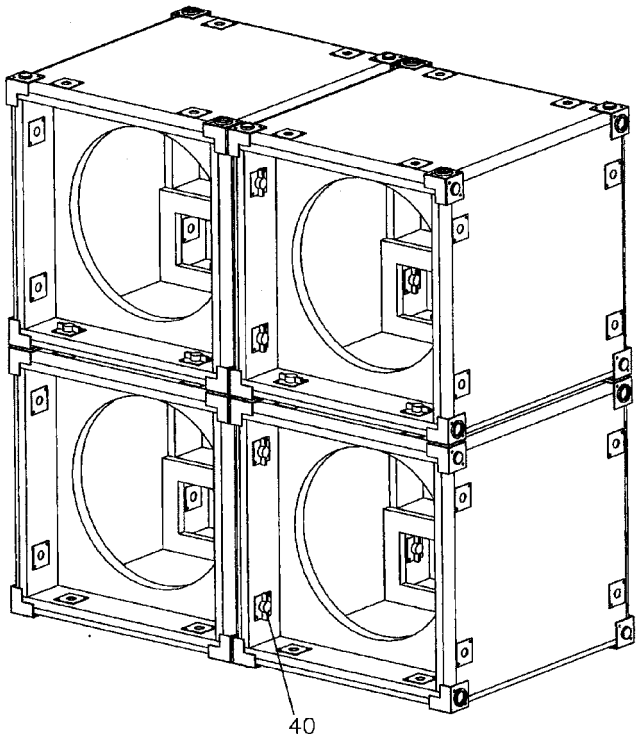
through a small opening **17**. ''Tests performed on a prototype apparatus have shown the surprising result that the speaker will not, or to a very little extent, affect the performance of the antenna.''—GLA

7,201,251

43.38.Tj MODULAR SPEAKER CABINET

Derrick Lynn Baird, Knoxville, Tennessee
10 April 2007 (Class 181/145); filed 7 July 2003

Sixteen quick connect devices **40**, eight in the front and eight in the back of a speaker cabinet, allow for the stacking of speaker cabinets in both



the horizontal and vertical directions. This allows a musician to bring just one or two for rehearsal, but more for performance.—NAS

7,221,688

43.38.WI METHOD AND APPARATUS FOR RECEIVING A DIGITAL AUDIO BROADCASTING SIGNAL

Philip H. Vanness, assignor to iBiquity Digital Corporation
22 May 2007 (Class 370/529); filed 31 July 2002

A technique is described that transmits both FM radio and digital content within the current FM radio bandwidth. As proposed, older FM demodulators would still receive regular radio broadcasts, but enhanced receivers would get additional digital content hidden within the upper- and lower-frequency sidebands. This content enhances the quality of the FM broadcast. It also allows radio stations to slowly transition from traditional analog to all digital broadcasting.—GFE

7,224,939

43.38.WI AUDIO BROADCAST RECEIVING APPARATUS AND METHOD

Kazuo Takayama and Shuji Sugahara, assignors to Fujitsu Ten Limited
29 May 2007 (Class 455/45); filed in Japan 21 November 2003

This impressively difficult-to-read patent appears to describe the same system as United States Patent 7,221,688, reviewed above.—GFE

7,231,303

43.40.Le VIBRATION SENSOR AND METHOD FOR MONITORING THE CONDITION OF ROTATING COMPONENTS AND BEARINGS

Erich Griessler *et al.*, assignors to i-for-T GmbH
12 June 2007 (Class 702/34); filed in Germany 13 April 2002

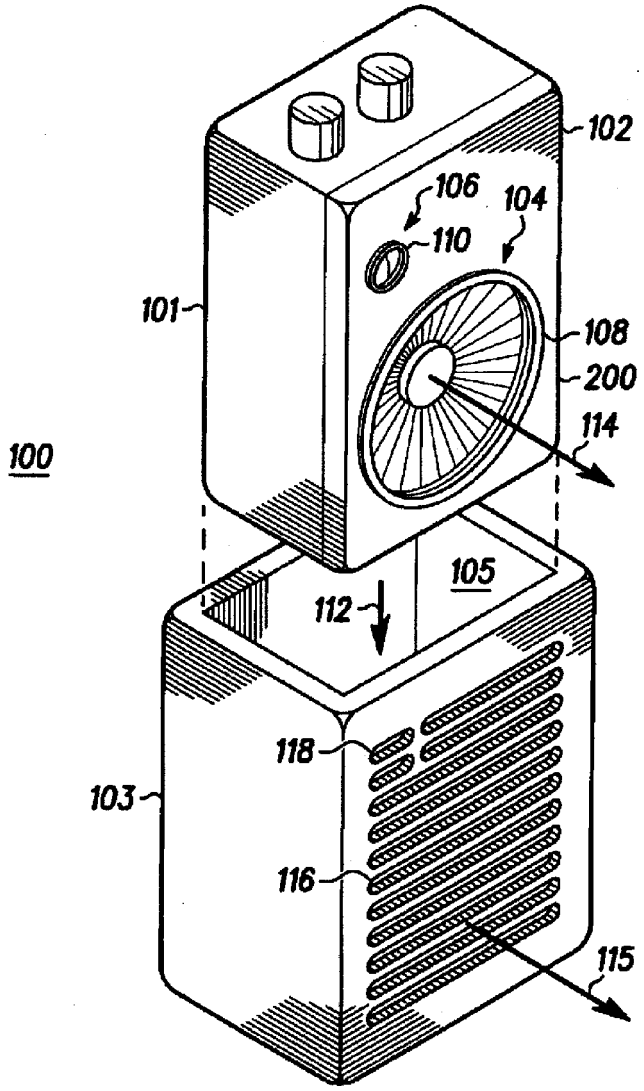
A monitoring package, intended to be mounted near a machine element that is to be monitored, includes a sensor, analog-to-digital converter, signal conditioning electronics, and evaluation means for converting signals to a state value via signal analysis and diagnostics algorithms. The package can also include display, alarm, and switching capabilities. Use of such packages is claimed to avoid the need for transmitting signals from many monitoring stations to a central computer.—EEU

7,206,422

43.40.Rj TRANSDUCER ASSEMBLY APPARATUS

Jody H. Akens *et al.*, assignors to Motorola, Incorporated
17 April 2007 (Class 381/189); filed 26 February 2003

The authors claim radio **100** that is assembled along two perpendicular axes. Microphone **106** is assembled along axis **114** onto a seal and electrical



contacts. Speaker **104** is assembled on the same axis and is retained by chassis **102**. Chassis **102** is assembled to housing **103** on axis **112**.—AJC

7,234,153

43.40.Tm VIBRATION DAMPER FOR DAMPENING VIBRATIONS AT LOW FREQUENCIES

Erik Jensen, assignor to ISIC A/S
19 June 2007 (Class 720/651); filed in Denmark 8 August 2001

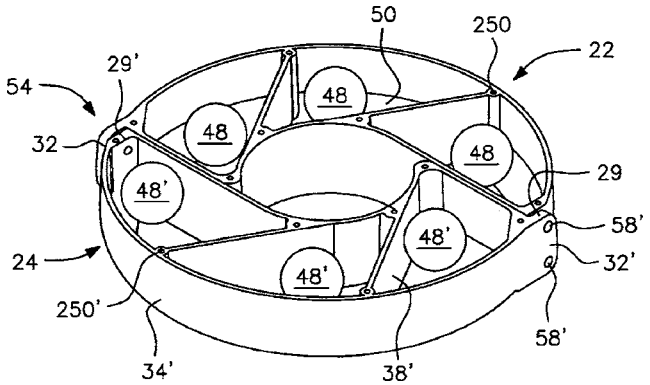
This patent relates to isolation of CD players and the like from jolts and vibrations of vehicles in which the players may be mounted. The isolation system consists of several pairs of circular wire springs, with the springs connected so that their planes are at right angles to each other, thus providing low stiffness in all directions.—EEU

7,232,017

43.40.Tm POLE VIBRATION DAMPING ASSEMBLY AND METHOD

Ray C. Minor, assignor to Kearney-National Incorporated
19 June 2007 (Class 188/378); filed 1 April 2004

This damper, intended to be mounted on light poles and similar devices in order to reduce their wind-induced vibrations, consists of a series of wedge-shaped compartments that contain lead spheres **48** and **48′**. The



device is constructed in two parts, **22** and **24**, that can be placed around a pole and bolted together. One of these dampers is intended to be mounted near the top of a pole in order to suppress vibrations in the fundamental mode. A second such device is to be mounted near the pole's midlength in order to suppress the second mode.—EEU

7,222,704

43.40.Vn DAMPING OF VIBRATIONS

John B. Pearson and Philip A. Emery, assignors to BAE Systems plc
29 May 2007 (Class 188/378); filed in United Kingdom 1 February 2002

Attenuation of resonant vibrations of a relatively flexible structure that is situated near a relatively rigid structure is accomplished by means of a force generator that acts between the two structures. Signals from sensors attached to the flexible structure are fed to a controller, which causes the force generator to produce forces that oppose the flexible structure's motion.—EEU

7,231,052

43.40.Vn VIBRATION-ISOLATING AND SOUND ISOLATING SYSTEM FOR VEHICLE

Hideki Matsuoka and Toshio Inoue, assignors to Honda Giken Kogyo Kabushiki Kaisha
12 June 2007 (Class 381/71.4); filed in Japan 18 July 2002

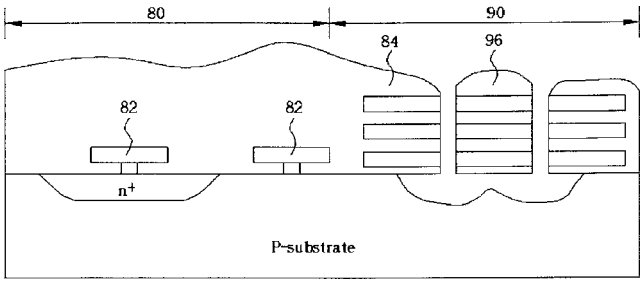
An active isolator between an automotive vehicle's engine and frame is controlled on the basis of a crank pulse signal from the engine. The output of a speaker in the vehicle compartment is controlled based on the rotational speed of the engine and on the noise sensed by a microphone located in the compartment. The system uses two different controllers, which act independently.—EEU

7,206,265

43.40.Yq DEVICE CAPABLE OF DETECTING VIBRATION/SHOCK

Pei-Ching Kuo and Cheng-Yao Liao, assignors to Lite-On It Corporation
17 April 2007 (Class 369/44.32); filed in Taiwan 4 September 2003

A disk drive read-write head vibration and shock detection sensor **80-90** is claimed where a floating electrode **96** is etched between



comb fingers **84**. Photo diodes **82** etched into the p-substrate material provide the read-write signals. Vibration will cause motion of **96** among fingers **84** which will induce a warning signal.—AJC

7,225,062

43.40.Yq DEVICE FOR INDICATING THE VIBRATION LEVEL OF A VEHICLE

Pierre-Antoine Aubourg, assignor to Eurocopter
29 May 2007 (Class 701/3); filed in France 22 December 2003

This device is intended to provide warning of excessive vibrations, for example to the pilot of a helicopter. Each of the signals from a number of vibration sensors is resolved into frequency bands. The resulting filtered components are weighted, compared to stored reference values, and combined to generate an evaluation signal. The various evaluation signals then are weighted and combined into a single signal that is compared to an alarm threshold.—EEU

7,225,108

43.40.Yq CROSS CORRELATION DIAGNOSTICS TOOL FOR VIBRATION ANALYSIS

Burton R. Clarke and Ryan P. Allgaier, assignors to Caterpillar Incorporated
29 May 2007 (Class 702/183); filed 10 November 2005

Detection and analysis of problems in a machine during operation is accomplished using vibration sensors, acoustic sensors, and transducers for measuring other parameters, such as temperature and pressure. Signals received from the various sensors are correlated and all the resulting signals are cross-correlated. Anomalies are determined from the strongest cross-correlation, possibly using neural network or autoregression algorithms. The analysis means may be located remotely from the machine, receiving signals via a communications link.—EEU

7,225,109

43.40.Yq METHOD AND APPARATUS TO DIAGNOSE MECHANICAL PROBLEMS IN MACHINERY

Donald Karl Frerichs *et al.*, assignors to ABB Incorporated
29 May 2007 (Class 702/185); filed 13 January 2005

A model of a process in which machinery is employed is developed by use of a technique such as advanced pattern recognition, and this model is used to generate predicted values for a number of the process' operating parameters and of the machinery's vibration parameters. Statistical process control methods are used to determine whether the difference between the predicted and measured values exceeds a statistical limit. Alarms or notifications are developed via a rules set.—EEU

7,228,846

43.50.Gf APPARATUS, SYSTEM, AND METHOD FOR REDUCING PRESSURE PULSATIONS AND ATTENUATING NOISE TRANSMISSION IN A FUEL SYSTEM

Jason T. Kilgore, assignor to Siemens VDO Automotive Corporation
12 June 2007 (Class 123/447); filed 3 June 2004

A fuel line conveying liquid fuel from a supply to a nozzle has one or more helical turns. Pressure pulses in the fuel system can cause the helical turns to uncoil, reducing the transmission of these pulses.—EEU

7,226,656

43.55.Ev THERMOFORMABLE ACOUSTIC SHEET

Michael William Coates and Marek Kierzkowski, assignors to I.N.C. Corporation
5 June 2007 (Class 428/221); filed in Australia 19 July 2000

This acoustic sheet uses a glass fiber web with adhesive fibers so that when it is heated, the composite sheet is compressed into a moldable form. The sheet can be used under the hood of an automobile or for similar applications around an engine.—CJR

7,231.094

43.58.Kr LOW-POWER SIGNAL PROCESSING USING MEMS

James A. Bickford and Paul Ward, assignors to The Charles Stark Draper Laboratory, Incorporated
12 June 2007 (Class 382/248); filed 1 October 2002

The authors disclose a farfetched concept for analog computation of the Fourier transforms of images using an array of MEMS resonators. There is nothing new about this except for the notion that it could actually work, and the applications of such a device to image processing. For instance, there is nothing to address how the device would be tuned so that its frequency bands were the ones that you would be interested in examining, so it would be similar to having a microprocessor with a single clock speed. It is also unexplained how one makes resonators having a frequency span on the order of 1024:1 on a single substrate. On the other hand, it could calculate a Fourier transform really fast. Perhaps this would be a good project for that pesky graduate student.—JAH

7,225,405

43.58.Ta SYSTEM AND METHOD FOR AUDIO CREATION AND EDITING IN A MULTIMEDIA MESSAGING ENVIRONMENT

John W. Barrus *et al.*, assignors to Ricoh Company, Limited
29 May 2007 (Class 715/716); filed 9 January 2002

The self-evident idea of dragging or recording audio files into email, text, or presentations was misappropriated for this patent.—GFE

7,227,944

43.60.Dh ECHO REDUCTION FOR A HEADSET OR HANDSET

John F. Gerhardt *et al.*, assignors to Plantronics, Incorporated
5 June 2007 (Class 379/390.02); filed 22 March 2006

A method is described to reduce echoes common in network telephony. A known audio signal is transmitted and its cross-talk measured in order to better extract the echoes.—GFE

7,231,051

43.60.Jn DETECTION OF VIEWING DIRECTION BY MICROPHONE

Florent Paviot and Siegfried Rothe, assignors to DaimlerChrysler AG
12 June 2007 (Class 381/56); filed in Germany 17 April 2002

This proposed driver safety device attempts to detect the drivers look direction from their speaking direction and then take defensive actions.—GFE

7,221,765

43.66.Ts SYSTEM AND METHOD FOR INDIVIDUALIZED TRAINING OF HEARING AID USERS

Josef Chalupper and Reinier Kortekaas, assignors to Siemens Audiologische Technik GmbH
22 May 2007 (Class 381/60); filed in the European Patent Office 12 April 2002

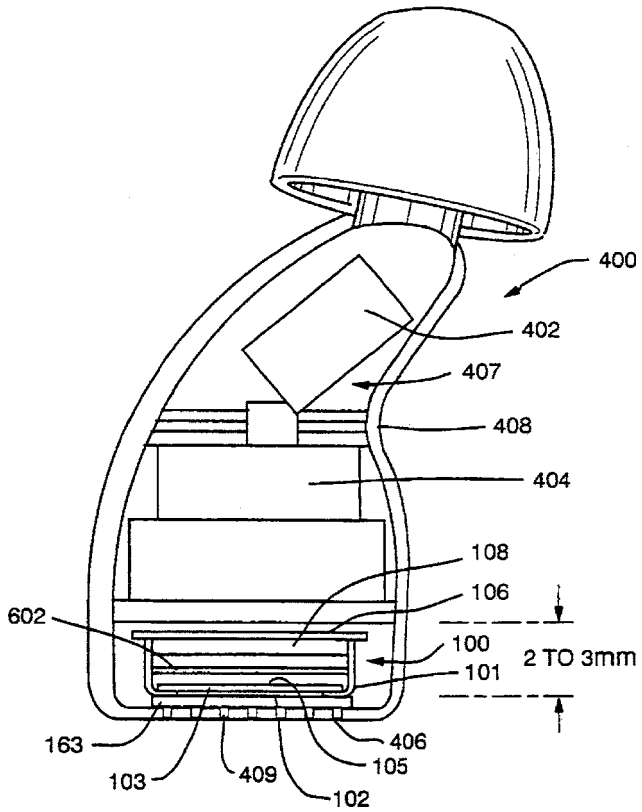
The object is to provide individualized training procedures for hearing aid wearers. A databank stores a number of training courses associated with different hearing devices and is connected to a multimedia training device for acquiring a user profile. The system selects an appropriate level training course according to the user profile via a data network such as the Internet.—DAP

7,221,768

43.66.Ts HEARING AID WITH LARGE DIAPHRAGM MICROPHONE ELEMENT INCLUDING A PRINTED CIRCUIT BOARD

Walter P. Sjursen *et al.*, assignors to Sarnoff Corporation
22 May 2007 (Class 381/322); filed 21 February 2006

An electrically-conductive microphone housing for a disposable hearing aid is sealed via a printed circuit board packaging the signal processing



amplifier circuitry, thereby providing EMI shielding for all of the electrical components. The single large diaphragm in the microphone may be divided into several smaller active diaphragm areas.—DAP

7,221,769

43.66.Ts HEARING AID ADAPTED FOR DISCRETE OPERATION

Martin Bondo Jorgensen, assignor to Sonion Roskilde A/S
22 May 2007 (Class 381/328); filed in Denmark 24 September 1998

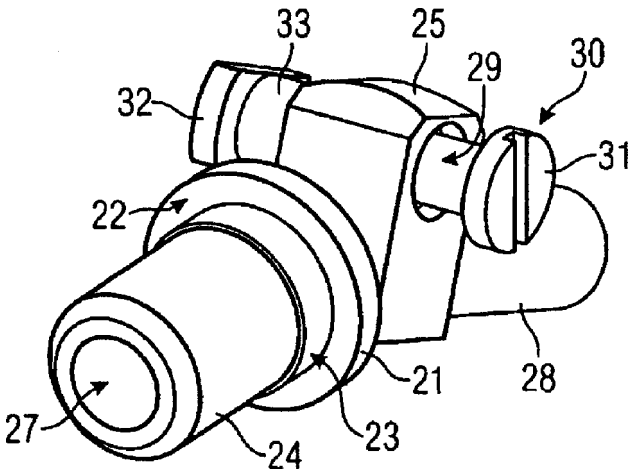
A hearing aid wearer is given the capability of turning the hearing aid power on and off, adjusting the volume control, and selecting a particular program memory by applying pressure with a finger to a predetermined part of the outer ear on/in which the hearing aid is worn. The hearing aid, which may fit behind the ear or in the ear canal, contains a sensing mechanism to respond to force from finger pressure and a controlling mechanism consisting of a stiff lever, which may also be used to remove the hearing aid from the ear canal.—DAP

7,221,770

43.66.Ts CONNECTION PIECE FOR HEARING DEVICE SUPPORT HOOK

Werner Fickweiler and Michael Sattler, assignors to Siemens Audiologische Technik GmbH
22 May 2007 (Class 381/330); filed in Germany 22 July 2003

A one-piece metal tubular connector for a detachable ear hook on a behind-the-ear hearing aid incorporates a fastening section between two tube sections that are connected together and are hollowed out for a sound channel. One of the tube sections is curved and attaches to a stop having a



sealing surface that also absorbs force from the ear hook. A fastening clip with a screw holds the assembly firmly to the hearing aid housing. The advantage is said to be that a reduced cross-section of the hearing aid housing is allowed in the region of the ear hook connection.—DAP

7,224,815

43.66.Ts HEARING AID DESIGN

Albert A. Maltan *et al.*, assignors to Advanced Bionics Corporation
29 May 2007 (Class 381/322); filed 1 September 2004

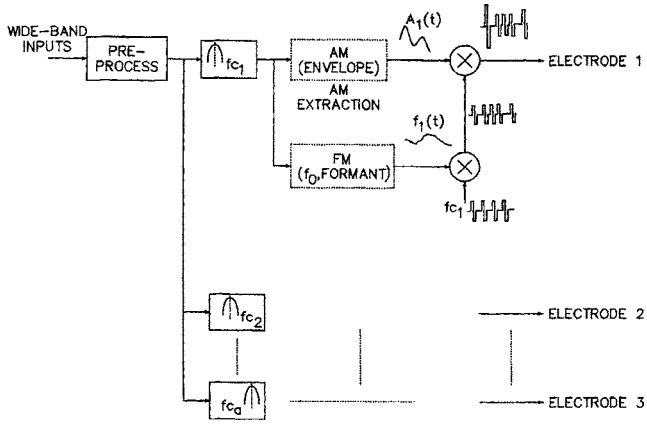
Components of a hearing aid are packaged into a module that is shaped for insertion into a tunnel made through the soft tissue connecting the retroauricular space with the ear canal. The microphone may be located in the behind-the-ear portion of the module and the receiver (speaker) is positioned so that the sound output is directed into the ear canal. Additionally, the signal processing circuitry within the module may perform voice command recognition and circuitry may be included for transmitting sounds picked up by the microphone via a wireless link to a remote device.—DAP

7,225,027

43.66.Ts COCHLEAR IMPLANTS AND APPARATUS/METHODS FOR IMPROVING AUDIO SIGNALS BY USE OF FREQUENCY-AMPLITUDE-MODULATION-ENCODING (FAME) STRATEGIES

Fan-Gang Zeng and Kai-Bao Nie, assignors to Regents of the University of California
29 May 2007 (Class 607/57); filed 27 August 2002

A combination AM-FM encoding strategy is used to process signals to be sent to a cochlear implant. The strategy encodes changes in signals rather than the absolute signals. Band-specific frequency modulations, which may or may not carry fundamental frequency information, are extracted and used to modulate the carrier frequency in each band. The pulse train is first frequency modulated and then the result is amplitude modulated. The advantage is that a reduced bandwidth of only about 5 kHz is required for 8–10 channels of information without degrading sound quality.—DAP

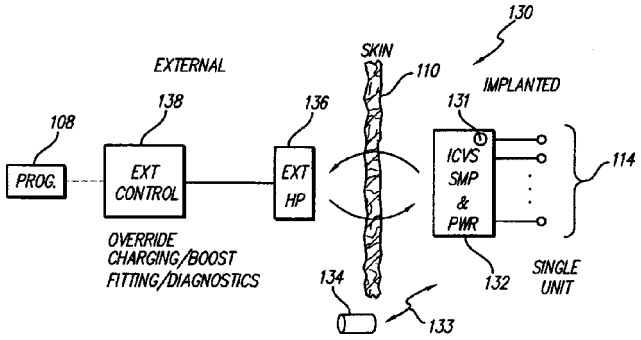


7,225,028

43.66.Ts DUAL COCHLEAR/VESTIBULAR STIMULATOR WITH CONTROL SIGNALS DERIVED FROM MOTION AND SPEECH SIGNALS

Charles C. Della Santina and Michael A. Faltys, assignors to Advanced Bionics Corporation
29 May 2007 (Class 607/57); filed 20 May 2005

Systems and methodology are recommended to simultaneously enable both auditory and vestibular sensations in patients with combined deafness and vestibular hypofunction. A hybrid cochlear/vestibular implantable device provides vestibular stimulation with spatial information including acceleration, velocity and/or position of the head or body, and an estimate of the gravity vector. This information is derived from motion sensors such as rotational and linear accelerometers that can sense movement of the patient's head.—DAP



7,226,406

43.66.Ts AT LEAST PARTIALLY IMPLANTABLE HEARING SYSTEM

Gerd M. Müller and Hans Leysieffer, assignors to Cochlear Limited
5 June 2007 (Class 600/25); filed in Germany 25 September 2000

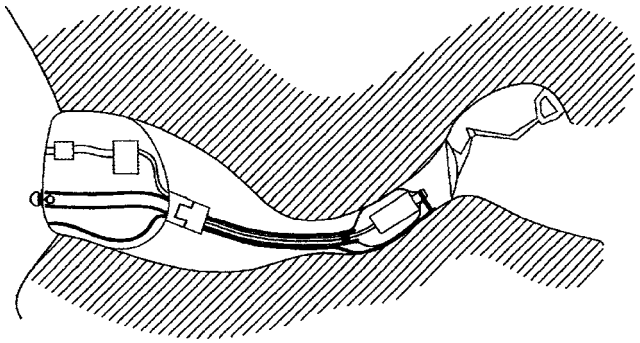
In order to enable easier replacement and accurate positioning of the implanted output transducer, a mechanical positioning micromanipulator is attached permanently to the cranial vault of the person being implanted. A releasable coupling may be utilized for snap-in output transducer replacement.—DAP

7,227,968

43.66.Ts EXPANDSIBLE RECEIVER MODULE

Aart Zeger van Halteren *et al.*, assignors to Sonion Roskilde A/S
5 June 2007 (Class 381/328); filed 24 June 2002

A hearing aid receiver (speaker) is packaged in an expandible housing designed to be positioned deeply into the bony portion of the ear canal. The



housing expansion is accomplished via inflation of a balloon-like device with the receiver acting as a pump via a tubular connection to the housing.—DAP

7,231,055

43.66.Ts METHOD FOR THE ADJUSTMENT OF A HEARING DEVICE, APPARATUS TO DO IT AND A HEARING DEVICE

Bohumir Uvacek and Herbert Bachler, assignors to Phonak AG
12 June 2007 (Class 381/312); filed 24 October 2001

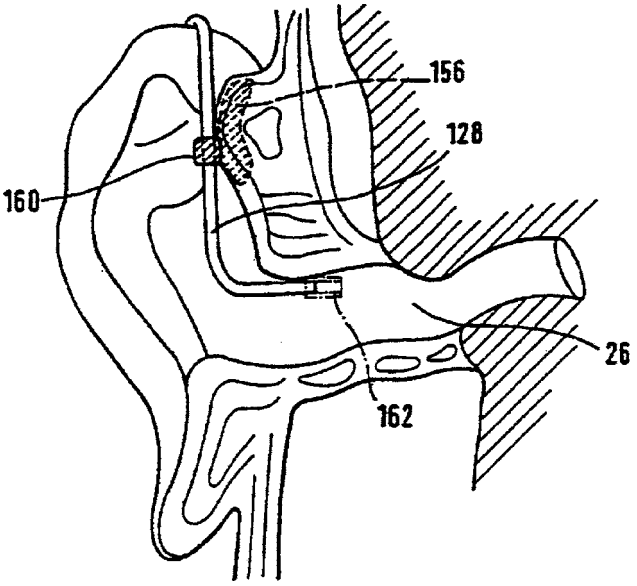
A hearing device is adjusted with an apparatus for psychoacoustic parameters while it is on the wearer. For example, the appropriate loudness in critical bands is provided based on the difference between a normal hearing loudness model and the loudness perceived by the hearing impaired individual.—DAP

7,233,676

43.66.Ts OTOPLASTY FOR BEHIND-THE-EAR (BTE) HEARING AIDS

Erich Bayer, Straubing, Germany
19 June 2007 (Class 381/328); filed in Germany 14 October 1999

This patent addresses an earpiece design for an open behind-the-ear hearing aid fitting that is said to be less visible and more comfortable while retaining more of the natural resonance of the ear than previous designs.



Retention of the earpiece occurs via an arc-shaped clip around the outer edge of the concha external ear and leaves the lower half of the wearer's ear canal open.—DAP

7,231,257

43.66.Ts COCHLEAR IMPLANT SOUND PROCESSING METHOD AND SYSTEM

Hugh McDermott and Colette McKay, assignors to The University of Melbourne
12 June 2007 (Class 607/57); filed in Australia 28 February 2003

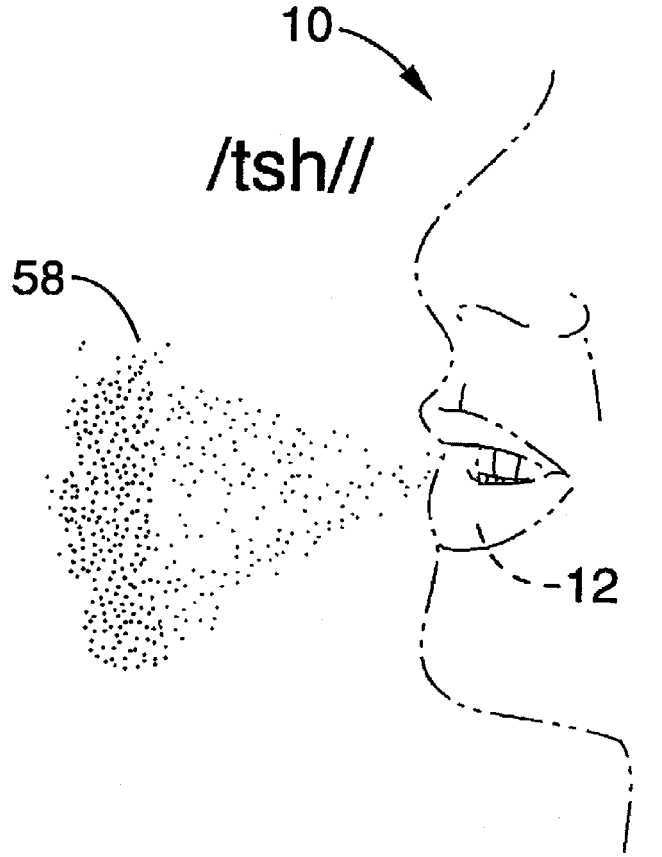
A signal processor generates a temporal pattern of stimulation pulses at a specific rate to selected electrodes within an implanted electrode array. The interleaved periods of the stimulation pulses are adjusted in accordance with a conversion function related to one or more estimated fundamental frequencies that are extracted from the signal.—DAP

7,225,129

43.71.Ky VISUAL DISPLAY METHODS FOR IN COMPUTER-ANIMATED SPEECH PRODUCTION MODELS

Dominic W. Massaro *et al.*, assignors to The Regents of the University of California
29 May 2007 (Class 704/261); filed 20 September 2001

In order to facilitate speech therapy, a talking head model is proposed



that not only visualizes face movement, but, breath, voice vibration, and other features.—GFE

7,069,214

43.72.Ar FACTORIZATION FOR GENERATING A LIBRARY OF MOUTH SHAPES

Jean-Claude Junqua, assignor to Matsushita Electric Industrial Company, Limited
27 June 2006 (Class 704/235); filed 12 March 2002

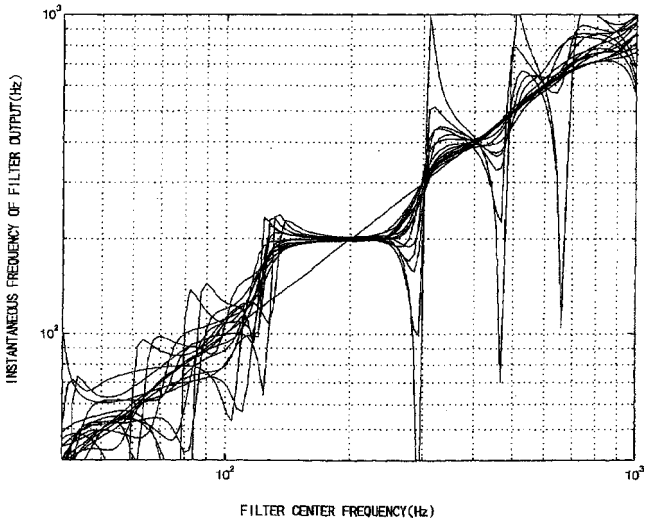
Here is yet another application of eigenvector analysis to recover a set of individual characteristics from a known (trained) population data set. The application is mouth shapes as measured in some unspecified manner for a known population of speakers. The mouth shapes are parameterized in some equally unspecified manner, resulting in a database of shape parameters vs speakers. Any one of a variety of analysis methods is then applied to determine speaker-specific eigenvectors within the data space. Suggested analysis methods include principal components analysis, independent components analysis, linear discriminant analysis, factor analysis, or singular value decomposition. An optional rank reduction may then be applied, at least for some of the analysis methods. Given a few mouth shape parameters for a new speaker, the entire mouth shape space for that speaker may then be recovered from the best fitting eigenvector for that speaker.—DLR

7,085,721

43.72.Ar METHOD AND APPARATUS FOR FUNDAMENTAL FREQUENCY EXTRACTION OR DETECTION IN SPEECH

Hideki Kawahara and Toshio Irino, assignors to Advanced Telecommunications Research Institute International
1 August 2006 (Class 704/258); filed in Japan 7 July 1999

Described is a speech fundamental frequency analyzer based on a process of instantaneous frequency extraction in individual filters of a close-spaced filter bank. The bank would typically cover a frequency range of 20–800 Hz with up to 24 bands/octave of individual band filters consisting of Gaussian-shaped Gabor functions. In the analysis, individual harmonics



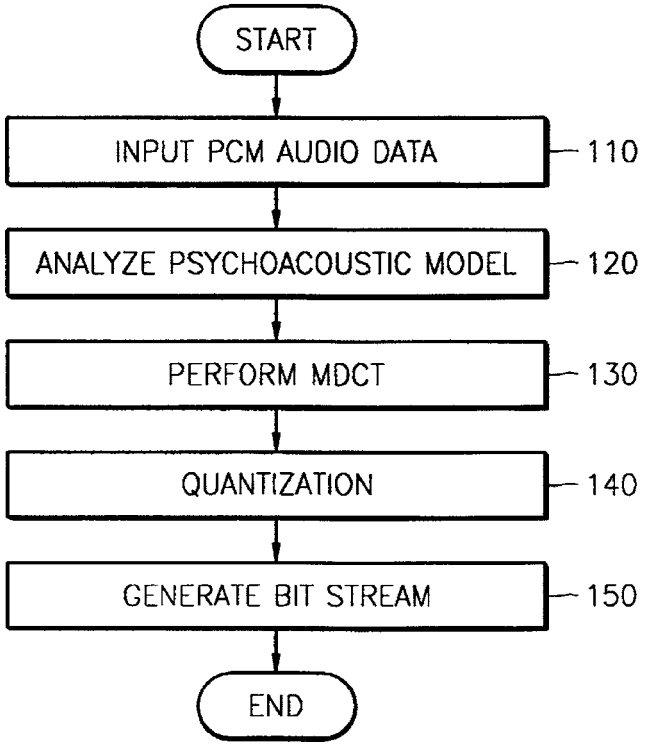
are recovered in groups of adjacent bands and their instantaneous frequencies are computed by partial differentiation in both time and frequency. The analysis allows a precise determination of the moment of excitation of each harmonic up to the 800 Hz top end. All math derivations are presented in detail, along with a summary of performance for male and female speakers.—DLR

7,225,123

43.72.Gy METHOD FOR COMPRESSING AUDIO SIGNAL USING WAVELET PACKET TRANSFORM AND APPARATUS THEREOF

Ho-jin Ha, assignor to Samsung Electronics Company Limited
29 May 2007 (Class 704/200.1); filed in Republic of Korea 16 February 2002

Using a psychoacoustic model, the perceptual energy level in each of 32 bands is compared to a preset threshold to determine whether a modified discrete cosine transform (MDCT) or a wavelet packet transform (WPT) processing window is used. If the level of the calculated perceptual energy is



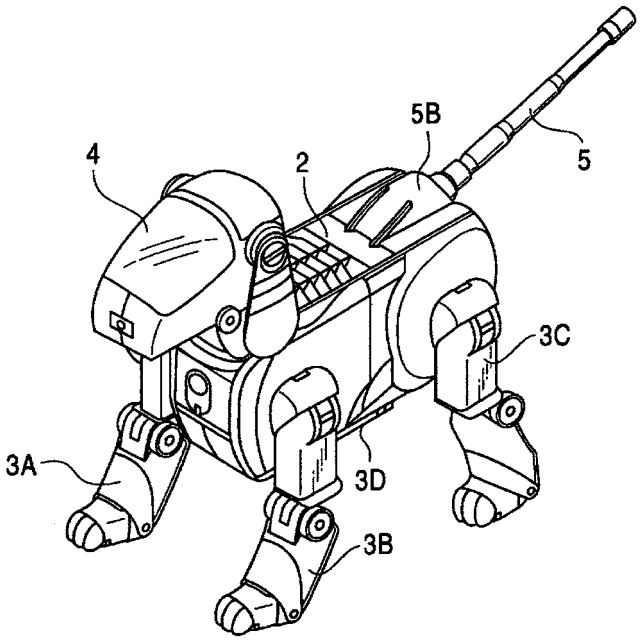
greater than the threshold, a WPT window (for attack type signals) is selected or, if less than the threshold, an MDCT window (for steady state signals) is selected. The sound quality is said to improve on previous MDCT-based methods, even at low bitrates.—DAP

7,065,490

43.72.Ja VOICE PROCESSING METHOD BASED ON THE EMOTION AND INSTINCT STATES OF A ROBOT

Yasuharu Asano and Hongchang Pao, assignors to Sony Corporation
20 June 2006 (Class 704/275); filed in Japan 30 November 1999

This patent describes a robotic toy with speech synthesis capability that includes a sensitivity to various emotional states; being able to adjust the nature of the output speech dependent on internal states or settings corresponding roughly to human emotions. These internal states may be expressed in the form of changes in the speech pitch, speaking rate, or may also involve changes in the vocabulary selected to express a particular condition. The greater part of the patent text and of the claims, deals with the sensory side of emotions, the ability of the toy to sense changes in the environment and, as a result, to make appropriate changes in its internal



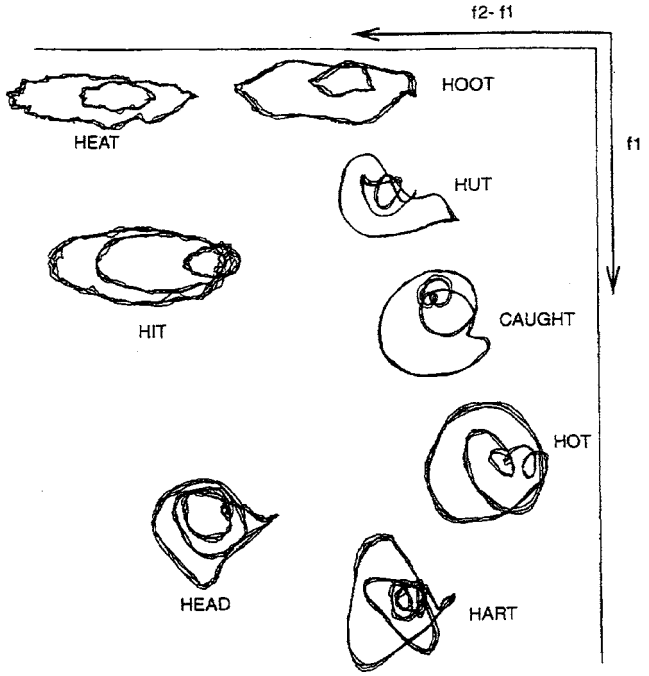
state. Sensors allow the toy to detect changes in how it is being handled, touched, etc., as well as to detect patterns in the the voice of the interacting human, such as pitch alterations or speaking rate changes.—DLR

7,069,217

43.72.Ja WAVEFORM SYNTHESIS

Stephen McLaughlin and Michael Banbrook, assignors to British Telecommunications PLC
27 June 2006 (Class 704/269); filed in United Kingdom 15 January 1996

This method of waveform generation for speech synthesis is said to provide a smooth and continuously varying speech output signal, more like human speech than is provided by a typical segment concatenation synthesizer. The method, applicable only for voiced speech, consists of constructing trajectories in a low-dimensional state space by computing each next trajectory point based on the current state and a continually moving and randomized target point in the desired vowel space, such that a vowel of the desired vowel quality and duration is generated. The figure shows some typical trajectories as measured in the f1 vs f2-f1 vowel space.—DLR



7,225,126

43.72.Kb SYSTEM AND METHOD FOR PROCESSING SPEECH FILES

Julia Hirschberg and Stephen Whittaker, assignors to AT&T Corporation
29 May 2007 (Class 704/235); filed 12 June 2001

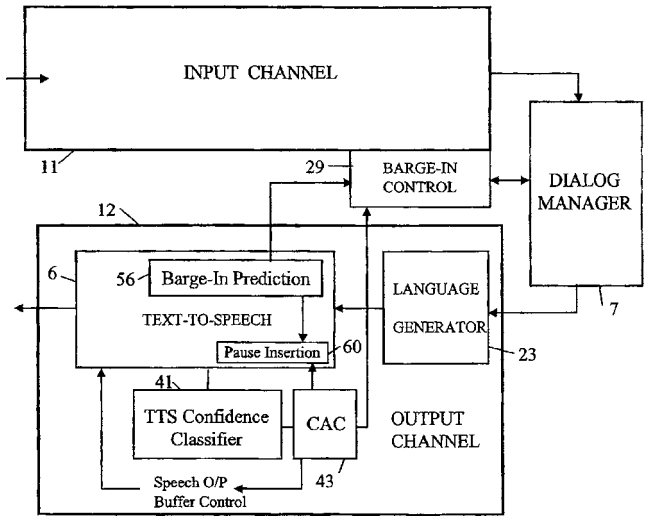
This painfully obvious patent allows people to shorten their voicemail messages into sound bites that they could then store or pass on.—GFE

7,062,440

43.72.Ne MONITORING TEXT TO SPEECH OUTPUT TO EFFECT CONTROL OF BARGE-IN

Paul St. John Brittan and Roger Cecil Ferry Tucker, assignors to Hewlett-Packard Development Company, L.P.
13 June 2006 (Class 704/266); filed in United Kingdom 4 June 2001

This patent covers another aspect of the talking computer system discussed in United States Patent 7,062,439 [reviewed in J. Acoust. Soc. Am. **122**(4), 1869 (2007)]. Here, the spotlight is on barge-in control, enhanced by advantageous cooperation between the recognition and synthesis subsystems. The barge-in system should not be overly sensitive, such that it might interrupt the speech output upon detection of extraneous noises, etc. In the words of the patent, it requires ''careful management.'' Two conditions are identified which can result in changes in the barge-in sensitivity setting. The more obvious case is when the speech output includes a prompt or other phrase requesting a user response. Another situation discussed here is the case in which the speech output quality is substandard or unusual in some respect, which might tend to elicit a user response. Either condition results in a change in the speech output settings, allowing the user to interrupt more easily.—DLR



7,069,213

43.72.Ne INFLUENCING A VOICE RECOGNITION MATCHING OPERATION WITH USER BARGE-IN TIME

Scott A. Thompson, assignor to NetByTel, Incorporated
27 June 2006 (Class 704/231); filed 9 November 2001

A simple, if infuriating, technique is presented which could perhaps improve the accuracy of a speech recognition system in a particular situation. The scenario is the use of voice response for making a purchase by telephone. As the field of possible choices is narrowed, the user is asked to ''choose an item from the following list.'' A list of possible choices is then read out and the time at which the user's response was detected is taken into account in identifying the user's response. The assumption is of course made that the pliant user will slavishly follow directions and speak the chosen word at just the right time and not say something else entirely—DLR

7,069,220

43.72.Ne METHOD FOR DETERMINING AND MAINTAINING DIALOG FOCUS IN A CONVERSATIONAL SPEECH SYSTEM

Daniel M. Coffman *et al.*, assignors to International Business Machines Corporation
27 June 2006 (Class 704/275); filed 13 August 1999

This conversational, interactive speech system uses a dialog manager in a way that is more general than is done in prior art systems. Rather than focusing on one particular application at a time, here, the dialog manager would not follow a given set of application rules, but would have a more general set of interaction guidelines. As an application is presented to be executed, the dialog manager would incorporate the goals of the new application into the ongoing dialog with the user, a dialog state referred to here as a ''multimodal history.'' The approach is said to allow the user to switch easily between different applications, and provides the system with a broader base from which to disambiguate input commands.—DLR

7,069,221

43.72.Ne NON-TARGET BARGE-IN DETECTION

Matthew D. Crane, Charlestown, Massachusetts *et al.*
27 June 2006 (Class 704/275); filed 26 October 2001

This speech recognition system would include a barge-in sound analyzer to determine whether a detected sound should in fact cause an interruption in the on-going speech output or whether the output should continue at the point of interruption, or at some prior point, such as a previous word or phrase boundary. The first task for the barge-in analyzer is to determine whether the interrupting sound was in fact speech. If it was, then additional analyses are performed to determine whether the speaker was actually the user and, if so, what was said. The current dialog state may also be used to help in making these decisions.—DLR

7,228,275

43.72.Ne SPEECH RECOGNITION SYSTEM HAVING MULTIPLE SPEECH RECOGNIZERS

Norikazu Endo *et al.*, assignors to Toyota InfoTechnology Center Company, Limited
5 June 2007 (Class 704/235); filed 13 January 2003

A computerized decision module selects between the outputs of a first speech recognizer and a second speech recognizer based on adjusted confidence scores reflecting accuracy estimates. The adjusted confidence scores are calculated by multiplying a speech selection parameter that is derived from the speaker's location, identity, accent, or speech context with the confidence score and adding another speech selection parameter derived from similar external data.—DAP

7,221,972

43.80.Vj ULTRASOUND SYSTEM WITH PROTOCOL-DRIVEN USER INTERFACE

John I. Jackson *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
22 May 2007 (Class 600/407); filed 16 July 2004

The processor in this ultrasound system includes a controller that accepts commands from an input device and enables a transition from one stage in an examination protocol to another stage in the protocol by using a single input.—RCW

7,223,238

43.80.Vj METHOD AND DEVICE FOR MARKING SKIN DURING AN ULTRASOUND EXAMINATION

Rebecca L. Swanbom, Paonia, Colorado
29 May 2007 (Class 600/437); filed 18 October 2004

The skin of a patient is marked by a device to denote a feature of interest during an ultrasound examination. The device has two parts. One part attaches the device to an ultrasound transducer. The other part marks the skin to denote the feature of interest. The two parts of the device can be rotated relative to each other.—RCW

7,223,240

43.80.Vj ULTRASONIC DIAGNOSTIC APPARATUS

Masaru Murashita, assignor to Aloka Company, Limited
29 May 2007 (Class 600/443); filed in Japan 14 November 2002

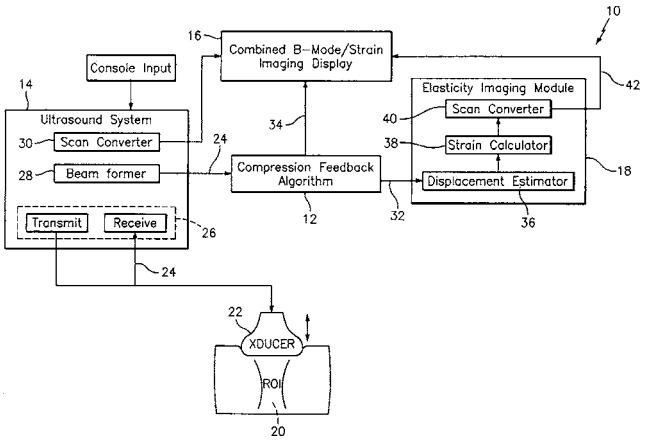
Three-dimensional ultrasonic image data are acquired and stored in a memory. A predetermined threshold and an inversion of the brightness scale are used to display the data. This converts a high-brightness organ wall to a low brightness and shows a low-brightness cavity with a high brightness to facilitate observation of the cavity.—RCW

7,223,241

43.80.Vj METHOD AND APPARATUS FOR ELASTICITY IMAGING

Emil G. Radulescu, assignor to Aloka Company, Limited
29 May 2007 (Class 600/443); filed 16 December 2004

A computationally efficient algorithm is used to analyze images obtained from a hand-held transducer that is part of an ultrasound imaging system. The algorithm selects advantageous precompression and postcompression frame pairs from which elasticity images are derived with optimal



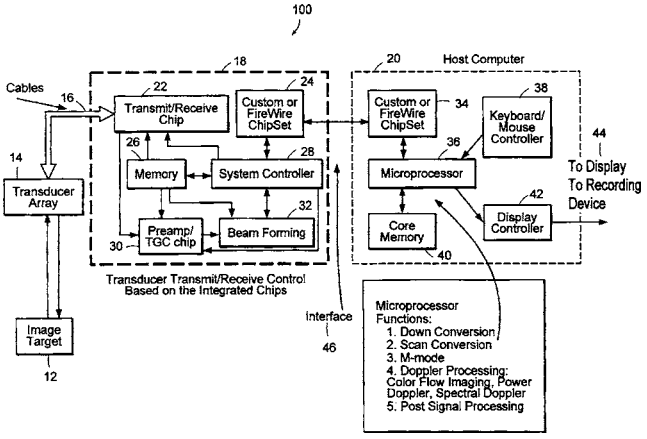
dynamic ranges and signal-to-noise ratios. The real-time operation of the algorithm facilitates operator training and reduces elasticity image artifacts.—RCW

7,223,242

43.80.Vj ULTRASOUND IMAGING SYSTEM

Xingbai He *et al.*, assignors to Teratech Corporation
29 May 2007 (Class 600/454); filed 27 September 2002

This system contains a processing module and associated memory that together map serial data to a vector representation, calculate an autocorrelation of the data and a phase shift of the autocorrelation to find a range of Doppler velocities, and produce an image for display.—RCW



7,226,416

43.80.Vj SYSTEM AND METHOD FOR PHASE INVERSION ULTRASONIC IMAGING

Glen McLaughlin and Ting-Lan Ji, assignors to ZONARE Medical Systems, Incorporated
5 June 2007 (Class 600/443); filed 21 August 2002

Multiple sets of transmit pulses that differ in amplitude, frequency, phase, or pulse width are transmitted, received, and combined to form an average signal that is used to produce an ultrasound image.—RCW

7,229,411

43.80.Vj IMAGING, THERAPY, AND TEMPERATURE MONITORING ULTRASONIC SYSTEM

Michael H. Slayton and Peter G. Barthe, assignors to Guided Therapy Systems, Incorporated
12 June 2007 (Class 600/439); filed 10 July 2002

A single transducer assembly is connected to subsystems for imaging, therapy, and monitoring temperature. The imaging subsystem produces an image of a treatment region. The therapy subsystem excites the transducer with a high power to heat the treatment region. The temperature monitoring subsystem maps and monitors the temperature of the treatment region. The transducer assembly can be configured to provide three-dimensional imaging, therapeutic treatment, and temperature monitoring data.—RCW

7,229,412

43.80.Vj VIEWING SYSTEM HAVING MEANS FOR PROCESSING A SEQUENCE OF ULTRASOUND IMAGES FOR PERFORMING A QUANTITATIVE ESTIMATION OF FLOW IN A BODY ORGAN

Marie Jacob *et al.*, assignors to Koninklijke Philips Electronics N.V.
12 June 2007 (Class 600/454); filed in the European Patent Office 28 December 2001

Mitral regurgitation is assessed by processing a sequence of three-dimensional ultrasound color flow images to determine flow velocity in a volume, construct isovelocity surfaces, and compute the surface of an orifice through which the flow passes. Peak velocity of flow through the orifice is computed as a function of the flow velocity. The flow volume is also determined.—RCW

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

Performance prediction of passive time reversal communications (L)

H. C. Song[a] and W. S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238, USA

Sea-Moon Kim[b]
Korea Ocean Research and Development Institute, Daejeon 305-343, Korea

(Received 27 March 2007; revised 7 August 2007; accepted 9 August 2007)

A recent paper [M. Stojanovic J. Acoust. Soc. Am. **117**, 1173–1185 (2005)] presented theoretical performance bounds of time reversal communications. In this letter, the performance of time reversal communications is evaluated using at-sea experimental data and compared to the theory for two different approaches; (1) time reversal alone and (2) time reversal combined with channel equalization. It is found that approach (1) shows a good agreement between theory and data. On the other hand, approach (2) indicates that the data is about 3–5 dB below the theory which assumes perfect knowledge of the channel, an infinite number of taps to remove the intersymbol interference, and no need for phase tracking. Taking these into consideration, the theoretical performance can provide a useful upper bound for predicting performance of time reversal communications.
 [DOI: 10.1121/1.2782820]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg [DRD] Pages: 2517–2518

The impact of spatial diversity in passive time reversal communications[1] has been addressed using at-sea experimental data from transmissions between a single transmitter and a multielement receiver array [see Fig. 1(a)]. The performance of two different approaches was investigated in terms of output signal-to-noise ratio (SNR) versus the number of receiver elements M: (1) time reversal alone (TR) and (2) time reversal combined with channel equalization (TR+EQ). Earlier, Stojanovic[2] developed theoretical performance bounds for various approaches including (1) and (2), under the assumption that the time-invariant channel transfer functions are known to the receiver, an infinite number of taps are used for linear equalization, and no phase tracking is required. Furthermore, recently it has been confirmed[3] that approach (2) offers nearly optimal performance in theory when the receive array provides appropriate spatial diversity.

The objective of this letter is to compare the performance between theory and data. For this purpose, data from a recent experiment conducted in a shallow region of the Mediterranean Sea is chosen for two specific reasons. First, the data exhibit a high input SNR across the array (e.g., 12–19 dB), allowing for a clean capture of the channel impulse response (CIR) from a probe signal preceding a communication data packet. Second, the environment was very stable such that the channel remained relatively time invariant during the transmission of each data packet (about 10 s) with both the transmitter and receivers fixed, except that there is a small Doppler shift due to mismatch in sampling rate[1] ($f_d=-0.054$ Hz). As a result, the data can be evaluated under conditions similar to those assumed in the theory. An example of the CIR from a probe at 96 m depth in a flat region of 120 m deep water received to a 32-element vertical receiver array (VRA) at 4 km range is displayed in Fig. 1. The delay spread is about 80 ms resulting in an intersymbol interference (ISI) of 40 symbols when the symbol rate is $R=1/T=500$ Hz. The complexity of the channel is beneficial for time reversal communications.

We begin with analysis of the at-sea experimental data. The experimental setup is the same as the one described in Ref. 1 for the stationary source case and we briefly review the basic signaling scheme. The source level was 179 dB *re* 1 μPa and the shaping pulse signal was a 15 ms, 2.5–4.5 kHz chirp with a Hanning window, resulting in an effective 100 ms, 3–4 kHz chirp. The duration of the chirp after compression (matched filtering) is $T=2$ ms equivalent to the symbol period. Since the signal bandwidth is $B=1$ kHz, the excess bandwidth[4] is 100% and the roll-off factor $\alpha=1$.

The resulting performance of time reversal communications from the data using quadrature phase-shift keying (QPSK) modulation is illustrated in Fig. 2: (1) TR (▲) and (2) TR+EQ (●). The M elements are selected sequentially from the bottom. A fractionally spaced decision-feedback equalizer (DFE) with tap spacing of $(1/4)T$ is applied to

[a]Author to whom correspondence should be addressed. Electronic mail: hcsong@mpl.ucsd.edu
[b]Work done while at MPL.

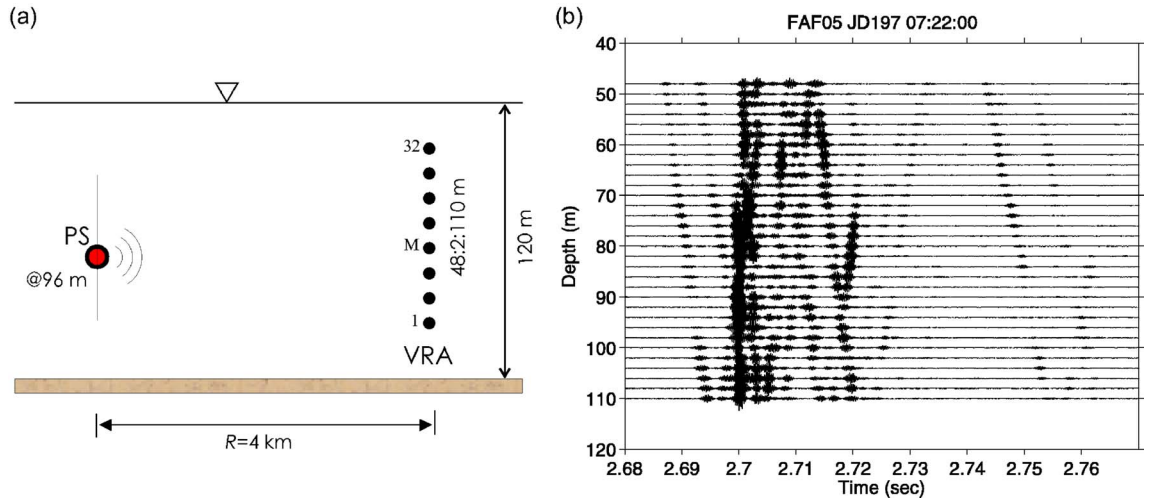


FIG. 1. (Color online) (a) Schematic of passive time reversal communication. (b) The channel impulse responses (CIR) observed by a 32-element vertical receiver array (VRA) from a probe source (PS) at 96 m depth at 4 km range in 120 m deep water.

approach (2). The number of taps for the feedforward and feedback portions of the DFE are 20 and 10, respectively. Overall, approach (2) outperforms approach (1) by 3–5 dB, consistent with the results reported in Ref. 1. It also is interesting to observe that the performance of approach (2) (●) follows closely the characteristics of the input SNR (*) in Fig. 2.

Now the theoretical performance bounds are evaluated from the analytic expressions developed in Ref. 2: Eq. (33) for TR (△) and Eq. (46) for TR+EQ (○). The numerical computation involves three components: (i) the CIR of Fig. 1(b) represented by $\gamma(f)$ defined in Eq. (12) of Ref. 2, (ii) a raised cosine spectrum[4] $X(f)$ with an excess bandwidth of $\alpha=1$, and (iii) $E/N_0=37$ dB which is described in the Appendix. The resulting theoretical performance bounds are superimposed in Fig. 2 for comparison purposes. Two important observations can be made. First, approach (1) shows a good agreement between the theory and data. Second, in contrast with approach (1), the performance of data using approach (2) is below the theory by approximately 3–5 dB.

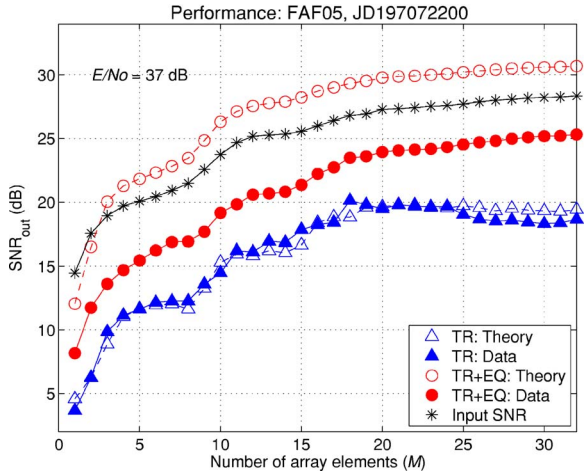


FIG. 2. (Color online) Performance comparison between data (filled) and theory (hollow) versus the number of receiver elements M: (1) time reversal alone (TR, triangles) and (2) time reversal combined with equalization (TR+EQ, circles). The receiver elements are selected from the bottom. The input SNR (*) is superimposed as a reference.

This is not surprising because the theory assumes perfect knowledge of the channel, infinite number of linear equalizer taps to remove the ISI, and no need for phase tracking. For example, in the presence of noise, too many taps effectively increase the noise level so that a minimal number of taps are used in practice.[4] Taking these into consideration, the theoretical performance can provide a useful upper bound for predicting performance of time reversal communications.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-06-1-0128. In addition, SeaMoon Kim would like to thank the Korea Research Foundation for support under Grant KRF-2005-214-D00402 with funding provided by the Korean Government (MOEHRD).

APPENDIX: ESTIMATION OF E/N_0 FOR THEORETICAL PERFORMANCE BOUND

The transmitted energy per symbol E is given by

$$E=\sigma_s^2 T, \tag{A1}$$

where σ_s^2 is the variance of the transmitted signal and T is the symbol interval. Assuming that the noise at the receiver is white with a power spectral density of N_0, the noise variance σ_n^2 is given by

$$\sigma_n^2=N_0 B, \tag{A2}$$

where B is the signal bandwidth. Note that the variance of signal (σ_s^2) and noise (σ_n^2) can be estimated directly from the transmitted signal and received data, respectively. Finally, E/N_0 is obtained from the above two equations such that

$$E/N_0=(BT)\sigma_s^2/\sigma_n^2=2(\sigma_s^2/\sigma_n^2)\simeq 37\ \text{dB}, \tag{A3}$$

where $B=1$ kHz and $T=2$ ms.

[1]H. Song, W. Hodgkiss, W. Kuperman, W. Higley, K. Raghukumar, T. Akal, and M. Stevenson, "Spatial diversity in passive time reversal communications," J. Acoust. Soc. Am. **120**, 2067–2076 (2006).
[2]M. Stojanovic, "Retrofocusing techniques for high rate acoustic communications," J. Acoust. Soc. Am. **117**, 1173–1185 (2005).
[3]H. C. Song and S. Kim, "Retrofocusing techniques in a waveguide for acoustic communications," J. Acoust. Soc. Am. **121**, 3277–3279 (2007).
[4]J. Proakis, *Digital Communications* (McGraw-Hill, New York, 2001).

Flat-response sound source technique for using the two-microphone method in an impedance tube (L)

Xiaodong Jing[a] and Xiaofeng Sun
School of Jet Propulsion, Beijing University of Aeronautics and Astronautics, Beijing 100083, P. R. China

(Received 4 January 2007; revised 5 August 2007; accepted 7 August 2007)

In this letter, to improve the acoustic impedance measurement in a tube using the two-microphone method, a sound source technique is developed that is capable of generating incident sound waves of uniform intensity over the entire frequency range of interest. The basic principle is to use a digitally synthesized compensation input signal to achieve a flat response of the sound source provided by one or more loudspeakers. The present sound source technique has been demonstrated to be very effective and easily implemented with the current impedance tube testing systems.
 [DOI: 10.1121/1.2781581]

PACS number(s): 43.58.Bh, 43.55.Ev, 43.20.Mv [AJZ] Pages: 2519–2521

I. INTRODUCTION

Being an accurate and efficient way to measure the acoustic impedance of sound-absorbing materials or structures, the impedance tube test by using the two-microphone method has found wide application in both acoustics research and engineering. Over the past decades, in order to improve or refine this method, many investigations have been conducted regarding various aspects such as error analysis,[1] microphone mismatching,[2] damping or flow effects,[3,4] etc. The present work attempts to make a contribution by considering the influence of the nonflatness of the sound source response and thus finding a solution to this problem.

The frequency response of the sound source in an impedance tube should be as flat as possible for obvious reasons. First, if in some situations considerably low-amplitude sound pressures occur in the nonuniform spectrum of the sound source, this results in the loss of signal-to-noise ratio and a corresponding decrease of the accuracy in the measurement. Furthermore, it is quite reasonable to require that the acoustic impedance of a sample be measured as a function of frequency when all the other conditions are kept unchanged. As for this concern, the intensity of the incident sound wave is an importance influential factor. This is especially true for sound-absorbing structures that are sensitive to the incident sound intensity, such as the widely used Helmholtz resonator or perforated liners of resonant type.[5] However, as a result of either the inherent nonuniform loudspeaker response or the attenuation effect of the absorptive materials inserted into the tube for resonance modification,[6] it is common that the intensity of the incident sound wave varies, and sometimes considerably, over the frequency range of interest. An equalizer can be used to compensate for the decrease of the sound intensity at high frequencies, but the method becomes inadequate if one wants to obtain a uniform incident sound spectrum. With the sound source technique commonly used in an impedance tube test, it is very difficult to establish the one-dimensional sound field of uniform incident sound intensity that is needed for accurately measuring the acoustic impedance by the two-microphone method.

In this letter, a novel sound source technique is developed that is capable of generating incident sound wave of uniform intensity over the entire frequency range of interest in an impedance tube test. The basic principle is to use a digitally synthesized predistorted input signal to achieve a flat response of the loudspeaker sound source. Similar methods have been successfully used to produce desired short sound pulses in the time-domain measurements.[7] Here, an extension is made to apply the method to a harmonic measurement in an impedance tube. The digital control method and the experimental results are presented in the following sections.

II. THE BASIC PRINCIPLES

As shown in Fig. 1, a one-dimensional sound field is established inside the impedance tube when the sound source emits sound waves at frequencies lower than the cutoff frequency of the higher-order modes of the waveguide. Two microphones are used to measure the sound pressures at two axial positions separated by a distance of s, such that the one-dimensional sound field can be decomposed into incident and reflected waves. The origin of the x coordinate is taken to be at the surface of the test sample to facilitate the analysis. The harmonic time dependence $e^{i\omega t}$ is assumed, where t is time, ω is angular frequency, and $i=\sqrt{-1}$. If the damping effect due to the air viscosity is ignored, the sound pressure of the one-dimensional sound field is given by

$$p(x) = p_i e^{-ikx} + p_r e^{ikx}, \tag{1}$$

where k is the wave number, and p_i and p_r are the amplitudes of the incident and reflected sound waves, respectively. Therefore, the sound pressures at the positions of the two microphones are

$$p_1 = p_i e^{-ikx_1} + p_r e^{ikx_1}, \tag{2a}$$

$$p_2 = p_i e^{-ikx_2} + p_r e^{ikx_2}. \tag{2b}$$

It is straightforward to obtain p_i from Eqs. (2a) and (2b),

[a]Electronic mail: jingxd@buaa.edu.cn

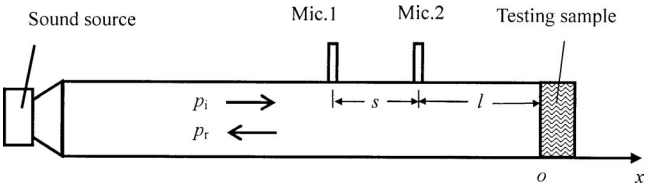


FIG. 1. Illustration of measuring acoustic impedance in a tube by the two-microphone method.

$$p_i = p_1 e^{-ik(l+s)} \frac{1 - H_{12} e^{-iks}}{1 - e^{-2iks}}, \tag{3}$$

where $H_{12}=p_2/p_1$ is the transfer function between the sound pressures measured by the two microphones. Assuming that the electrical signal driving the sound source is $z(f)$, the transfer function of the sound source connected to the impedance tube, H_s, can be computed from

$$H_s(f) = \frac{p_i(f)}{z(f)}, \tag{4}$$

where f is the frequency. Thus, in order to produce incident sound of uniform intensity, i.e., constant $|p_i(f)|$ over the frequency range of interest, the input into the sound source should be compensated by the following sythesized gain:

$$g(f) = \frac{1}{|H_s(f)|}. \tag{5}$$

It can be noted that, avoiding the usage of the phase of p_1 that is yet to be determined, Eq. (5) can accurately compute the sythesized input gain from the absolute value of p_1, the transfer function H_{12}, and the microphone locations. In certain circumstances, the digital compensation cannot be properly done if $|H_s(f)|$ is too small and thus $g(f)$ has very large values at some frequencies. For these outlier frequencies, the gain is clipped to a limit value $g_{\max}$ that is proportional to the average gain value over all the frequencies.

III. THE EXPERIMENT

The experiment is carried out in an impedance tube similar to the small tube setup of the B&K model 4206 waveguide.[8] A 29 mm small tube is connected to a loudspeaker sound source through a 100 mm large tube as a transition. The frequency coverage of the setup is from 500 to 6000 Hz, limited by the spacing of the two microphones and the diameter of the small tube for the lower and upper frequencies, respectively. A multitone source excitation is used that is composed of linearly increasing frequencies at 50 Hz intervals over the range of interest. In the experiment, the magnitude of the signal at the first microphone and the transfer function between the two microphones are measured, and are used to determine the magnitude of the incident sound wave from Eq. (3). The transfer function H_s of the sound source is then obtained from the analysis in Eq. (4). Finally, the input gain is synthesized as indicated in Eq. (5) and is used to compensate the sound source to emit incident sound waves of uniform intensity.

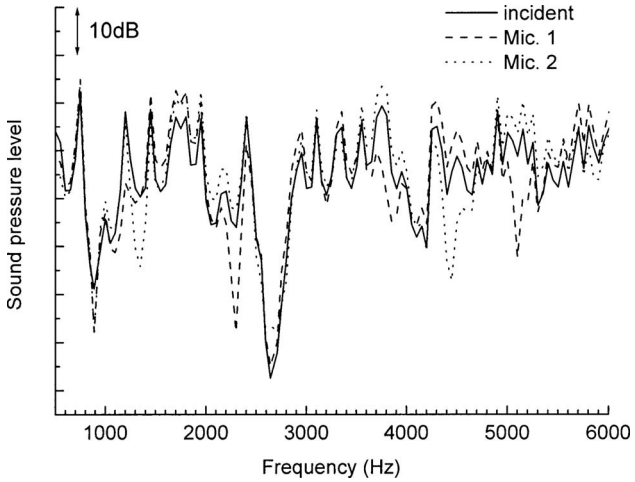


FIG. 2. The spectra of the incident sound and the measured signals at each microphone for a normal (no source compensation) measurement.

IV. RESULTS AND DISCUSSION

In a normal measurement without source compensation, as shown in Fig. 2, the sound pressure spectra measured by the two microphones are significantly nonuniform. The corresponding incident sound spectrum, as described by Eq. (3), is also nonuniform, with the sound pressure level difference between the lowest and highest points being approximately 60 dB. Two narrow regions of extremely low SPL occur near 900 and 2600 Hz, either due to the nonuniform response of the loudspeaker sound source or the resonance of the whole source-tube system.

To remedy this unfavorable situation, we use the synthesized gain to compensate for the response of the source-tube system, and the result computed from Eq. (5) is presented in Fig. 3. As expected, the gain function has large values near 900 and 2600 Hz. However, it is difficult to effectively realize the source compensation if the gain values are too large, so an upper limit is set to be five times of the average gain values over all the frequencies, and it can be seen that the gain curve is clipped to this limit at about 2600 Hz. As shown in the source-compensated measurement in Fig. 4, the sound pressure of the incident sound wave is kept at the

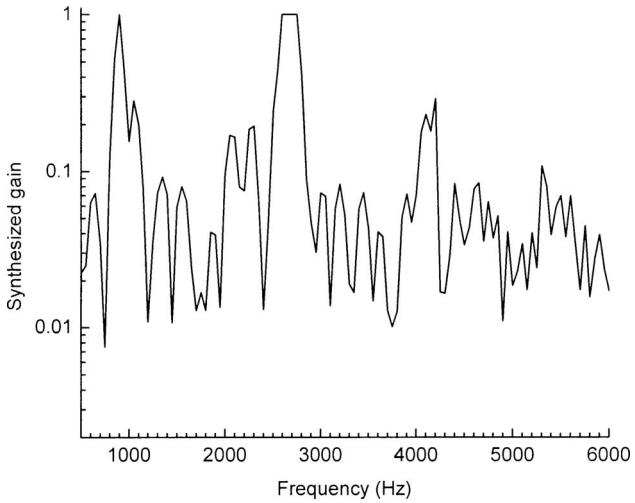


FIG. 3. The synthesized source gain as a function of frequency.

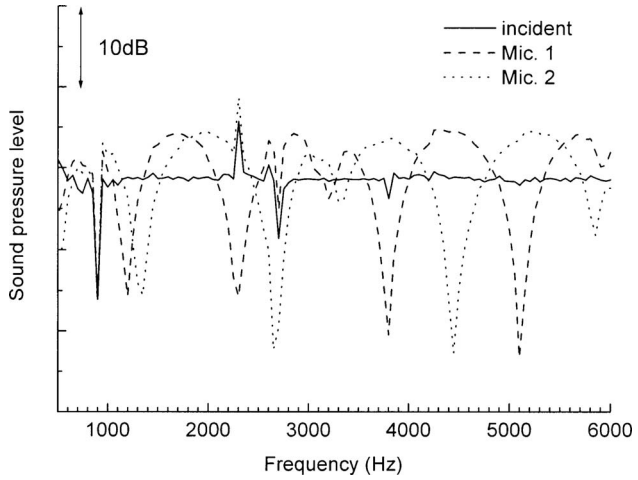


FIG. 4. The spectra of the incident sound and the measured signals at each microphone for a source-compensated measurement.

same level within a small error of about 2 dB except at three discrete frequency points, 900, 2300, and 2600 Hz; and the difference between the highest and the lowest SPL values is considerably reduced over the entire frequency range. As a result, the regular standing wave pattern can be clearly seen from the measurements of the two microphones inside the tube. Thus it is demonstrated that the present synthesized compensation method is successfully applied to achieve a flat response of the sound source in the impedance tube measurement.

To further demonstrate the usefulness of the present method, the acoustic impedance of a perforated liner sample is measured when the source is compensated, and the result is compared with that obtained in a normal measurement in Fig. 5. The theoretical prediction of Guess's model[9] is also presented for comparison purposes. The geometrical parameters of the sample are: The orifice diameter is 2 mm, the plate thickness is 0.5 mm, the open area ratio is 5%, and the cavity depth is 60 mm. Being mainly reactive, the acoustic liner with large orifices has a resistance that is too small to be reliably measured. Thus, only the acoustic reactance of the sample, both measured and computed, is presented. As

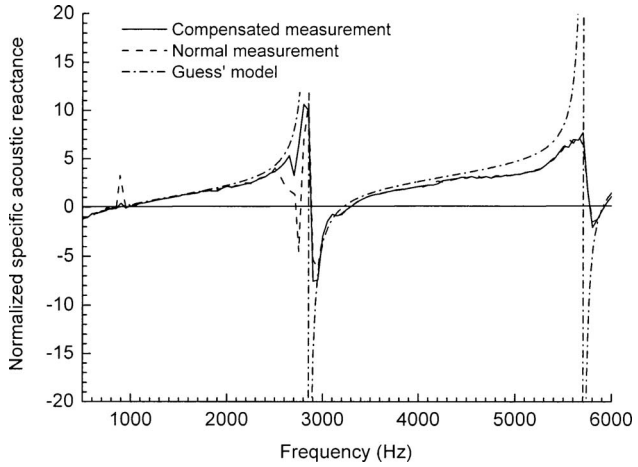


FIG. 5. Comparison among the source-compensated measurement, the normal measurement, and the prediction of Guess' model for the acoustic reactance of the perforated liner sample.

shown in Fig. 5, the reactances obtained from the normal and compensated measurements are nearly identical above 3000 Hz, but tend to have values less than those predicted with Guess's model. Below 3000 Hz, the normal measurement markedly deviates from the theoretical prediction near the frequencies of 900 and 2600 Hz; in contrast, measurements with the compensated source generally show good agreement with Guess's model. Thus, by means of the source-compensation method, the improvement of the measurement accuracy is obvious at frequencies of low source response.

V. CONCLUDING REMARKS

In this letter, a flat-response sound source technique has been developed for application in an impedance tube test using the two-microphone method. With this technique, it is possible to generate incident sound wave of uniform intensity over the entire frequency range of interest. The basic principle is that a digitally synthesized input gain function is used to compensate the response of the sound source. A demonstration experiment has been conducted in which the sound pressure of the incident sound wave is maintained at a constant level within a small error of about 2 dB, except at limited frequency points. It has also been shown that the present method can improve the measurement accuracy when the response of the sound source is considerably low. Besides, the present method has the potential benefit of maintaining the desired source condition for measurements with samples that are nonlinear, i.e., dependent on incident sound intensity. Being easily and conveniently implemented without the need for extra equipment, it is suggested that the present technique be incorporated into impedance tube testing systems to improve the measurement accuracy.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support for this work from the National Natural Science Foundation of China (10302003, 10472009) and the Fok Ying Tung Educational Foundation (91052).

[1]F. Seybert and B. Soenarko, "Error analysis of spectral estimates with application to the measurement of acoustic parameters using random sound fields in ducts," J. Acoust. Soc. Am. **69**, 1190–1198 (1981).
[2]J. Kruger and M. Quickert, "Determination of acoustic absorber parameters in impedance tubes," Appl. Acoust. **50**, 79–89 (1997).
[3]W. T. Chu, "Extension of the two-microphone transfer function method for impedance tube measurements," J. Acoust. Soc. Am. **80**, 347–348 (1986).
[4]M. L. Munjal and A. G. Doige, "The two-microphone method incorporating the effects of mean flow and acoustic damping," J. Sound Vib. **137**, 135–138 (1990).
[5]U. Ingard and S. Labate, "Acoustic circulation effects and the nonlinear impedance of orifices," J. Acoust. Soc. Am. **22**, 211–218 (1950).
[6]ASTM, "Standard test method for impedance and absorption of acoustical materials using a tube, two microphones and a digital frequency analysis system," *Annual Book of ASTM Standard* (ASTM, Philadelphia, 1998), **E 1050–90**.
[7]X. Jing and K.-Y. Fung, "Generation of desired impulses," J. Sound Vib. **297**, 616–626 (2006).
[8]Bruel & Kjaer Corporation, "Product data: Impedance tube kit (50 Hz –6.4 kHz)–Type 4206," http://www.bksv.com/pdf/Bp1039.pdf. Last viewed 9/27/07.
[9]W. Guess, "Calculation of perforated plate liner parameters from specified acoustic resistance and reactance," J. Sound Vib. **40**, 119–137 (1975).

Estimation of tissue's elasticity with surface wave speed (L)

Xiaoming Zhang[a] and James F. Greenleaf
Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, Minnesota 55905

(Received 30 April 2007; revised 23 August 2007; accepted 27 August 2007)

The mechanical response of tissues to external forces has gained considerable interest in medical diagnosis. One approach to imaging tissue elastic properties is to apply an external force on the surface of the body. Another approach is to generate a localized force inside the tissue with the radiation force of ultrasound. In this paper, a new method is developed to estimate tissue's elasticity based on surface wave speed measurement. The theory of surface wave speed is developed for estimating tissue's elasticity. Experiments are carried out on a tissue-mimicking ultrasound phantom. An amplitude modulated ultrasound signal of a few hundred hertz is used to generate a localized force in the phantom. The surface wave fields are measured with a laser vibrometer. It shows that the surface wave speed can be used for inversely estimating tissue's elastic properties based on tissue's surface measurement. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2785045]

PACS number(s): 43.80.Ev [CCC] Pages: 2522–2525

I. INTRODUCTION

The mechanical response of tissues to external forces has gained considerable interest in medical diagnosis. One approach to imaging tissue elastic properties is to apply an external force on the surface of the body.[1–3] Another emerging field is to measure the mechanical properties of tissue with noninvasive ultrasound radiation force inside the tissue. Examples of newly developed imaging modalities include shear wave elasticity imaging,[4] supersonic shear imaging,[5] acoustic radiation force impulse imaging,[6] and vibro-acoustography.[7] In these new imaging modalities, the shear wave is used to estimate elastic properties of tissue. In vessels, the bending wave method has been developed for noninvasively estimating elastic properties of arteries using the radiation force of ultrasound.[8] The resonance can also be used in arteries for quantitative measurement of the local elastic modulus of arteries.[9,10] A complete theory of surface wave has been developed for calculating the field and radiation impedance of mechanical radiators on the surface of a semi-infinite isotropic solid.[11] An analytic solution of surface wave in a half-space viscoelastic medium is obtained for application in medical diagnosis.[12] The radiation impedance of a finite circular piston on a half-space viscoelastic medium is investigated and verified with experiments.[13] Experiments of nonlinear surface elastic waves in a biological tissue are carried out.[14] Vibration in the tissue is generated by a vibrator and measured by an accelerometer within the frequency range of 3–400 Hz. A theory of nonlinear surface waves in soft, weakly compressible elastic medium is developed.[15] It is found that the quadratic nonlinearity of surface waves in weakly compressible elastic medium is due to the geometric nonlinearity.

In this letter, the surface Rayleigh wave and its propagation speed on the tissue's surface are investigated for estimating the tissue's elastic properties. To generate the surface wave, a localized radiation force of ultrasound is remotely and noninvasively applied on or close to the body's surface. This force can generate both a shear wave as well as a compression wave inside tissue. However, only the surface wave can propagate on the body's surface. Theoretical and experimental analyses of surface wave speed as well as its estimation of tissue's elasticity are presented. This new method shows the possibility of estimating the tissue's elastic properties from the surface wave measurement.

II. THEORY

We consider the problem of wave propagation on the surface of tissue. In our method, the surface wave is generated noninvasively by the radiation force of ultrasound. This problem can be analyzed as the wave propagation in a semi-infinite viscoelastic medium under harmonic excitation by a uniformly distributed force on a circular surface area. The equation of wave propagation in an isotropic and elastic medium is[11,12]

$$(\lambda + 2\mu)\,\nabla\nabla\cdot\boldsymbol{u} - \mu\,\nabla\times\nabla\times\boldsymbol{u} = \rho\frac{\partial^2\boldsymbol{u}}{\partial t^2}, \tag{1}$$

where $\boldsymbol{u}$ is the displacement, ρ is the density of the medium, and λ and μ are, respectively, the Lams constants of the medium. In this letter, we consider inverse estimating of the elasticity of tissue from surface wave speed measurements. The viscous properties of the tissue can be measured by the decay of surface wave amplitude with distance, but is not discussed.

The problem is solved in the cylindrical polar coordinate system (Fig. 1). Consider a harmonic force excitation with a uniform stress on the surface of the medium in the circular region of $r \leq a$. Miller and Pursey[11] derive the following in-

[a]Author to whom correspondence should be addressed. Electronic mail: zhang.xiaoming@mayo.edu

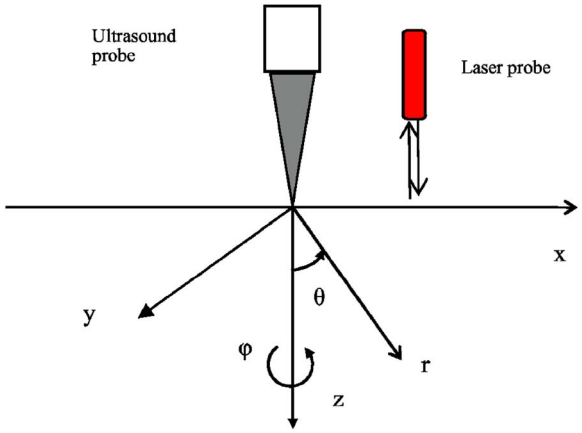


FIG. 1. (Color online) Ultrasound based surface wave model for measuring elastic properties of tissue. The surface wave is generated noninvasively by the piston-like localized radiation force of ultrasound. The surface wave is measured noncontact with a laser vibrometer.

tegral expression for the displacement field in the r and z directions at any location in and on the surface of the medium,

$$u_z' = \frac{a'}{\mu}\int_0^\infty \frac{\sqrt{(\xi'^2-1)}J_1(\xi' a')}{F_0(\xi')}\{2\xi'^2 e^{-z'\sqrt{(\xi'^2-\eta^2)}} + (\eta^2 - 2\xi'^2)e^{-z'\sqrt{(\xi'^2-1)}}\}J_0(\xi' r')d\xi'$$

$$u_r' = \frac{a'}{\mu}\int_0^\infty \frac{\xi' J_1(\xi' a')}{F_0(\xi')}\{2\sqrt{(\xi'^2-1)}\sqrt{(\xi'^2-\eta^2)} \times e^{-z'\sqrt{(\xi'^2-\eta^2)}} + (\eta^2 - 2\xi'^2)e^{-z'\sqrt{(\xi'^2-1)}}\}J_1(\xi' r')d\xi'. \quad (2)$$

For the convenience of evaluating Eq. (2), the results are normalized by measuring lengths in units of $1/k_1$, so that $\xi' = \xi/k_1$, $a' = ak_1$, $z' = zk_1$, $r' = rk_1$, where ξ is the integration parameter in the wave number domain. The divisor function of the integration functions is $F_0(\xi') = (2\xi'^2 - \eta^2)^2 - 4\xi'^2\sqrt{(\xi'^2-1)}\sqrt{(\xi'^2-\eta^2)}$, where $\eta = k_2/k_1 = \sqrt{\{2(1-\sigma)/(1-2\sigma)\}}$, $k_1 = \omega\sqrt{\rho/(\lambda+2\mu)}$, $k_2 = \omega\sqrt{\rho/\mu}$, σ being Poisson's ratio for the medium, k_1 and k_2 denote the wave numbers for compression and shear wave propagation, respectively, and J_0 and J_1 refer to Bessel functions of the first kind.

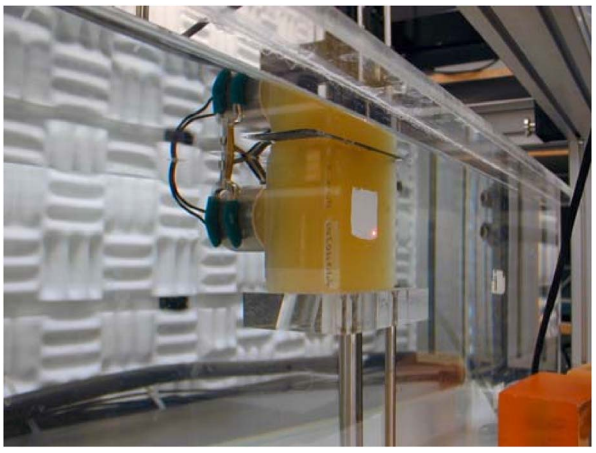


FIG. 2. (Color online) Experimental setup for measurement of the surface wave in an ultrasound phantom. The ultrasound transducer is behind the phantom which generates a force on the front surface of the phantom. The surface vibration on the front surface with paint is measured with a laser vibrometer.

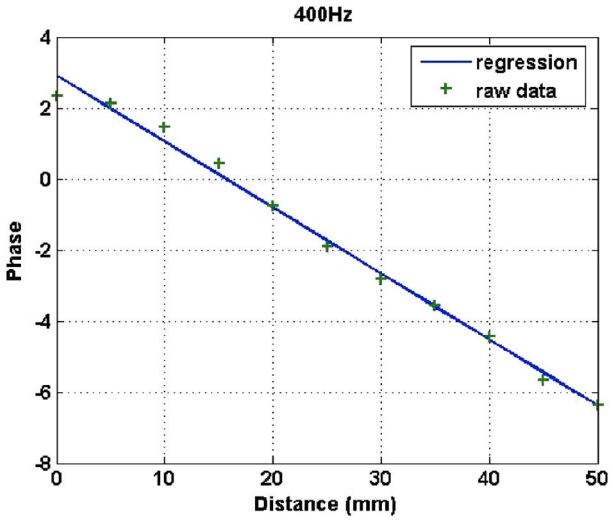


FIG. 3. (Color online) Regression equation is $\Delta\widehat{\phi} = -185.9803\Delta r + 2.9069$ with a 95% confidence interval and the wave speed is $c_s = 13.5137$ at 400 Hz.

The wave displacement field can be solved with Eq. (2); however, we develop the inverse method of using the surface wave speed to estimate the elastic modulus of tissue. Our objective is to develop the relation between the surface wave speed and the tissue's elasticity. The surface wave speed can be solved by letting the divisor function $F_0(\xi') = 0$, which results in

$$(2\xi'^2 - \eta^2)^2 - 4\xi'^2\sqrt{(\xi'^2-1)}\sqrt{(\xi'^2-\eta^2)} = 0. \quad (3)$$

This equation is a bi-fourth equation of ξ', however, only the solution for which ξ' is real, $\xi' > \eta$ and $\xi' > 1$, is the right solution for the surface wave. We solve the surface wave for different Poisson's ratios representing materials from metals to soft tissues.

III. EXPERIMENT

Experiments are carried out on a commercial tissue mimicking ultrasound phantom (ATS Laboratories, Bridgeport, CT 06608). The phantom under test is a rubber-based material. The usable life of the phantom is greater than 7–10 years. The phantom's dimensions are $10 \times 10 \times 7$ cm^3 with different density of scatters. An amplitude modulated ultrasound tone burst of a few hundred hertz is used to generate a localized force on the surface of the phantom. The surface wave fields are measured with a laser vibrometer. The experimental setup is shown in Fig. 1.

The phantom is placed in a water tank (Fig. 2). A 3 MHz confocal transducer of 10 cm focal length produces the radiation force remotely on the surface of the phantom. The focal size of ultrasound beam is about 0.7 mm in diameter, which generates almost a point force on the phantom. To measure the surface wave of the phantom, the measured surface area is painted for laser detection of surface vibration.

TABLE I. Wave number ratios among-shear wave, surface wave, and compression wave for different materials.

Poisson's ratio	k_2/k_1	k_s/k_1
0.25	1.73	1.88
0.33	2.00	2.14
0.40	2.45	2.60
0.45	3.32	3.50
0.49	7.14	7.49
0.499	22.38	23.43

IV. RESULTS AND DISCUSSION

The surface wave speed in a semi-infinite medium can be solved from Eq. (3) for different material properties. The Poisson's ratios are studied from 0.25 to 0.499, which represent materials from metal to soft tissues. It is quite interesting to establish the relationship between the surface wave speed to shear wave speed and compression wave speed. Table I shows the ratios of wave numbers between the shear wave and compression wave k_2/k_1 and that of the surface wave to compression wave k_s/k_1 for different materials, where k_s denotes the wave number of surface wave. It shows that the surface wave number is largest and the compression wave number is the smallest. The three wave numbers are well separated in the wave number domain. The surface wave number is quite close to the shear wave number. However, both the surface wave number and shear wave number are much larger than the compression wave number. As the Poisson's ratios increase from 0.25 to 0.499, the ratio between the shear wave number and the compression wave number increase significantly from 1.88 to 23.43. It is well known that the compression wave travels much faster than the shear wave in soft tissues. The wavelength of compression wave is so large in tissues that the compression wave is not used in elastic imaging of tissues. However, a much smaller wavelength of the shear wave than the compression wave makes the shear wave widely used for elastic imaging of tissues.

The ratio of the shear wave speed to surface wave speed c_2/c_s is shown in Table II for different materials, where c_2 and c_s denote, respectively, the speed of shear and surface waves. It shows that the shear wave is a little faster than the surface wave, but the ratio between the shear wave speed to surface wave speed is almost constant over a quite large range of material. The speed ratio is 1.09 for metal material of Poisson's ratio 0.25. The speed ratio reduces a little to 1.05 for polymer materials and soft tissues with Poisson's ratio between 0.45 and 0.499. Therefore, the surface wave speed can be related to the elastic modulus of tissue as

$$c_s = \frac{1}{1.05}\sqrt{\frac{E}{3\rho}}, \tag{4}$$

where E is Young's modulus of tissue material.

TABLE II. Wave velocity ratio between the shear wave and surface wave for different materials.

Poisson's ratio	c_2/c_s
0.25	1.09
0.33	1.07
0.40	1.06
0.45	1.05
0.49	1.05
0.499	1.05

In this letter, the surface wave on the tissue's surface is generated and measured. The surface wave can be generated by either a contact or noncontact method. For example, the surface wave on the skin can be generated by a mechanical shaker or an ultrasound transducer. Noncontact surface wave generation can be made with an ultrasound transducer through degassed water. The surface wave can be measured contact using accelerators or noncontact by optical or ultrasound probes. In our experiments, the surface wave is generated noncontact by the radiation force of ultrasound. The surface wave is measured noncontact by a Polytec VibraScan Laser Vibrometer system.

The objective is to measure the surface wave speed from which the elasticity of tissue is inversely estimated. The viscous property of tissue can be measured from the decay of surface wave amplitude with distance,[8] but is not discussed in this letter. The surface wave propagates symmetrically from the piston-like ultrasound excitation force, therefore, the phase of the surface wave changes symmetrically with the radial distance, except in a small area of the source near field. The phase change of surface wave with the radial distance is used to measure the surface wave speed.[8] The wave number of surface wave propagation in the radial direction is $k_s = \omega/c_s$, where c_s is the surface wave speed. The derivative of the surface wave phase with respect to radial distance r is measured and the surface wave speed is solved by

$$c_s = 2\pi f|\Delta r/\Delta\phi|, \tag{5}$$

where Δr is the radial distance of two measuring positions, $\Delta\phi$ is the wave phase change over distance Δr in radius, and f is the frequency in hertz.

The phase change over a given distance can be used to calculate the wave speed. To improve accuracy, we measured the phase change in 5 mm increments over a long distance of 50 mm. Figure 3 shows the wave phase change with radial distance, where phase is measured relative to the excitation signal. The excitation signal is a 100 Hz amplitude modulated ultrasound tone burst. We can see from Fig. 3 that the slope of phase with distance is well behaved, which suggests that an accurate and reliable wave speed can be obtained.

A statistical model is developed to measure the wave speed that uses all the data of the phase change with distance. The regression of the phase change $\Delta\phi$ and the distance Δr is analyzed. The regression curve can be obtained by "best fitting" a linear relationship of the phase change $\Delta\phi$ and the distance Δr in the least mean squares (LMS) sense and its equation is

$$\widehat{\Delta\phi} = -a\Delta r + \beta, \tag{6}$$

where $\widehat{\Delta\phi}$ denotes the value of $\Delta\phi$ on the regression for a given distance of Δr.

The surface wave velocity can be estimated by

$$c_s = 2\pi f|\Delta r/\Delta\widehat{\phi}| = 2\pi f/\alpha, \quad (7)$$

where c_s is the estimation of wave speed from the regression analysis that should be better than just using two position measurement.

The regression analysis for the 400 Hz tone burst wave in the phantom is shown in Fig. 3. The regression equation, with the LMS of a 95% confidence interval, is $\Delta\widehat{\phi} = -185.9803\Delta r + 2.9069$ and the estimation of the wave speed is $c_s = 13.5137$ m/s at 400 Hz. Regression analyses for the 100, 200, and 300 Hz waves provide the wave speed of 12.2315, 12.7360, and 13.0641 m/s, respectively. We can see some wave velocity dispersion with frequency, but it is not as big as the bending wave in vessels.[8] The average surface wave speed is 12.8863 m/s for the wave speed between 100 and 400 Hz. The Young's modulus of phantom can be estimated from the surface wave speed by

$$E = 3\rho(1.05c_s)^2. \quad (8)$$

With a surface wave speed of 12.8863 m/s, the Young's modulus is estimated to be $E = 549$ kPa. In this calculation, the mass density is assumed $\rho = 1000$ kg/m^3. This commercial ultrasound phantom is a rubber-based material. Unfortunately, the company does not provide the elastic properties of the phantom. However, the estimation of $E = 549$ kPa is well between arterial tissues, which are in the range of a few hundred kilopascal[8,10,16] and rubber materials, which are typically in the range of megapascal.[17] Independent tests of the phantom's Young's modulus can be made later with a servo-hydraulic testing machine (Instron Corp., Canton, MA).

There are some advantages of the present method. One advantage of using wave propagation is that the equations are generally local and do not require boundary conditions if the wave reflections can be ignored or controlled. In addition, the wave equations are amenable to applying inverse solutions, that is, finding the material properties given the wave characteristics such as wave speed and dispersion. In our method, we can control wave propagation, such as the frequency of wave, so we can measure the local tissue material properties with high spatial and temporal accuracy. For example, the surface wave speed can be measured in 5 mm, and therefore, allows us to measure the local tissue's elasticity. Another advantage is that we use the surface wave speed not the wave fields (such as displacements) to inversely estimate the elasticity of tissue. The simple relation of the surface wave speed and the tissue's elasticity is developed. One obvious advantage is that the surface wave speed can be measured on the surface of tissue, which is much easier than the measurement of shear wave speed inside the tissue. This is especially useful to evaluate surface tissue such as skin. However, we develop the relation between the surface wave speed and the shear wave speed, which may be useful to evaluate tissues near the surface or the skin.

V. CONCLUSION

In this letter, the surface wave speed is proposed for estimation of tissue's elasticity. The surface wave of tissue is generated noninvasively by localized radiation force of ultrasound. The surface wave of tissue is measured noncontact by a laser vibrometer. This method is noninvasive for both surface wave generation and detection. It has the advantage of measurement on the surface of tissue. The theory shows that the surface wave speed is quite close to the shear wave speed for a large range of materials from metals to soft tissues. The shear wave is a little faster than the surface wave. The ratio of the shear wave speed to the surface wave speed is 1.05 for soft tissues. Experiments are carried out on a commercial tissue-mimicking ultrasound phantom. The surface wave speed of the phantom is measured with a statistical model from which the Young's modulus is estimated. The surface wave speed method might become a medical tool in evaluating the elasticity of surface tissues such as skin.

ACKNOWLEDGMENTS

The authors thank Randall R. Kinnick for experimental support and Jennifer L. Milliken for secretarial assistance.

[1] R. M. Lerner, S. R. Huang, and K. J. Parker, "Sonoelasticity images derived from ultrasound signals in mechanically vibrated tissues," Ultrasound Med. Biol. **16**, 231–239. (1990).
[2] J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: A quantitative method for imaging the elasticity of boilogical tissues," Ultrason. Imaging **13**, 111–134 (1991).
[3] T. J. Hall, Y. Zhu, and C. S. Spalding, "In vivo real-time freehand palpation imaging," Ultrasound Med. Biol. **29**, 427–435 (2003).
[4] A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).
[5] J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 396–409 (2004).
[6] K. Nightingale, S. McAleavey, and G. Trahey, "Shear wave generation using acoustic radiation force: In vivo and ex vivo results," Ultrasound Med. Biol. **29**, 1715–1723 (2003).
[7] M.Fatemi and J. F. Greenleaf, "Ultrasound stimulated vibro-acoustic spectroscopy," Science **280**, 82–85 (1998).
[8] X. Zhang, R. R. Kinnick, M. Fatemi, and J. F. Greenleaf, "Noninvasive method for estimating of complex elastic modulus of arterial vessels," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 642–652 (2005).
[9] X. Zhang, M. Fatemi, R. R. Kinnick, and J. F. Greenleaf, "Noncontact ultrasound stimulated optical vibrometry study of coupled vibration of arterial tubes in fluids," J. Acoust. Soc. Am. **113**, 1249–1257 (2003).
[10] X. Zhang and J. F. Greenleaf, "Estimation of complex arterial elastic modulus from ring resonance excited by ultrasound radiation force," Ultrasonics **44**, e169–e172 (2006).
[11] G. F. Miller and H. Pursey, "The field and radiation impedance of mechanical radiations on the free surface of a semi-infinite isotropic solid," Proc. R. Soc. London, Ser. A **223**, 521–541 (1954).
[12] T. J. Royston, H. A. Mansy, and R. H. Sandler, "Excitation and propagation of surface waves on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **106**, 3678–3686 (1999).
[13] X. Zhang and T. J. Royston, "Radiation impedance of a finite circular piston on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **109**, 795–802 (2001).
[14] B. N. Klochkov, "Nonlinear vibroacoustic processes at the surface of a biological tissue," Acoust. Phys. **46**, 621–623 (2000).
[15] E. A. Zabolotskaya, Y. A. Llinskii, and M. F. Hamilton, "Nonlinear surface waves in soft, weakly compressible elastic media," J. Acoust. Soc. Am. **121**, 1873–1878 (2007).
[16] W. W. Nichols and M. F. O'Rourke, *McDonald's Blood Flow in Arteries*, 3rd ed. (Arnold, London, 1990).
[17] X. Zhang and J. F. Greenleaf, "Generation of a torsion wave and measuring its propagation velocity in the circumferential direction of arterial wall," Ultrasonics **44**, e165–e168 (2006).

Gaussian representation of high-intensity focused ultrasound beams

Joshua E. Soneson[a] and Matthew R. Myers[b]
Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland 20993

(Received 7 November 2006; revised 6 April 2007; accepted 14 August 2007)

A method for fast numerical simulation of high-intensity focused ultrasound beams is derived. The method is based on the frequency-domain representation of the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, and assumes for each harmonic a Gaussian transverse pressure distribution at all distances from the transducer face. The beamwidths of the harmonics are constrained to vary inversely with the square root of the harmonic number, and as such this method may be viewed as an extension of a quasilinear approximation. The technique is capable of determining pressure or intensity fields of moderately nonlinear high-intensity focused ultrasound beams in water or biological tissue, usually requiring less than a minute of computer time on a modern workstation. Moreover, this method is particularly well suited to high-gain simulations since, unlike traditional finite-difference methods, it is not subject to resolution limitations in the transverse direction. Results are shown to be in reasonable agreement with numerical solutions of the full KZK equation in both tissue and water for moderately nonlinear beams. [DOI: 10.1121/1.2783124]

PACS number(s): 43.25.Ts, 43.80.Gx [MFH] Pages: 2526–2531

I. INTRODUCTION

High intensity focused ultrasound (HIFU) has become an important modality in the development of noninvasive surgical procedures. Devices using HIFU to ablate diseased tissue, coagulate deep bleeds, and ameliorate chronic pain have all been proposed. In order to determine the safety and efficacy of these devices, as well as optimize design characteristics, models for simulating HIFU beams need to be developed.

Modeling HIFU beams is challenging for the following three reasons. First, the governing Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation[1] is nonlinear, and analytic solutions are difficult or impossible to derive. Thus computer models offer the only practical means of predicting HIFU beam phenomena. Second, the KZK equation has three independent variables (in a cylindrical geometry), two spatial and one temporal. Even on modern computer architectures, integrating three-dimensional partial differential equations (PDEs) is a time-consuming task. Third, there is great disparity in the characteristic scales of the problem. For example, the beamwidth at the transducer face may be 10 cm, while at the focus the beam is as little as 1 mm wide. Another disparity is present in the characteristic time scales: A typical HIFU sonication occurs on the order of 1 s, while the acoustical cycle is approximately 1 μs, often with thin shock fronts which develop from an initially harmonic wave form. A model which resolves all these scales would be far too complex for any practical use. Therefore it is necessary to use approximation techniques resulting in models which are simple enough for practical purposes yet retain the physics of interest.

To simulate long duration (on the order of 1 s) HIFU sonications, it is more convenient to study the problem in the frequency domain. In this approach, the fast time scale is removed from the KZK equation, resulting in a system of coupled PDEs in two spatial dimensions, one PDE for each harmonic field. This technique works well when higher harmonic generation is inhibited by absorption, so that a relatively small number of PDEs is sufficient to describe the beam. If absorption is weak (such as in water) and/or does not sharply increase with frequency (as is the case in biological tissues), energy flows up the spectrum and very high harmonics are excited. This necessitates simulations which account for hundreds or perhaps even thousands of harmonics. One method for accelerating simulations is to use an algorithm which adaptively determines the number of harmonics required, such as that used by Khokhlova *et al.*[2] Another sophisticated method[3–6] overcomes the need for a very large number of harmonics by matching a relatively small number of computed harmonics with the known Fourier expansion of a sawtooth wave to account for shockwave development. This way, only the first few tens of harmonics are computed explicitly and the rest of the spectrum is extrapolated using known formulas. The work of Curra *et al.*[6] was performed using a model Gaussian transducer with a gain (the ratio of the pressure at the focus to the pressure at the transducer face in lossless linear media) value of $G=20$. Typical HIFU devices for medical applications operate at around $G=50$ or higher. For example, the geometry of the ExAblate 2000 (InSightec, Haifa, Israel) system for ablating uterine fibroids[7] results in $G=62$. Without the use of special techniques, simulating high-gain beams at high power, where stronger nonlinear interaction leads to the generation of very high harmonics, is complicated by the intrinsic disparity in the spatial scales. The problem may be attacked using a very fine computational grid which resolves all the scales, but this

[a]Electronic mail: joshua.soneson@fda.hhs.gov
[b]Electronic mail: matthew.myers@fda.hhs.gov

results in impractically intensive computations. A more sophisticated technique to overcome the disparity in the spatial scales is to employ a nonuniform computational mesh which is fine near the focus to resolve the small scales, and coarse in other regions to speed the computation.[4,8,9] In particular, this approach was used by Hart and Hamilton[9] to study nonlinear effects in media whose absorption depends quadratically on frequency. In that work simulations were performed at gain values up to $G=72.5$, but the strong absorption at higher frequencies limited high-harmonic generation. Many optimizations have been developed to speed up simulations of high-intensity focused beams by introducing convergent spatial coordinates and using asymptotic methods to reduce the number of harmonics in simulations. However, even with these optimizations, the simulations are still time-consuming.

The subject of this paper is the introduction of a fast frequency-domain technique which is capable of approximating HIFU beams at moderate power levels and high gain. This method allows the system of coupled PDEs describing the harmonic fields to be reduced to a system of coupled ordinary differential equations (ODEs), one ODE for each harmonic, which may be rapidly integrated on any modern workstation using standard numerical techniques. This is done by approximating the transverse pressure profile of each harmonic field using a Gaussian function. In actual HIFU systems, transducers with uniform shading or phased arrays are commonly used. However, the Gaussian approximation simplifies the analysis due to the lack of interference effects, but still serves to estimate pressure, intensity, and nonlinear effects in the focal region. Moreover, it automatically satisfies the boundary conditions (1) that the field vanishes as the radial coordinate approaches infinity; and (2) that the transverse beam profile is symmetric about the axis. This property eliminates numerical artifacts associated with artificial boundaries and obviates implementations of absorbing boundary conditions. The algorithm is faster than an analogous PDE method by a factor roughly equal to the number of points required to discretize the PDEs in the radial direction. Therefore, if 1000 points are required for a given PDE simulation, an algorithm using the Gaussian representation runs about 1000 times faster.

The paper is organized as follows: In Sec. II, the simple ODE system is derived from the KZK equation; Sec. III describes the integration method and presents comparisons of the Gaussian representation technique with a traditional finite-difference KZK method; and concluding remarks are made in Sec. IV.

II. MODEL EQUATIONS

In many cases of practical importance, including most HIFU regimes, the processes of sound diffraction, absorption, and nonlinear interaction can be accurately described by the KZK equation,[1]

$$\frac{\partial^2 p}{\partial z \partial t} = \frac{c_0}{2}\nabla_r^2 p + \mathcal{L}p + \frac{\beta}{2\rho_0 c_0^3}\frac{\partial^2 p^2}{\partial t^2}, \tag{1}$$

where p is deviation from the ambient pressure, z and r are the axial and radial coordinates, respectively, $\nabla_r^2 = \partial^2/\partial r^2 + r^{-1}\partial/\partial r$ is the transverse Laplacian, t is time measured in the reference frame propagating at the small-signal sound speed c_0, ρ_0 is the mass density of the propagation medium, $\beta = 1 + B/2A$ is the nonlinear parameter, and $\mathcal{L}$ is a linear operator which accounts for absorption and sound velocity dispersion. This operator is sometimes given by $\mathcal{L} = (\delta/2c_0^3)\partial^3/\partial t^3$ (where δ is the diffusivity of sound), which assumes the absorption depends on frequency quadratically, as is the case for propagation in water. However, in the present analysis the operator is expressed generally to account for propagation in tissue, where absorption depends almost linearly on frequency and is accompanied by sound velocity dispersion.

If the pressure field may be regarded as periodic in time, as is the case for HIFU sonications containing more than a few hundred acoustical cycles, then the pressure may be factored as

$$p(r,z,t) = \frac{1}{2}\sum_{n=1}^{\infty}[A_n(r,z)e^{in\omega t} + A_n^*(r,z)e^{-in\omega t}]. \tag{2}$$

Here ω is the angular frequency, A_n is the complex amplitude function corresponding to the nth harmonic, and the (*) denotes complex conjugation. Integrating Eq. (1) with respect to time t and using the factorization (2) results in the frequency-domain representation of the KZK equation. Rescaling the variables according to $\zeta = z/d$, $\rho = r/a$, and $u_k = A_k/p_0$ (where d and a are the transducer focal length and effective radius, respectively, and p_0 is the peak pressure deviation at the transducer face) yields a dimensionless system of PDEs for the rescaled amplitude functions:

$$\frac{\partial u_n}{\partial \zeta} + \frac{i}{4nG}\nabla_\rho^2 u_n + \gamma_n u_n = \frac{inN}{4}\sum_{m=1}^{\infty} u_m(u_{n-m} + 2u_{m-n}^*), \quad n = 1,2,\ldots . \tag{3}$$

In this expression, $G = z_0/d$ is the linear pressure gain (in terms of the Rayleigh distance $z_0 = \pi a^2 f/c_0$, with transducer frequency f), γ_n is a complex number representing the operator $\mathcal{L}$ in the frequency domain, and $N = 2\pi p_0 \beta d f/\rho_0 c_0^3$ is the coefficient of nonlinearity. In Eq. (3), it is understood that only amplitude functions with positive subscripts (representing positive harmonics) are included in the convolution sum.

The relationship between absorption and frequency may be described using a power law over a large range of frequencies, so that

$$\alpha(f) = \alpha_0(f/f_0)^\eta, \tag{4}$$

where α_0 is the absorption at some reference frequency f_0, usually 1 MHz. The absorption α is proportional to the real part of the γ parameter. To ensure the system satisfies causality, the corresponding imaginary part of γ, which accounts for sound velocity dispersion, must be determined using the Kramers-Kronig relations, or a local approximation thereof.[10] The power-law absorption curve results in the following expression for the γ parameter:

$$\gamma(f) = d\alpha(f) - \frac{2id}{\pi}\begin{cases}(\eta-1)^{-1}[\alpha(f)-\alpha(f_0)f/f_0], & \eta \neq 1\\ \alpha(f)\log(f/f_0), & \eta = 1.\end{cases} \tag{5}$$

Boundary conditions corresponding to a parabolic approximation of a spherically converging, Gaussian-distributed monochromatic wave are given by[11]

$$u_1(\rho,\zeta=0) = \exp[(iG-1)\rho^2], \tag{6a}$$

$$u_n(\rho,\zeta=0) = 0, \quad n=2,3,\ldots, \tag{6b}$$

$$\frac{\partial u_n}{\partial \rho}(\rho=0,\zeta) = 0 = u_n(\rho\to\infty,\zeta), \quad n=1,2,\ldots. \tag{6c}$$

Equations (3) and (6) are sufficient to determine the entire pressure field, and have been used successfully in many investigations of ultrasound beam phenomena. However, to reduce the computation required to integrate the system of PDEs, the harmonic fields are approximated using the following Gaussian representation:

$$u_n(\rho,\zeta) = a_n(\zeta)\exp[-Gnb_n(\zeta)\rho^2], \tag{7}$$

where the Gaussian amplitude and width parameters, a_n and b_n, are complex-valued. This representation automatically satisfies the boundary conditions (6c), and reflects the trend that the width of the nth harmonic field scales like $n^{-1/2}$, which may be observed in a quasilinear analysis.[11] Inserting Expression (7) into Eq. (3) results in

$$\frac{1}{a_n}\frac{da_n}{d\zeta} - Gn\frac{db_n}{d\zeta}\rho^2 + ib_n(Gnb_n\rho^2-1) + \gamma_n = \frac{inN}{4a_n}\sum_{m=1}^{\infty} a_{n-m}a_m \exp\{G[nb_n-(n-m)b_{n-m}-mb_m]\rho^2\} + 2a_{m-n}^* a_m \exp\{G[nb_n-(m-n)b_{m-n}^*-mb_m]\rho^2\}, \quad n=1,2,\ldots. \tag{8}$$

At $\rho=0$, this equation reduces to

$$\frac{da_n}{d\zeta} + (\gamma_n - ib_n)a_n = \frac{inN}{4}\sum_{m=1}^{\infty} a_m(a_{n-m}+2a_{m-n}^*), \quad n=1,2,\ldots, \tag{9}$$

resulting in a coupled system of ODEs for the Gaussian amplitude parameters. At $\rho\to\infty$, the terms proportional to ρ^2 dominate and the equation reduces to

$$\frac{db_n}{d\zeta} = ib_n^2. \tag{10}$$

Equation (10) is independent of the harmonic number n, so the general width parameter b may be used. Equation (10) may be solved exactly, and applying the boundary condition $b(0)=1/G-i$ gives[11,12]

$$b(\zeta) = \frac{1-iG}{G-(G+i)\zeta}. \tag{11}$$

This may be substituted into Eq. (9) to obtain the final model equations

$$\frac{da_n}{d\zeta} + \left[\gamma_n - \frac{G+i}{G-(G+i)\zeta}\right]a_n = \frac{inN}{4}\sum_{m=1}^{\infty} a_m(a_{n-m}+2a_{m-n}^*), \quad n=1,2,\ldots. \tag{12}$$

Thus in using the Gaussian representation, the system of PDEs has been reduced to a system of ODEs which may be quickly integrated using standard numerical methods.

III. SIMULATIONS

The integration of Eq. (12) is performed by first letting $\mathbf{a}=(a_1,a_2,\ldots)^T$, and rewriting the equation in the vector form

$$\frac{d\mathbf{a}}{d\zeta} + \begin{pmatrix}\gamma_1-ib & & & & \\ & \gamma_2-ib & & & \\ & & \gamma_3-ib & & \\ & & & \gamma_4-ib & \\ & & & & \ddots\end{pmatrix}\mathbf{a} = 2iN\begin{pmatrix}0 & a_1^* & a_2^* & a_3^* & \cdots\\ a_1 & 0 & 2a_1^* & 2a_2^* & \cdots\\ a_2 & 2a_1 & 0 & 3a_1^* & \cdots\\ a_3 & 2a_2 & 3a_1 & 0 & \\ \vdots & \vdots & \vdots & & \ddots\end{pmatrix}\mathbf{a}. \tag{13}$$

This system is integrated using a globally second-order split-step algorithm[13] in which the linear part is solved using the trapezoid rule and the nonlinear part is solved using a second-order Runge-Kutta method. To account for the rapidly changing characteristic length scales encountered near the focus, the solution is computed on a nonuniform grid which is fine near the focus and coarse in other regions. The nodes of the nonuniform grid $\zeta_k, k=1,2,\ldots,K$ are obtained from nodes $\xi_k, k=1,2,\ldots,K$ distributed uniformly over $[0,2]$ via the transformation

$$\zeta_k = 1 + \sinh[\sinh^{-1}(G)(\xi_k-1)]/G, \tag{14}$$

where the gain G parametrizes the redistribution of the grid. As G increases, the grid becomes finer near $\zeta=1$ and coarser as $\zeta\to 0$ and $\zeta\to 2$.

To further accelerate computation, the number of harmonics necessary for the simulation is determined adaptively and varied as a function of ζ. The technique is similar to that used by Khokhlova *et al.*[2] In this implementation, the number of harmonics computed j is incremented by a certain number of additional harmonics Δj when $|a_j|>\varepsilon$. Similarly, j is decremented by Δj when $|a_j|<\varepsilon$. The example simulations discussed here were computed using $\varepsilon=10^{-6}$ and $\Delta j=1$, 2, or 5, depending on simulation requirements.

Two scenarios are used to compare the Gaussian representation technique with a finite-difference method. In these comparisons, the finite-difference code is used to integrate the frequency-domain KZK equation on the domain $\rho\in[0,3]$, $\zeta\in[0,1.1]$ using fixed grid spacing $\Delta\rho=5\times10^{-4}$, $\Delta\zeta=2.5\times10^{-4}$ and the boundary conditions of Eq. (6). The code uses second-order finite-difference approximations for

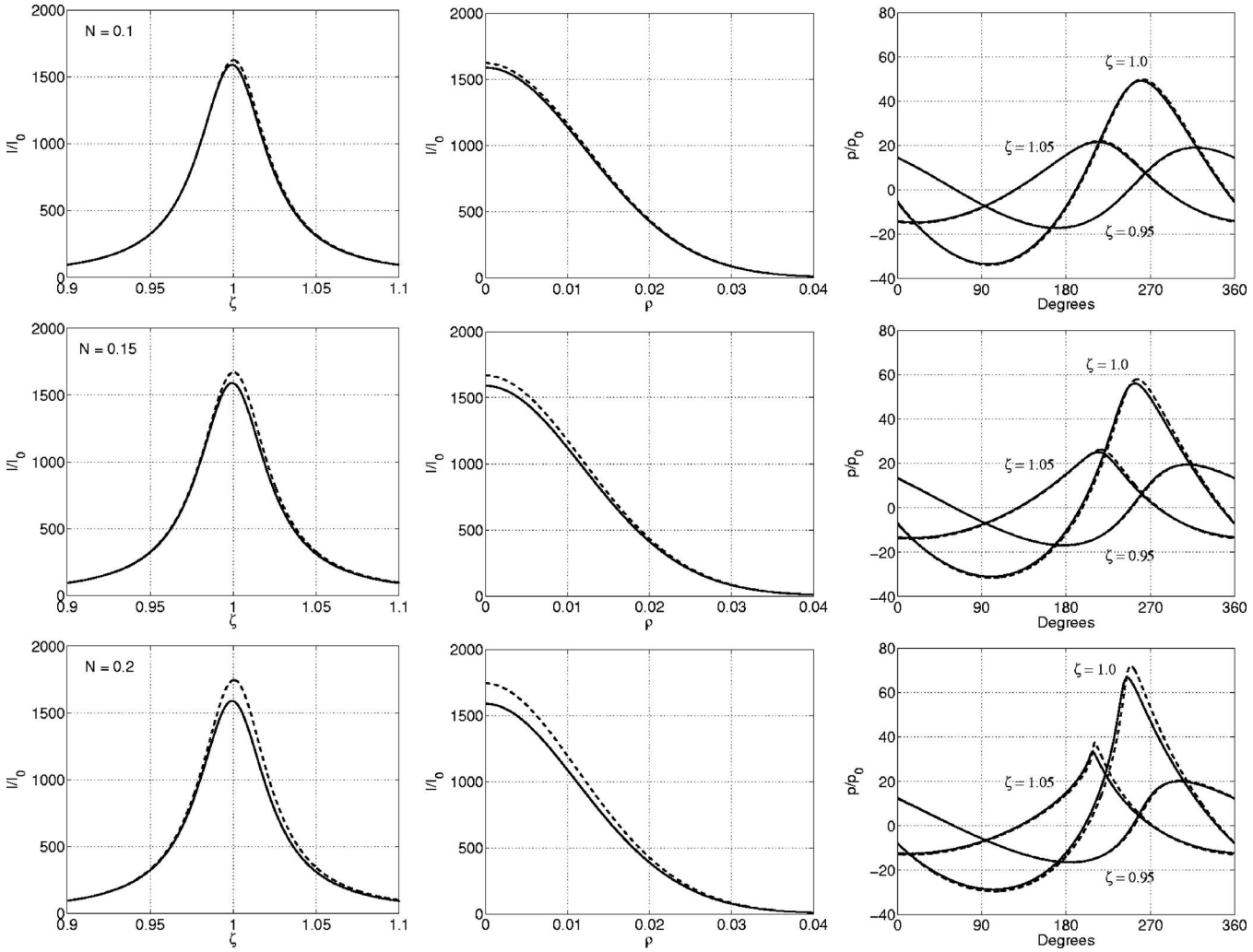


FIG. 1. Comparison of the Gaussian representation method (solid line) and full KZK (dashed line) simulations for HIFU propagation in water with $G=40$. From top to bottom, the rows show $N=0.1$, $N=0.15$, and $N=0.2$. Each row shows the normalized axial intensity, the normalized transverse focal intensity, and the normalized pressure wave forms.

the radial derivatives and operator splitting to solve the linear and nonlinear parts of the KZK equation separately. The linear part is integrated using the Crank-Nicolson technique while the nonlinear part is integrated using a second-order Runge-Kutta algorithm.

The first comparison consists of HIFU beam propagation in water at $G=40$. This corresponds to a transducer with a 4 cm effective radius and 10 cm focal length radiating at 1.18 MHz. The absorption is given by

$$\gamma_n = 3.41 \times 10^{-3} n^2. \tag{15}$$

The values of the nonlinear coefficient considered are $N=0.1$, 0.15, and 0.2, corresponding to 13, 30, and 53 W total acoustic power, respectively. Results from these simulations are shown in Fig. 1, where the normalized axial intensity, normalized radial intensity in the focal plane, and normalized pressure wave form are presented. The normalized time-average intensity is calculated in the plane-wave approximation

$$\frac{I(\rho,\zeta)}{I(0,0)} = \sum_{n=1}^{n_{\max}} |a_n(\zeta)|^2 \exp\{-2Gn\,\mathrm{Re}[b(\zeta)]\rho^2\}. \tag{16}$$

These plots show that the Gaussian representation agrees very well with the finite-difference method in the prefocal region. The solutions diverge at the focal and postfocal regions as N increases. This is due to the fact that the harmonic fields deviate from both the true Gaussian form and the $n^{-1/2}$ width trend as intensity (and propagation distance) increases, manifesting as suppressed energy transfer to higher harmonics and resulting in underpredicted pressures and intensities. Still, the results are qualitatively similar and nonlinear effects on the wave form shape are well represented.

Figure 2 shows results from the second comparison, consisting of a HIFU beam propagating in generic tissue whose absorption depends linearly on frequency. Using the same transducer geometry, the frequency required for $G=40$ is 1.23 MHz. The tissue parameters are chosen as $c_0=1550$ m/s, $\rho_0=1050$ kg/m^3, and $\beta=4.5$, and the absorption is $\alpha=0.55$ dB/cm/MHz, which, using Eq. (5), results in

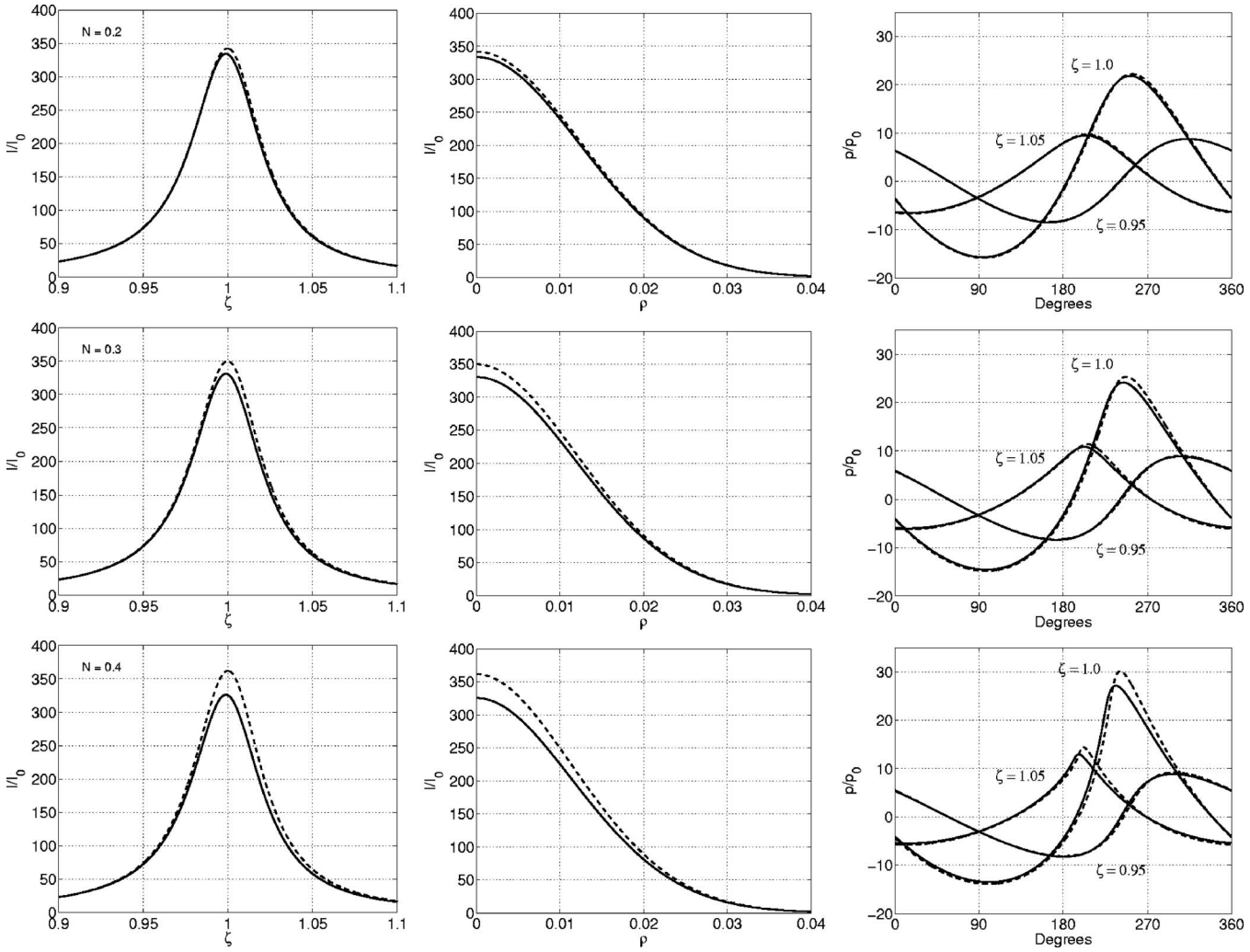


FIG. 2. Comparison of the Gaussian representation method (solid line) and full KZK (dashed line) simulations for HIFU propagation in tissue with $G=40$. From top to bottom, the rows show $N=0.2$, $N=0.4$, and $N=0.6$. Each row shows the normalized axial intensity, the normalized transverse focal intensity, and the normalized pressure wave forms.

$$\gamma_n = [1 - 2i \ln(1.23n)/\pi]n. \tag{17}$$

Results are shown for $N=0.2$, 0.3, and 0.4, corresponding to 39, 88, and 156 W total acoustic power, respectively. Due to stronger absorption in tissue, the Gaussian representation performs better at higher values of the nonlinear coefficient than it does in water. A similar trend of accurate prefocal approximation, and qualitative focal and postfocal representation is observed. Simulation parameters for the water and tissue runs are presented in Table I.

TABLE I. Simulation parameters.

N	Medium	Modes (j)	Δj	Grid points (K)	Time (s)[a]
0.1	Water	23	1	501	25.1
0.15	Water	50	2	501	25.9
0.2	Water	258	2	1001	69.2
0.2	Tissue	21	1	501	6.5
0.3	Tissue	42	1	501	6.8
0.4	Tissue	145	5	501	8.9

[a]MATLAB code implemented on Intel XEON 3.8 GHz processor running RedHat Linux.

To showcase the method's ability to handle high-gain simulations, Fig. 3 shows the results of a 3.3 MHz beam simulation in water using the same transducer geometry as mentioned earlier. Here, $G=112$ and $N=0.2$, corresponding to 7 W total acoustic power. The finite-difference code used for comparison in the previous figures was not able to obtain a solution at this frequency. The Gaussian representation method requires a maximum of 1120 modes and takes 16 min 38 s to compute this solution over 4001 grid points.

IV. CONCLUSIONS

In this work, a Gaussian approximation is used to reduce the frequency-domain representation of the KZK equation, consisting of a system of PDEs, to a simple system of ODEs. This approach works for transducers with Gaussian shading, automatically satisfies the boundary conditions at $r=0$ and $r\to\infty$, and significantly reduces the amount of computation required to perform simulations of HIFU beams. A comparison is made with a traditional finite-difference KZK code

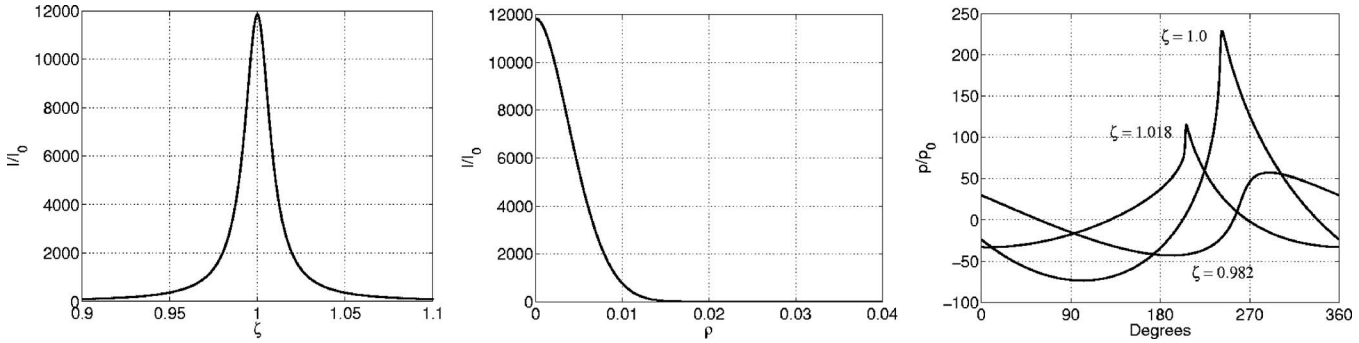


FIG. 3. Simulation of a high-frequency beam in water with $G=112$ and $N=0.2$. Left to right: Normalized axial intensity, normalized transverse focal intensity, and normalized pressure wave forms.

and the Gaussian approximation is shown to agree reasonably well in moderately nonlinear regimes, while requiring far less computer time.

While the application of this algorithm is limited to transducers with Gaussian shading, it is useful as an approximation of transducers which radiate uniform pressure distributions. System parameters could be modified to more accurately approximate beam propagation from uniform transducers in the presence of diffraction, absorption, and nonlinear interaction, to rapidly predict bioeffects in tissues, although this work is still ongoing. Moreover, this algorithm may prove useful to quickly explore the space of HIFU system parameters before committing to a time-intensive, general KZK simulator.

ACKNOWLEDGMENTS

This work was partially supported by DARPA through the Deep Bleeder Acoustic Coagulator Program.

[1]M. F. Hamilton and C. L. Morfey, "Model Equations," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 3, pp. 40–63.
[2]V. A. Khokhlova, R. Souchon, J. Tavakkoli, O. A. Sapozhnikov, and D. Cathignol, "Numerical modeling of finite-amplitude sound beams: Shock formation in the near field of a cw plane piston source," J. Acoust. Soc. Am. **110**, 95–108 (2001).
[3]Yu. A. Pishal'nikov, O. A. Sapozhnikov, and V. A. Khokhlova, "A modification of the spectral description of nonlinear acoustic waves with discontinuities," Acoust. Phys. **42**, 362–367 (1996).
[4]V. A. Khokhlova, M. A. Averkiou, S. J. Younghouse, M. F. Hamilton, and L. A. Crum, "Fast spectral algorithm for modeling of focused sound beams in a highly nonlinear regime," Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA, 20–26 June 1998, Vol. **4**, pp. 2875–2876.
[5]S. S. KashCheeva, O. A. Sapozhnikov, V. A. Khokhlova, M. A. Averkiou, and L. A. Crum, "Nonlinear distortion and attenuation of intense acoustic waves in lossy media obeying a frequency power law," Acoust. Phys. **46**, 170–177 (2000).
[6]F. P. Curra, P. D. Mourad, V. A. Khokhlova, R. O. Cleveland, and L. A. Crum, "Numerical simulations of heating patterns and tissue temperature response due to high-intensity focused ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1077–1089 (2000).
[7]K. Hynynen and N. McDannold, "MRI guided and monitored focused ultrasound thermal ablation methods: A review of progress," Int. J. Hyperthermia **20**, 725–737 (2004).
[8]Y. S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[9]T. S. Hart and M. F. Hamilton, "Nonlinear effects in focused ultrasound beams," J. Acoust. Soc. Am. **84**, 1488–1496 (1988).
[10]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **63**, 696–701 (1981).
[11]M. F. Hamilton, "Sound Beams," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 8, pp. 233–261.
[12]G. Du and M. A. Breazeale, "Theoretical description of a focused Gaussian ultrasonic beam in a nonlinear medium," J. Acoust. Soc. Am. **81**, 51–57 (1986).
[13]G. Strang, "On the construction and comparison of difference schemes," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **5**, 506–517 (1968).

The 2.5-dimensional equivalent sources method for directly exposed and shielded urban canyons

Maarten Hornikx[a] and Jens Forssén
Department of Civil and Environmental Engineering, Division of Applied Acoustics, Chalmers University of Technology, SE-41296 Göteborg, Sweden

(Received 1 March 2007; revised 16 August 2007; accepted 20 August 2007)

When a domain in outdoor acoustics is invariant in one direction, an inverse Fourier transform can be used to transform solutions of the two-dimensional Helmholtz equation to a solution of the three-dimensional Helmholtz equation for arbitrary source and observer positions, thereby reducing the computational costs. This previously published approach [D. Duhamel, J. Sound Vib. **197**, 547–571 (1996)] is called a 2.5-dimensional method and has here been extended to the urban geometry of parallel canyons, thereby using the equivalent sources method to generate the two-dimensional solutions. No atmospheric effects are considered. To keep the error arising from the transform small, two-dimensional solutions with a very fine frequency resolution are necessary due to the multiple reflections in the canyons. Using the transform, the solution for an incoherent line source can be obtained much more efficiently than by using the three-dimensional solution. It is shown that the use of a coherent line source for shielded urban canyon observer positions leads mostly to an overprediction of levels and can yield erroneous results for noise abatement schemes. Moreover, the importance of multiple façade reflections in shielded urban areas is emphasized by vehicle pass-by calculations, where cases with absorptive and diffusive surfaces have been modeled. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783197]

PACS number(s): 43.28.Js, 43.20.El, 43.28.En [VEO] Pages: 2532–2541

I. INTRODUCTION

Sound propagation in urban areas has been studied extensively during the last few decades. A main key in urban sound propagation research at a scale up to hundreds of meters is, according to the number of publications, believed to lie in understanding the sound field in a single urban canyon.[1–8] Studies were also addressed to sound propagation in several streets[9,10] and from one street to a shielded canyon, where sound propagates over a single building block.[11–14] The last category is of special interest here. Sound pressure levels at sides of dwellings that are directly exposed to road traffic noise are difficult to reduce to the desired level regarding annoyance and adverse health effects. Therefore, access to relatively quiet areas became a strategy towards offering inhabitants of city centers a healthy living environment regarding road traffic noise.[15] Such areas are often shielded from direct road traffic noise, e.g., courtyards. For this purpose, accurate models have been developed to understand the behavior of sound propagation from a directly exposed urban street canyon to a parallel canyon shielded from direct road traffic noise.[12,14] The idealized situation of parallel urban canyons can be considered as typical for large city centers and the geometry extends the widely studied cases of single and double noise screens. Scattering and absorption of sound waves at the façades of the canyons were shown to play an important role in weakening the sound propagation from the directly exposed to the shielded canyon, and meteorological effects, such as refraction and scattering by atmospheric turbulence, were found to influence the sound pressure levels in the shielded canyon significantly.[12,14] The results of these studies were based on the solution of the two-dimensional (2D) wave equation in the time domain[12] and frequency domain.[14] These solutions therefore represent the response from a coherent line source. An incoherent line source is, however, a better model for traffic on a straight road. For situations with a single screen, the attenuation of a coherent line source was shown to be larger than that of an incoherence line source.[16,17] It is therefore relevant to investigate whether these differences also appear in the parallel canyons case.

Possessing a point source solution offers the possibility to calculate the response from an incoherent line source. A pass-by of a single vehicle to compare various geometrical situations can then also be modeled. A solution of the three-dimensional (3D) Helmholtz equation (or wave equation) in the situation of canyons is computationally still expensive, however. When assuming the length of the streets as infinite and the cross section invariant over this length, a transform of solutions of the 2D Helmholtz equation to a solution of the 3D Helmholtz equation is possible. The approach, which is called a 2.5D method, was first presented by Duhamel[16] and used later by Salomons *et al.*,[18] Duhamel and Sergent,[19] Jean *et al.*,[17] Godinho *et al.*,[20] and Defrance and Jean,[21] among others. The method has been applied to compute point source solutions for various types of noise screens. Methods like the boundary element method (BEM) and the image sources method (ISM) were used to calculate the 2D solutions. In this paper, the transform is applied to obtain a point source solution in the case of parallel urban canyons. Road traffic noise is modeled by a sound source in one canyon, called the directly exposed canyon, and the observer

[a]Electronic mail: maarten.hornikx@chalmers.se

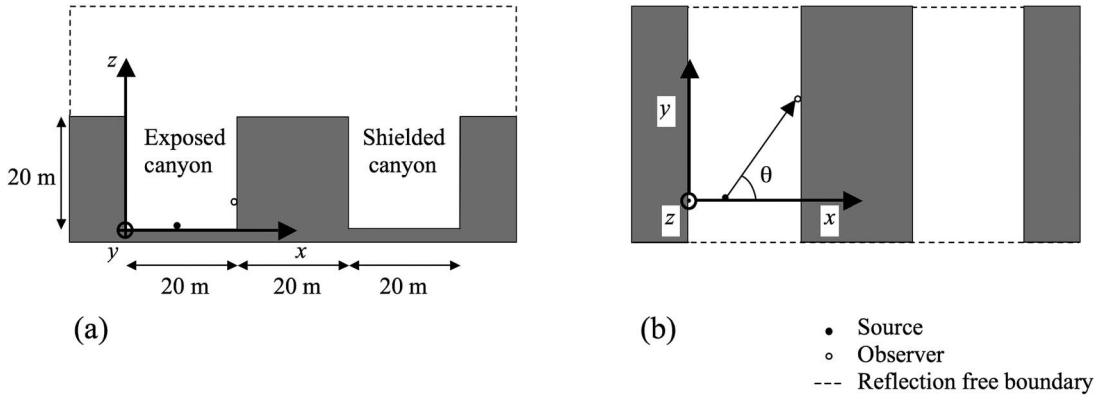


FIG. 1. Geometry of parallel urban street canyons with the used coordinate system. Arbitrary source and observer positions are shown: (a) cross section; and (b) top view.

positions are located in the same canyon or in a parallel shielded canyon. The equivalent sources method (ESM), previously developed to model sound propagation in geometries with urban canyons in the frequency domain,[14] is used to generate the 2D solutions. No atmospheric effects are included in the calculations of this study.

This paper is organized as follows. In Sec. II, a transform of solutions of a 2D Helmholtz equation to that of a 3D Helmholtz equation is presented. Compared to previous work, including atmospheric absorption in the transform is novel. The relevant sources of error in the transform are investigated for the free field solution. The accuracy of the method is further discussed by studying the case of a point source and an observer between two semi-infinite walls, an extreme case of a narrow street canyon. Then, the transform is applied to the ESM for the parallel canyons case with observer positions both in the directly exposed and in the shielded canyon. The transformed ESM is used for calculations in Sec. III. In Sec. III A, the pass-by of a single vehicle is modeled for a single screen and the parallel canyon situation, where several façade properties in the canyons are studied. Section III B focuses on the difference between the excess attenuation of a coherent and an incoherent line source for the various cases of urban street canyons.

II. THE 2.5D ESM

A. Transform from a 2D to a 3D solution of the Helmholtz equation

The geometry under study is shown in Fig. 1. A point source and an observer are located in a situation of parallel urban canyons. The observer may be placed in either the directly exposed or the shielded canyon. Except for the point source, the geometry is assumed to be invariant in the y direction: a 2.5D geometry. The purpose is to solve the 3D wave equation for this geometry in the frequency domain in a fast way. Solving a 3D equation for a 2.5D geometry is here further called a 2.5D method. It is done by applying a Fourier transform to the 3D Helmholtz equation with respect to the y coordinate. The resulting equation is set equivalent to a 2D Helmholtz equation, which is then solved efficiently. To obtain the desired 2.5D solution, an inverse Fourier transform over wave number k_y is finally applied.

Three-dimensional sound propagation in an unbounded medium due to a mono frequency monopole with unity source strength is described by the Helmholtz equation. The time convention $e^{i\omega t}$ has been adopted with ω the angular frequency and t the time. The effect of air absorption is included by an imaginary part to the wave number. In Cartesian coordinates (x,y,z), we write

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}+(k-i\alpha(f))^2\right]p = -\delta(x-x_s)\delta(y-y_s)\delta(z-z_s), \tag{1}$$

where $p(x,y,z,k)$ is the complex pressure due to the monopole source located at (x_s,y_s,z_s), and $\alpha(f)$ the frequency dependent air absorption coefficient. Applying a Fourier transform in the y direction, we obtain

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}+((k-i\alpha(f))^2-k_y^2)\right]P = -e^{ik_y y_s}\delta(x-x_s)\delta(z-z_s), \tag{2}$$

with

$$P(x,k_y,z)=\int_{-\infty}^{\infty} p(x,y,z,k)e^{ik_y y}\,dy,$$

where $k=\sqrt{k_x^2+k_y^2+k_z^2}$. This equation is equivalent to the 2D Helmholtz equation

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}+(K-i\alpha_{2D}(f,k_y))^2\right]q = -\delta(x-x_s)\delta(z-z_s), \tag{3}$$

where $q(x,K,z)=Pe^{-ik_y y_s}$, $K=\sqrt{k^2-k_y^2}$, and $\alpha_{2D}(f,k_y)=i(\sqrt{(k-i\alpha(f))^2-k_y^2}-K)$. The solution to Eq. (1) is obtained by solving Eq. (3) and applying an inverse Fourier transform with respect to k_y. For future use, we write the transform pair

$$p(x,y,z,k)=\frac{1}{2\pi}\int_{-\infty}^{\infty} q(x,K,z)e^{ik_y(y_s-y)}\,dk_y, \tag{4}$$

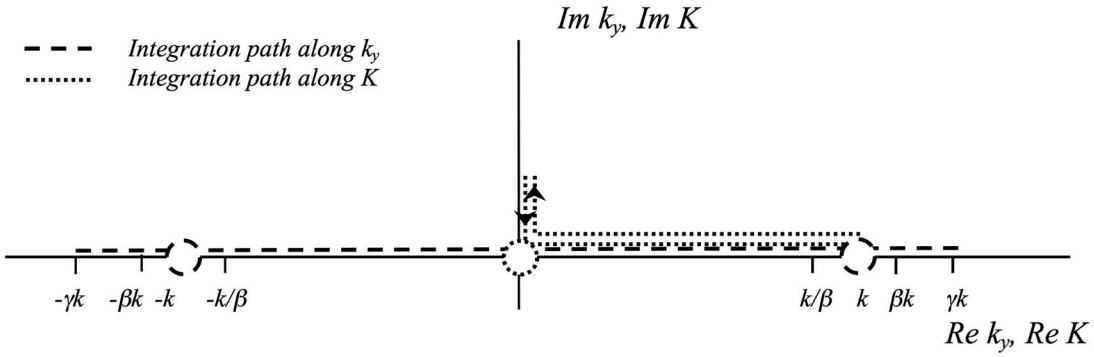


FIG. 2. Integration paths in the complex plane for the wave numbers k_y and K in Eq. (4) for unbounded wave propagation. (--) Integration along k_y with singularities at $k_y=-k$ and $k_y=k$; (··) integration along K with a singularity at $K=0$. Integration is not done around the singularities.

$$q(x,K,z)=\int_{-\infty}^{\infty} p(x,y,z,k)e^{-ik_y(y_s-y)}\,dy. \tag{5}$$

According to Eq. (4), the 2D solutions $q(x,K,z)$ over all wave numbers k_y are needed to calculate $p(x,y,z,k)$. Integration over all wave numbers k_y implies an integration path over a part of the real and imaginary axis of K for a certain wave number k [see Fig. 2].

The solution of Eq. (3) in an unbounded domain, is

$$q(x,K,z)=\begin{cases} -\dfrac{i}{4}H_0^{(2)}(Kr_{2D}) & K^2\geqslant 0 \\ \dfrac{1}{2\pi}K_0^{(1)}(|K|r_{2D}) & K^2<0 \end{cases}, \tag{6}$$

where $r_{2D}=\sqrt{(x-x_s)^2+(z-z_s)^2}$, $H_0^{(2)}$ the Hankel function of order zero and second kind and $K_0^{(1)}$ the modified Bessel function of order zero and first kind. For a real positive K value, the 2D solution is a traveling wave, whereas for imaginary K (i.e., for $k_y^2>k^2$), the Helmholtz equation gets the form of a diffusion equation approximately obeying an exponentially decaying Green's function. Salomons[18] showed that the major contribution of the integral in Eq. (4) lies around the point of stationary phase, which is situated at $k_y=k\sin\theta$, with $\theta=\arctan[(y-y_s)/r_{2D}]$ [see Fig. 1(b) for angle θ]. Since the point of stationary phase is bounded for k_y by $-k$ and k, integration limits with respect to k_y may be restricted to $-\gamma k$ and γk, with γ a number slightly larger than 1 [see Fig. 2]. Integration from $-\gamma k$ to γk implies integration over a singular value of q (both for the Hankel and modified Bessel function) in the origin of the complex plane [see Fig. 2]. This happens either when $r_{2D}=0$ or when $K=0$ [see Eq. (6)]. For the integration variable k_y, the singularity appears when $k_y=-k$ or $k_y=k$. The integral in Eq. (4) could be calculated applying residue calculus, yet since the 2D equation is solved numerically later (using the ESM), a numerical approach similar to the one presented by Salomons is used.[18] This approach embraces substitution of the integration variable for values around the singularity. The integration variable k_y is substituted by $u=\sqrt{k^2-k_y^2}$ for $|k/\beta|<|k_y|<|k|$ and by $v=\sqrt{k_y^2-k^2}$ for $|k|<|k_y|<|k\beta|$, where $\beta>1$ and a real number. The integral can now be divided into four parts along the real k_y axis (since q is independent of the sign of k_y). Solutions of q are calculated in each part with a (different) frequency resolution Δf and are spline interpolated to an equidistant wave number resolution Δk

$$\begin{aligned} p(x,y,z,k)=&\frac{1}{2\pi}\int_0^{k/\beta} q(x,K,z)\cdot(e^{ik_y(y_s-y)}-e^{-ik_y(y_s-y)})dk_y \\ &+\frac{1}{2\pi}\int_{0^+}^{k\sqrt{1-\beta^{-2}}} q(x,u,z)\cdot(e^{i\sqrt{k^2-u^2}(y_s-y)} \\ &-e^{-i\sqrt{k^2-u^2}(y_s-y)})\frac{u}{\sqrt{k^2-u^2}}du \\ &+\frac{1}{2\pi}\int_{0^+}^{k\sqrt{\beta^2-1}} q(x,iv,z)\cdot(e^{i\sqrt{k^2+v^2}(y_s-y)} \\ &-e^{-i\sqrt{k^2+v^2}(y_s-y)})\frac{v}{\sqrt{k^2+v^2}}dv+\frac{1}{2\pi}\int_{k\beta}^{k\gamma} q(x,K,z) \\ &\cdot(e^{ik_y(y_s-y)}-e^{-ik_y(y_s-y)})dk_y. \end{aligned} \tag{7}$$

Then, the integrals are approximated by sums

$$\begin{aligned} p_n(x,y,z,k)=&\frac{1}{2\pi}\sum_{n=1/2}^{L-1/2} q(x,\sqrt{k^2-(n\Delta k_y)^2},z) \\ &\cdot(e^{in\Delta k_y(y_s-y)}+e^{-in\Delta k_y(y_s-y)})\Delta k_y \\ &+\frac{1}{2\pi}\sum_{n=1/2}^{M-1/2} q(x,n\Delta u,z)\frac{n\Delta u}{\sqrt{k^2-(n\Delta u)^2}} \\ &\cdot(e^{i\sqrt{k^2-(n\Delta u)^2}(y_s-y)}+e^{-i\sqrt{k^2-(n\Delta u)^2}(y_s-y)})\Delta u \\ &+\frac{1}{2\pi}\sum_{n=1/2}^{N-1/2} q(x,in\Delta v,z)\frac{n\Delta v}{\sqrt{k^2+(n\Delta v)^2}} \\ &\cdot(e^{i\sqrt{k^2+(n\Delta v)^2}(y_s-y)}+e^{-i\sqrt{k^2+(n\Delta v)^2}(y_s-y)})\Delta v \\ &+\frac{1}{2\pi}\sum_{n=1/2}^{O-1/2} q(x,\sqrt{k^2-(k\beta+n\Delta k_y)^2},z) \\ &\cdot(e^{i(k\beta+n\Delta k_y)(y_s-y)}+e^{-i(k\beta+n\Delta k_y)(y_s-y)})\Delta k_y, \end{aligned} \tag{8}$$

with $L=k/\beta\Delta k_y$, $M=k\sqrt{1-\beta^{-2}}/\Delta u$, $N=k\sqrt{\beta^2-1}/\Delta v$, $O=(k\gamma-k\beta)/\Delta k_y$, and $p_n(x,y,z,k)$ the numerical 3D solution. The values for k, β, γ, Δk_y, Δu, and Δv should be chosen such that L, M, N, and O get integer values.

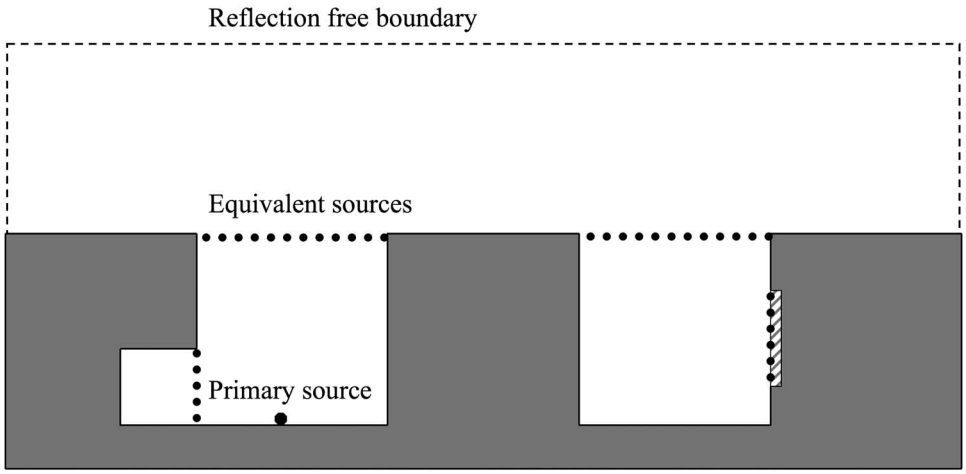


FIG. 3. Principle of the ESM. Equivalent sources are positioned at interfaces between cavities, impedance boundaries, and the free field.

The transform of a coherent line source solution to a point source solution can be applied to any method fulfilling Eq. (3) and boundary conditions [see Sec. II C]. For the geometry of Fig. 1, the ESM is used.[14] This method has been validated before by the BEM[14] and a scale model study.[22] In the ESM, the domain is decoupled into subdomains: closed cavities and a free space above a flat ground surface [see Fig. 3]. The domains are coupled by placing equivalent sources at the interfaces of the domains. First, the Green's functions from the primary source to the equivalent sources and among the equivalent sources in the subdomains are calculated. The Green's functions used here are a modal solution in the cavities and a semifree field Green's function above the canyons. To obtain cavity modes with finite amplitude, a small amount of damping has been used in the cavity Green's functions.[14] Then, the strengths of the equivalent sources at the subdomain interfaces are calculated using conditions for pressure and normal velocity over the interfaces. Absorption patches at the boundaries and niches (e.g., balconies) can be modeled by additional subdomains and equivalent sources.

For the imaginary frequencies, the Helmholtz equation (now a diffusion equation) is solved by a second-order accurate finite difference scheme. At the rigid boundaries, the pressure gradient is set to zero and at the far away reflection free boundaries of the computational domain, the pressure is set to zero. The size of the domain above the canyons was chosen to be 240 m $\times$ 140 m such that the results at the observer points converged.

The transform applied to the 2D ESM is here called the 2.5D ESM. All 2.5D ESM calculations in this paper have been made without including air absorption.

B. Accuracy in the free field situation

We will now discuss two aspects of the transform in Eq. (4) that are important regarding accuracy in the current application. Sound propagation in an unbounded domain will be considered first. The first aspect concerns the importance of the integration path for $|k_y| > |k|$ (i.e., along the imaginary axis of K). As mentioned, the solution of the Helmholtz equation for imaginary frequencies obeys a different type of equation and is not straightforward in some calculation models. Figure 4 shows the calculation error of the transform for two frequencies and two different y distances, both with and without the imaginary frequencies. The results are plotted as a function of r_{2D} (which is equal to the x distance, since the z coordinate of the source and observer are equal to zero). The error is defined as $20\log(|p_n|/|p_{\text{exact}}|)$, with $p_{\text{exact}} = e^{-ikr}/4\pi r$, and $r = \sqrt{r_{2D}^2 + (y - y_s)^2}$. The numerical results have been calculated using Eqs. (6) and (8). The results show that omitting the imaginary frequencies in the integral evaluation introduces fluctuations in the result for small r_{2D} values. The error is larger for smaller r_{2D} values and for lower frequencies. These effects are due to the slower decay of the function $K_0^{(1)}(Kr_{2D})$ in amplitude for smaller K and smaller r_{2D}. It is also visible that very close to $r_{2D} = 0$ m, there is even an error when imaginary frequencies are included. This error can be attributed to the location of the point of stationary phase. When θ goes towards 90° [see Fig. 1(b) for θ], the point of stationary phase shifts to the singularity point where

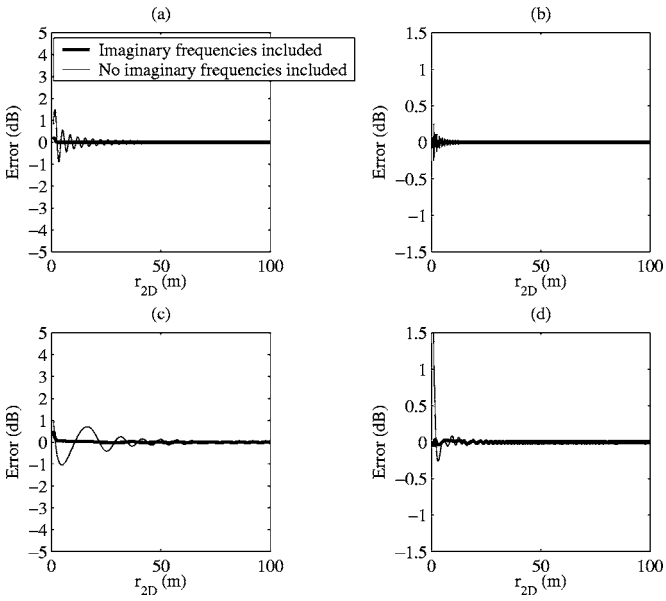


FIG. 4. Error in the transform in Eq. (4) for sound propagation in an unbounded domain, both with and without imaginary frequencies. The error is $20\log(|p_n|4\pi r)$, p_n calculated according to Eq. (8) with a frequency discretization of $\Delta f = 0.1$ Hz. Source at (0,0,0) and observer at $(x,y,0)$: (a) $f = 100$ Hz, $y = 0$ m; (b) $f = 1000$ Hz, $y = 0$ m; (c) $f = 100$ Hz, $y = 100$ m; and (d) $f = 1000$ Hz, $y = 100$ m.

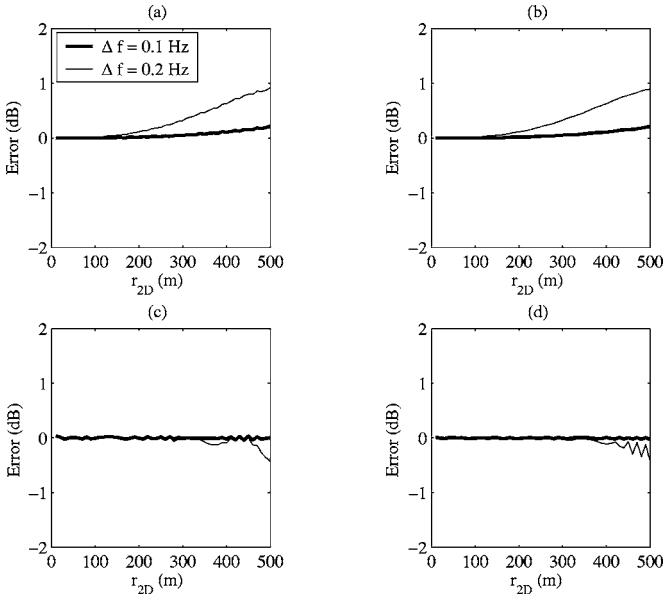


FIG. 5. Error in the transform in Eq. (4). Source at (0,0,0) and observer at $(x,y,0)$. Two frequency discretizations have been used for the evaluation of the integral. The calculated $r_{2D,\max}$ are 344 m for $\Delta f=0.2$ Hz and 688 m for $\Delta f=0.1$ Hz ($c_0=344$ m/s): (a) $f=100$ Hz, $y=0$ m; (b) $f=1000$ Hz, $y=0$ m; (c) $f=100$ Hz, $y=100$ m; and (d) $f=1000$ Hz, $y=100$ m.

the numerical integration is still more sensitive. The fluctuation of the error is slower for $y=100$ m than for $y=0$ m. This is most obvious at 100 Hz. The reason is again the location of the point of stationary phase; for $y=100$ m, the angle θ is larger at the same x position than for $y=0$ m, which implies that the point of stationary phase gets closer to the imaginary frequencies.

The second aspect concerns the frequency resolution Δf. The free field solution of the 2D Helmholtz equation that appears in the argument of the integral in Eq. (4) is oscillating over k_y. For a larger r_{2D}, this function oscillates more rapidly over k_y. To preserve accuracy of the numerical integration, a minimum frequency resolution corresponding to five points per period was therefore proposed,[1] $\Delta f = c_0/5r_{2D,\max}$, where $r_{2D,\max}$ is the maximum distance of interest. Figure 5 shows the error as function of the r_{2D} distance for two frequency resolutions and two frequencies. The error is defined as has been done for Fig. 4. It is obvious that the frequency dependence concerning r_{2D} is not strong. Further, we notice that a certain frequency discretization holds for a longer r_{2D} distance for $y=100$ m than for $y=0$ m. This is due to the fact that, at $y=0$ m, the point of stationary phase is located exactly at $k_y=0$. A constant frequency discretization, and consequently a constant discretization for K, yields a discretization for k_y that is biggest for $k_y=0$. Thus, with the point of stationary phase located at $k_y=0$, the introduced discretization error is largest here.

C. Accuracy in the directly exposed street canyon

We now turn to a situation with source and observer positions situated between two semi-infinite walls with the source placed at a ground surface [see Fig. 6]. The walls and

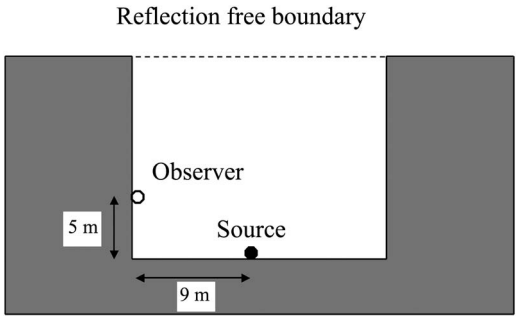


FIG. 6. Source at (9,0,0) and observer at $(0,y,5)$ above a rigid ground surface between two semi-infinite rigid façades.

the ground have an infinitely high acoustic impedance, further denoted as a rigid surface. In general, the boundary condition that belongs to Eq. (2) reads

$$\frac{\partial P}{\partial n} = -\frac{ik}{Z_s}P, \tag{9}$$

at the boundaries, with the frequency dependent normalized surface impedance

$$Z_s(f) = \frac{Z_{\text{boundary}}(f)}{Z_{\text{air}}},$$

where Z_{boundary} and Z_{air} are specific impedances of the boundary material and air. The 2D Helmholtz equation, Eq. (3), is completed by

$$\frac{\partial q}{\partial n} = -\frac{iK}{Z_{s,2D}}q \tag{10}$$

at the boundaries. For equivalence of Eqs. (9) and (10), we write

$$Z_{s,2D}(f,k_y) = \frac{Z_s(f)K}{k}. \tag{11}$$

The effect of air absorption has been neglected in this equivalence condition, and a source position of $y_s=0$ m is assumed. When the surface impedances are infinitely large, as is our main interest, the equivalence condition of Eq. (11) is automatically fulfilled. The situation can be considered as the limit case of a narrow street canyon. Therefore, and since analytical solutions in two and three dimensions are straightforward for this situation, it is worth studying the influence of the errors of the transform in this case. The source and observer positions correspond to those used in a scale model measurement study.[22] The 2D solution p_n is computed by a summation of 50 image sources using Eq. (6), whereas the 3D solution consists of a summation of the same number of image sources using the free field point source solution $e^{-ikr}/4\pi r$. Note that the error in the image sources method vanishes when the number of image sources is increased. Figure 7 [panels (ai) and (aii)] shows the sound pressure level relative to the free field level as function of frequency for one y position, and as function of y for $f=100$ Hz. The source and observer positions are given in Cartesian coordinates (x,y,z) according to Fig. 1. Panels (b)–(e) of Fig. 7 show the errors in the transform defined as $20\log(|p_n|/|p_{3D,\text{ISM}}|)$, where ISM refers to image sources

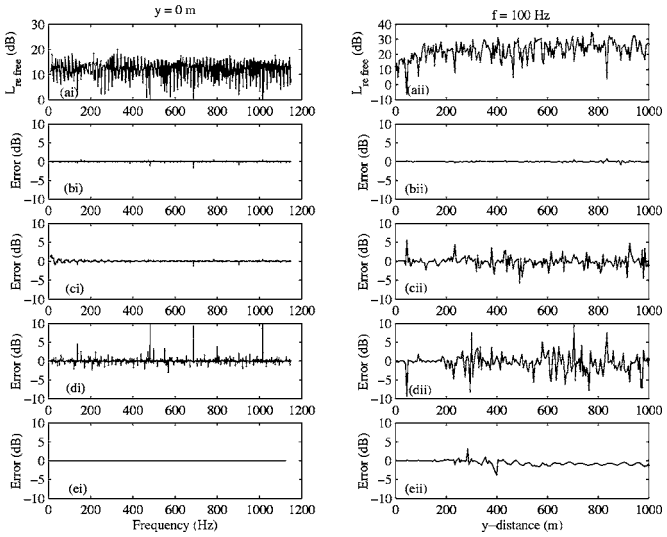


FIG. 7. Calculations for the source and observer positions as in Fig. 6. (ai) L_{refree} for source at (9,0,0) and observer at (0,0,5); (aii) L_{refree} for source at (9,0,0) and observer at $(0,y,5)$ at 100 Hz. Panels (b)–(e) show the errors in the transform (see text) for various choices in the numerical scheme to calculate p_n: (b) $\Delta f=0.1$ Hz; (c) $\Delta f=0.1$ Hz, no imaginary frequencies; (d) $\Delta f=0.2$ Hz; (e) $\Delta f=0.1$ Hz, $Z_{\text{facades}}=17.8$.

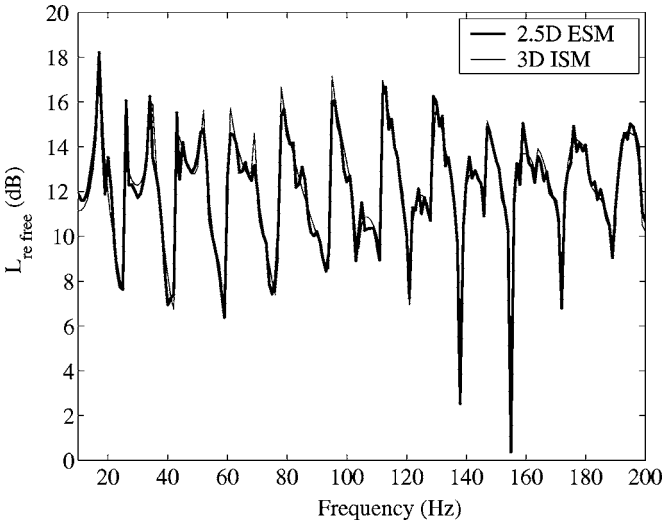


FIG. 8. L_{refree} with a point source at (9,0,0) and the observer at (0,0,5) in the geometry of Fig. 1. Imaginary frequencies have been included in the transform of the 2.5D ESM and 3D ISM calculations have been made without diffraction.

method. With a frequency resolution of $\Delta f=0.1$ Hz to calculate p_n, the errors over frequency are small and occur mostly at resonances [see Fig. 7(bi)]. This error increases with y distance [see Fig. 7(bii)], which is due to the fact that the contribution of higher order reflections to the sound pressure level increases with increasing y position. These higher order reflections have a larger path length, which results in a larger error [see Fig. 5]. Still, the errors at 1000 m are small. Omitting the imaginary frequencies leads to larger errors [see Figs. 7(ci) and 7(cii)], which are more pronounced at the lower frequencies, as observed before. Excluding the imaginary frequencies leads to substantial errors over the y distance at 100 Hz. These errors do not seem to increase substantially over the distance. This can be understood since all observer positions are located at the same r_{2D} position from the source. For a coarser frequency resolution of $\Delta f=0.2$ Hz, the errors increase for all frequencies compared to the $\Delta f=0.1$ Hz case [see Fig. 7(di)]. Also, the errors increase for $y>200$ m due to the influence of the higher order reflections at larger y position, [see Fig. 7(dii)].

When the impedance of the boundaries is finite, a distinct set of 2D solutions has to be calculated for each frequency in 3D (since $Z_{s,2D}$ is a function of f and k_y), which increases the computation time. To show the influence of an impedance boundary on the error, we use a real valued surface impedance of the boundaries of $Z_s=17.8$, corresponding to an absorption coefficient of 0.2 for normal wave incidence. The results are shown in Figs. 7(ei) and 7(eii). The 2D and 3D solutions have now been calculated by the following sums:

$$q(x,K,z) = -2\frac{i}{4}\sum_{n=0}^{M} Q_{\text{cyl}}^{n}(Z_{s,2D}(f,k_y),r_{2D,n})H_0^{(2)}(Kr_{2D,n}), \tag{12}$$

$$p(x,y,z,k) = 2\sum_{n=0}^{M} Q_{\text{sph}}^{n}(Z_s(f),r_n)\frac{e^{-ikr_n}}{4\pi r_n}, \tag{13}$$

where Q_{cyl} is the cylindrical wave reflection coefficient implemented according to Chandler-Wilde and Hothersall,[23] Q_{sph} the spherical wave reflection coefficient implemented according to Chien and Soroka,[24] and $r_{2D,n}^2=\sqrt{(z-z_s)^2+(x-x_{s,n})^2}$, with $x_{s,n}$ the x distance of the nth image source to the receiver. The above expressions are generally applicable for any (complex valued) impedance. For high frequencies and large distances, Q_{sph} and Q_{cyl} have both been substituted by the plane wave reflection coefficient, Q_{plane}. A frequency resolution of $\Delta f=0.1$ Hz has been used. Figure 7(ei) shows that the results over frequency are better than the corresponding results without absorption. An explanation is that the largest error in the latter case was due to higher order reflections (with distances larger than $r_{2D,\max}$ from the source), which play a smaller role when increasing the absorption. Over the y distance, the error increases from 200 m on. Inaccuracies from the first-order reflections were found to cause this error. Minor errors in the numerically calculated reflection coefficients could cause the calculated error, i.e., the error shown may overpredict the actual error due to the transform.

For the directly exposed canyon of Fig. 1, the 2.5D ESM is used. Figure 8 shows the sound pressure level relative to the free field level, L_{refree}. Results of the 2.5D ESM including imaginary frequencies are compared with calculations with the 3D ISM for the situation with two semi-infinite walls [see Fig. 6]. The agreement is fine and the small error was shown to disappear at higher frequencies. The small error at low frequencies can be attributed to the lack of the diffraction contributions in the ISM model. This is supported by a comparison between 2D results from the ESM, the ISM, and the BEM; see Fig. 9. The ISM results display deviations from the ESM and BEM results. Since the 2D solutions are shown, no transform errors are included. At a larger y dis-

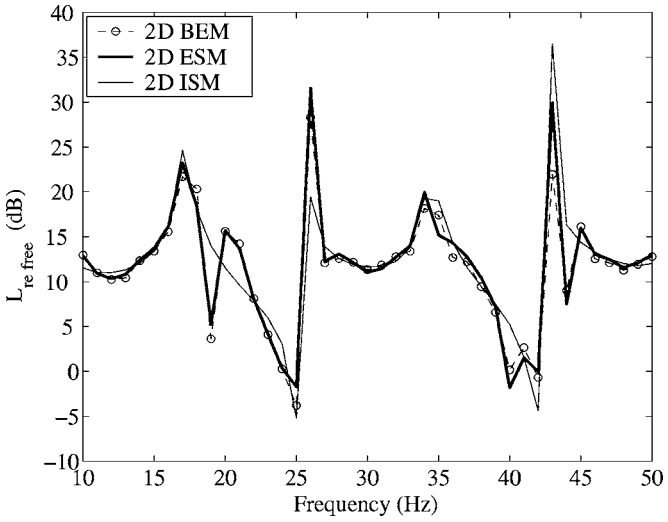


FIG. 9. L_{refree} with a coherent line source at (9,0,0) and the observer at (0,0,5) in the geometry of Fig. 1. The 2D ISM result has been calculated without diffraction.

tance, the deviations between results of the 3D ISM model and the transformed ESM are expected to increase, since the diffraction contribution is larger there.

For engineering use of the 2.5D ESM, errors are calculated in the 100 Hz 1/3-octave band for an observer point in the directly exposed street canyon and two different façade types; rigid façades and façades with rigid diffusion patches applied. In Fig. 10(a), the street canyons geometry with diffusion patches is shown. The patches have a dimension of 0.2 m by 2.0 m and are positioned according to a scale model study.[22] The case with diffusion patches is somewhat closer to a realistic case. Table I shows the required frequency descritization for the two cases, where a maximum error of 0.5 dB relative to a converged calculation with a frequency discretization of $\Delta f=0.05$ Hz has been used as a tolerance. The same frequency discretization of $\Delta f=0.4$ Hz can be used for an acceptable error for both situations. A larger Δf will be allowable for larger y distances. The reason has been mentioned before in the discussion of Fig. 5: the discretization in k_y is coarsest around $k_y=0$ and introduces the largest error if the observer is located at $y=y_s$.

TABLE I. Required frequency discretization (Hz) in the 2.5D ESM for a tolerance of 0.5 dB. Situation of Fig. 1 for rigid façades and Fig. 10(a) for the case with diffusion patches. Source coordinates are (9,0,0). For the incoherent line source calculations, $\Delta f=0.2$ Hz up to 50 Hz.

	Point source		Incoherent line source
	$f=100$ Hz, $y=0$ m		$f=1000$ Hz
Observer coordinates	(0,0,15)	(60,0,15)	(49,0,0)
Rigid façades	0.4	0.1	0.4
Diffusion patches	0.4	0.2	1.6

D. Accuracy in the shielded canyon

For the shielded canyon, no analytical solution exists. To investigate the sources of error in the transform in the shielded canyon of Fig. 1, the frequency resolution error is studied by convergence of the 2.5D ESM results with Δf, as for the directly exposed canyon. The source coordinates are (9,0,0) and the observer is located at coordinates (60,0,15). Since the shortest travel path between the source and nearest observer is approximately 67 m, and thereby much larger than the wavelength, the imaginary frequencies are omitted in the calculation. The required frequency discretizations for a tolerance of 0.5 dB are printed in Table I. In contrast to the directly exposed street canyon results, a finer frequency discretization is necessary for an acceptable accuracy. The influence of higher order reflections causes this difference. For the diffusion patches case, the required frequency discretization is coarser than for the case of rigid façades. The reason for this difference is that the higher order reflections are of more importance in the case of rigid façades.

Calculation time with the 2D ESM method increases with frequency. For the geometry with diffusion as in Fig. 10(a), a single observer position, $\Delta f=0.2$ Hz and for 1000 Hz, a 2.5D calculation at the shielded side takes approximately 42 h using MATLAB on a computer with a dual 2.66 GHz processor and a 4 GB memory. Six equivalent sources per wavelength are then used in the method and no parallel calculations are done. Since we have calculated 2D frequencies up to 1000 Hz, this computation time stands for the calculation of all 2.5D results up to 1000 Hz. When

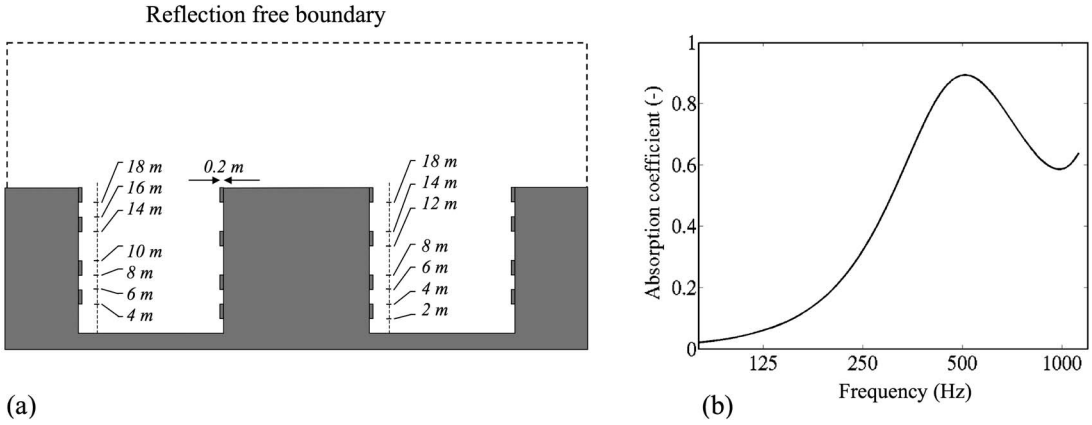


FIG. 10. (a) Cross section of two parallel street canyons with extra patches (absorption or diffusion). (b) The absorption coefficient for normal wave incidence used for the absorption patches. Diffusion patches are rigid.

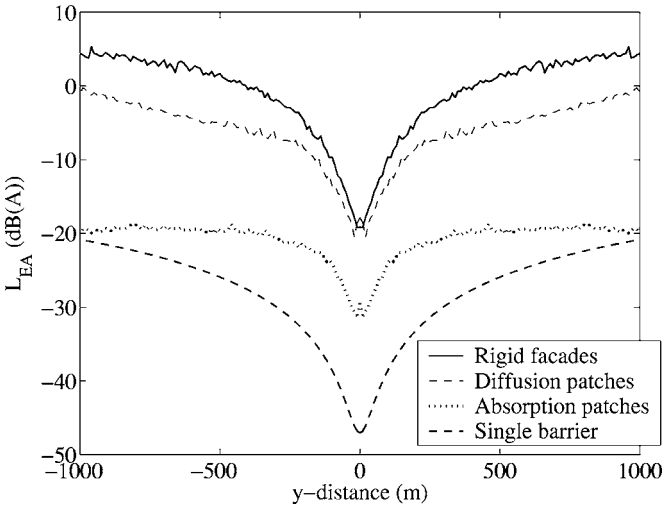


FIG. 11. Vehicle pass-by excess attenuation levels for the geometry of Fig. 1. Source at $(9,y,0)$ and observer at (49, 0, 0). Diffusion and absorption patches as in Fig. 10.

façades would consist of material with a finite impedance, 42 h would be the approximate calculation time for 1000 Hz only. Other frequencies require a new calculation over the 2D frequencies. In the absence of air absorption, however, cavity Green's functions have to be calculated only once per frequency.

III. POINT AND LINE SOURCES IN STREET CANYONS

The 2.5D ESM will now be used to investigate some aspects related to two types of sound sources of road traffic. The extreme cases of a single pass-by and a continuous traffic flow will be illuminated. Here we will consider the sound pressure only at a single point in the shielded canyon. In the transform integral in Eq. (5), no imaginary frequencies have been taken into account, and frequency resolutions in the transform of 0.1 Hz for rigid façades and 0.2 Hz for façades with absorption or diffusion patches have been used.

A. Vehicle pass-by

Figure 11 shows four curves of the excess attenuation level L_{EA} [the level minus the level without obstacle(s)] at observer point (49,0,0), due to a pass-by of a vehicle at $(9,y_s,0)$ [see Fig. 1 for the coordinates]. The results have been *A* weighted, and a road traffic noise spectrum has been used. A distribution of 90% light and 10% heavy vehicles has been assumed with a speed of 50 km/h and the road noise spectra have been taken from Danish measurement data.[25] The 1/3-octave bands 100–1000 Hz have been used. The curves represent a vehicle on a rigid ground surface in the following situations: a single building block as barrier (i.e., the central building block of Fig. 1), two parallel canyons with rigid façades (i.e., the "full" situation of Fig. 1), two parallel canyons with façade absorption, and two parallel canyons with partially diffusely reflecting façades [i.e., as in Fig. 10(a)]. The diffusion patches are rigid. Figure 10(b) shows the absorption coefficient for normal wave incidence used for the absorption patches. It originates from the absorption material used in a scale model study of parallel urban street canyons.[22]. The situation of a single building block is modeled using Pierce's solution for a thick barrier.[26]

The multiple reflections cause a level that is more than 20 dB (!) higher than a situation with a single wide barrier only. Diffusion and absorption have a clear effect on the level. The slope of the excess attenuation curves over distance—it increases for a large $|y|$ distance—is similar for all cases and governed by the changing diffraction angles. This implies that the level difference over the $|y|$ distance is smaller for the cases with barrier than without. It indicates why distant sources are important for sound pressure levels at a shielded side. The L_{EA} is larger for the shielded canyon case with rigid façades than in the free field at a large $|y|$ distance, an effect also described by Forssén and Hornikx.[27].

B. Coherent and incoherent line source

A traffic flow consisting of a continuous flow of vehicles with random distances may be modeled by an uncorrelated row of point sources, being an incoherent line source for predicting the equivalent level. Using a calculation model that generates the excess attenuation at the shielded side for a coherent line source might thus give incorrect results when aiming at predicting the excess attenuation due to road traffic noise from a traffic flow. Former studies of Duhamel[16] and Jean *et al.*[17] showed that calculating the excess attenuation of a single screen gives higher losses when modeling for a coherent line source instead of an incoherent line source. They showed that the excess attenuation of a screen due to an incoherent line source is smooth over frequency, whereas it showed a clear interference pattern for the coherent line source solution. Van Renterghem *et al.*[14] show an error in the same direction in the case of two parallel urban canyons. They did not, however, make use of point source solutions to assemble the incoherent line source solution. The excess attenuation for an incoherent line source can be found by integrating the squared absolute value of the calculated point source solutions $p_n(x,y,z,k)$ over y

$$L_{\mathrm{EA,inco,1}}(x,z,k) = 20\log\left(\frac{\int_{-\infty}^{\infty}|p_n(x,y,z,k)|^2\,dy}{\int_{-\infty}^{\infty}|p_{\mathrm{free}}(x,y,z,k)|^2\,dy}\right)$$

$$\approx 20\log\left(\frac{\sum_{n=-N}^{N}|p_n(x,n\Delta y,z,k)|^2\Delta y}{\sum_{n=-N}^{N}|p_{\mathrm{free}}(x,n\Delta y,z,k)|^2\Delta y}\right). \quad (14)$$

By Parseval's theorem, the energy of the integrals in Eqs. (4) and (5) are equal

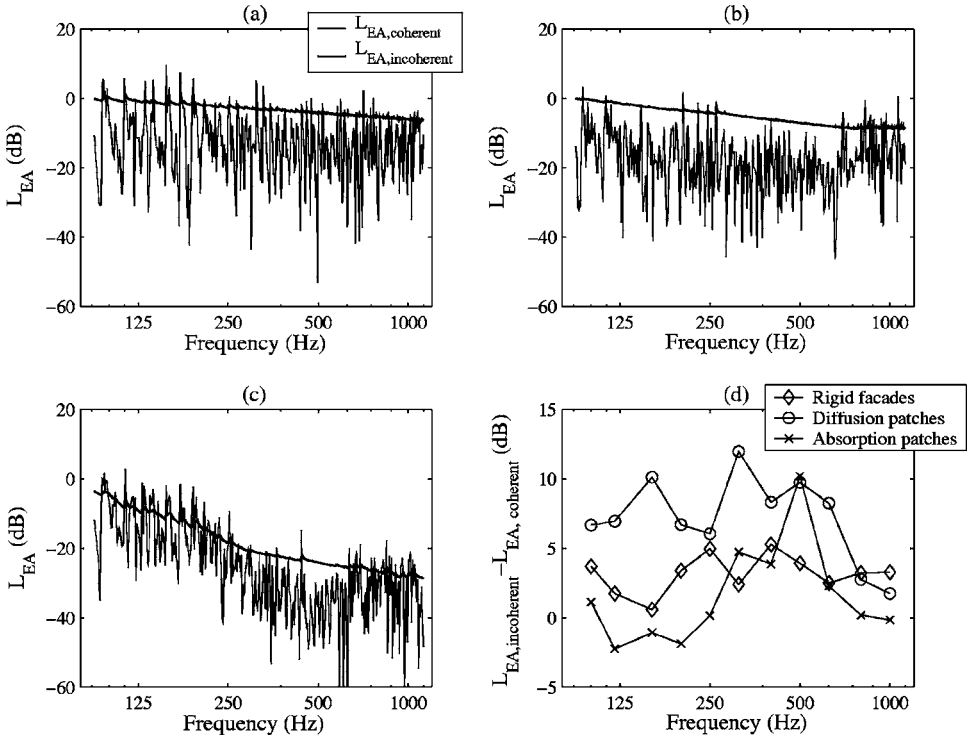


FIG. 12. L_{EA} results for the parallel street canyons case with the source at (9,0,0) and the observer at (49,0,0). The coherent line source solution is calculated with the 2D ESM and the incoherent line source solution according to Eq. (16): (a) rigid boundaries as in Fig. 1; (b) diffusion patches as in Fig. 10(a); (c) absorption patches as in Fig. 10; and (d) $L_{EA,incoherent} - L_{EA,coherent}$ for 1/3-octave bands.

$$\int_{-\infty}^{\infty} |p_n(x,y,z,k)|^2 \, dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} |q(x,K,z)|^2 \, dk_y. \tag{15}$$

Therefore, another way to calculate the incoherent line source solution is by using 2D solutions

$$L_{EA,inco,2}(x,z,k) = 20 \log \left(\frac{\frac{1}{2\pi} \sum_{n=-N}^{N} |q(x,\sqrt{k^2-(n\Delta k_y)^2},z)|^2 \Delta k_y}{\sum_{n=-N}^{N} |p_{free}(x,n\Delta y,z,k)|^2 \Delta y} \right). \tag{16}$$

Both expressions (14) and (16) converge when N goes to infinity and Δk_y and Δy go to zero. $L_{EA,inco,2}(x,z,k)$ can be calculated more efficiently by a coarser frequency discretization than the integral in Eq. (4), since the oscillating part is eliminated by taking the squared absolute values of q. When $|q|^2$ is a smooth function, e.g., as from Eq. (6), integration can be done very efficiently. For a resonant sound field however, the discretization has to be treated with more care. From Sec. III A, we know that at the shielded side sources at a larger y distance contribute to the level substantially. The integration limits have therefore to be rather large in Eq. (14). Equation (16) enables a largely reduced computational cost, and is therefore used for further calculations. At the shielded side, the major contribution in the summation in Eq. (16) comes from the lowest 2D frequencies, since they embody the first modes and are less reduced in amplitude by diffraction. The finest frequency discretization in the summand is thus required for the lowest frequencies. Therefore, the necessary frequency discretization is investigated where a fixed frequency discretization of $\Delta f = 0.2$ Hz for the lower frequency region up to 50 Hz has been chosen. The 1000 Hz 1/3-octave band results are shown in Table I for the cases with rigid façades, as in Fig. 1, and partially diffusely reflecting façades, as in Fig. 10(a). The errors have been calculated relative to a case with $\Delta f = 0.2$ Hz for the whole frequency range which gave a convergent result. A smaller Δf is necessary for the rigid case.

For calculations of an incoherent line source in a realistic environment (which has some absorption and diffusion), a proposal is to use $\Delta f = 0.2$ Hz up to $f = 88$ Hz and then 30 logarithmically spaced frequencies per 1/3-octave band. For the diffusion patches case, this approach yields a maximum error of 0.4 dB in the 1/3-octave band range 100–1000 Hz, whereas a maximum error of 0.9 dB is found for the rigid façade case. For the geometry of Fig. 10(a) with diffusion patches and the observer in the shielded canyon, the calculation time using this frequency discretization and frequency range is 23 h using the implemented 2.5D ESM method on a computer with a dual 2.66 GHz processor and a 4 GB memory.

The excess attenuation for a coherent line source and an incoherent line source for the parallel canyons of Fig. 1 are now compared for the cases of rigid boundaries, applied absorption patches, and applied diffusion patches. The cases with absorption and diffusion have again been modeled according to Fig. 10. The source position is at $(9,y,0)$ and the observer position at (49,0,0). As literature results for a single screen also show, the incoherent line source results are smoother over frequency than the coherent results; the interference effects are cancelled due to incoherence (see Fig. 12). In 1/3-octave bands, the coherent excess attenuation levels are lower than the incoherent attenuation levels for most cases and we see that the differences of the various cases differ by several dB. Even though these are only numerical results and evaluation together with measured data

has not been made here, it is clear that the influence of absorption and diffusion can be judged inaccurately using the coherent excess attenuation.

IV. CONCLUSIONS

The possession of point source solutions and incoherent line source solutions is of importance in urban sound propagation research when aiming to predict levels at shielded areas. By applying an inverse Fourier transform, coherent line source solutions of the two-dimensional equivalent sources method (2D ESM) were transformed to a point source solution for the situation of parallel urban canyons. This method is called the 2.5D ESM. The theoretical basis of the method should make it possible to model complex valued surface impedances. The integration paths of the transform integral over the imaginary frequencies were shown to be of importance for calculations in the directly exposed street canyon at the lower frequencies. In the shielded canyon, these integration paths may be omitted. The long travel distances by multiple reflections in the canyons impose a requirement of a high-frequency resolution for the evaluation of the transform integral. For a point source, a frequency discretization of $\Delta f = 0.4$ Hz gives satisfying results in the directly exposed canyon. In the shielded canyon for façades with some diffusion $\Delta f = 0.2$ Hz can be used, and $\Delta f = 0.1$ Hz should be used for rigid specular reflecting façades. Vehicle pass-by calculations show the importance of higher order reflections at the shielded canyon. The 2D solution can also be used for the calculation of an incoherent line source. This method is much more efficient regarding the computation time compared to a method making use of point source solutions. The excess attenuation for realistic traffic flows, modeled by incoherent line sources, shows that the effect of absorption and diffusion in street canyons may be predicted incorrectly when a coherent line source solution is used; i.e., the effect of mitigation may be overpredicted when modeling traffic along a straight road by a coherent line source.

ACKNOWLEDGMENTS

The work behind this paper has been supported by the Swedish Foundation for Strategic Environmental Research (MISTRA). The authors would like to thank Wolfgang Kropp for his ideas and fruitful discussions.

[1]K. V. Horoshenkov, D. C. Hothersall, and S. E. Mercy, "Scale modelling of sound propagation in a city street canyon," J. Sound Vib. **223**, 795–819 (1999).
[2]M. R. Ismail and D. J. Oldham, "A scale model investigation of sound reflection from building façades," Appl. Acoust. **66**, 123–147 (2005).
[3]J. Kang, "Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[4]J. Kang, "Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries," J. Sound Vib. **258**, 793–813 (2002).
[5]K. K. Lu and K. M. Li, "The propagation of sound in narrow street canyons," J. Acoust. Soc. Am. **112**, 537–550 (2002).
[6]J. Picaut, L. Simon, and J. Hardy, "Sound field modelling in a street with a diffusion equation," J. Acoust. Soc. Am. **106**, 2638–2645 (1999).
[7]J. Picaut and L. Simon, "A scale model experiment for the study of sound propagation in urban areas," Appl. Acoust. **62**, 327–340 (2001).
[8]J. Picaut, T. Le Polles, P. L'Hermite, and V. Gary, "Experimental study of sound propagation in a street," Appl. Acoust. **66**, 149–173 (2005).
[9]J. Kang, "Sound propagation in interconnected urban streets: A parametric study," Environ. Plan. B: Plan. Des. **28**, 281–294 (2001).
[10]J. Picaut, L. Simon., and J. Hardy, "Sound propagation in urban areas: A periodic disposition of buildings," Phys. Rev. E **60**, 4851–4859 (1999).
[11]M. Hornikx, "Towards a parabolic equation for modeling urban sound propagation," Proceedings 11th Long Range Sound Propagation Symposia, Fairly, Vermont, 2004.
[12]T. van Renterghem and D. Botteldooren, "Numerical simulation of sound propagation over rows of houses in the presence of wind," Proceedings of the 10th International Conference on Sound and Vibration, Stockholm, Sweden, June 2–3, 2003, 1381–1388.
[13]T. van Renterghem, E. Salomons, and D. Botteldooren, "Parameter study of sound propagation between city canyons with a coupled FDTD-PE model," Appl. Acoust. **67**, 487–510 (2006).
[14]M. Ögren, "Prediction of traffic noise shielding by city canyons," Ph.D. thesis, Chalmers University of Technology, Gothenburg, Sweden, 2004.
[15]T. Kihlman, *Handlingsplan Mot Buller (National Action Plan Against Noise)* (Allmänna Förlaget, Stockholm, Sweden, 1993).
[16]D. Duhamel, "Efficient calculation of the three-dimensional sound pressure field around a noise barrier," J. Sound Vib. **197**, 547–571 (1996).
[17]P. Jean, J. Defrance, and Y. Gabillet, "The importance of source type on the assessment of noise barriers," J. Sound Vib. **226**, 201–216 (1999).
[18]E. M. Salomons, A. C. Geerlings, and D. Duhamel, "Comparison of a ray model and a Fourier—boundary element method for traffic noise situations with multiple diffractions and reflections," Acust. Acta Acust. **226**, 35–47 (1997).
[19]D. Duhamel and P. Sergent, "Sound propagation over noise barriers with absorbing ground," J. Sound Vib. **218**, 799–823 (1998).
[20]L. Godinho, J. António, and A. Tadeu, "Sound propagation around rigid barriers laterally confined by tall buildings," Appl. Acoust. **63**, 595–609 (2002).
[21]J. Defrance and P. Jean, "Integration of the efficiency of noise barrier caps in a 3-D ray tracing method. Case of a T-shaped diffracting device," Appl. Acoust. **64**, 765–780 (2003).
[22]M. Hornikx, "Sound propagation to two-dimensional shielded urban areas," Licenciate of Eng., thesis, Chalmers University of Technology, Gothenburg, Sweden, 2006.
[23]S. N. Chandler-Wilde and D. C. Hothersall, "Efficient calculation of the Green's function for acoustic propagation above a homogeneous impedance plane," J. Sound Vib. **180**, 705–724 (1995).
[24]C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).
[25]H. Jonasson and S. Storeheier, "Nord 2000. New Nordic prediction method for road traffic noise," SP Rapport No. 2001:10, Borås, Sweden 2001.
[26]A. D. Pierce, "Diffraction of sound around corners and over wide barriers," J. Acoust. Soc. Am. **55**, 941–955 (1974).
[27]J. Forssén and M. Hornikx, "Statistics of A-weighted road traffic noise levels in shielded urban areas," Acta. Acust. Acust. **92**, 998–1008 (2006).

Vortex sound under the influence of a piecewise porous material on an infinite rigid plane

C. K. Lau[a] and S. K. Tang[b]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, P. R. China

(Received 2 May 2007; revised 17 August 2007; accepted 20 August 2007)

The vortex dynamics and the sound generation by an inviscid vortex in the presence of a finite length porous material on an otherwise rigid plane are studied numerically in the present study in an attempt to understand the sound generation near the surface of a wall lining in a lined duct. The combined effects of the effective fluid density and flow resistance inside the porous material, and the length and thickness of the porous material on the sound generation process are examined in detail. Results obtained demonstrate the sound pressure is longitudinal dipole and show how seriously the above-mentioned parameters are affecting the vortex sound pressure under the influence of the porous material. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783201]

PACS number(s): 43.28.Ra, 43.50.Nm, 43.50.Gf [GCL] Pages: 2542–2550

I. INTRODUCTION

Modern high rise buildings are usually heavily serviced. Building services equipment such as air handling units and fans are often used to provide an acceptable indoor environment to the building occupants. However, noise is generated at the same time as a result of flow instability containing regular fluctuations or turbulence.[1,2] It propagates into the interior of a building through the air conveying ductwork together with the low Mach number flow. The flow turbulence also interacts with any solid boundaries inside the ductworks and creates additional noise.[3] At low Mach number condition, which is the usual case in building services engineering, such turbulence-solid interaction is proved to be more important than that produced within the turbulent flow.[3]

The conventional practice to alleviate this kind of nuisance is to install dissipative silencers or wall linings.[4] These devices attenuate noise by the viscous actions within the porous material involved and the effective range of noise attenuation is between the middle to high frequencies. There are also studies related to the attenuation of sound by dissipative silencers in the presence of fluid flow (for instance, Cummings and Chang[5] and Peat and Rathi[6]), but the effects of flow turbulence are not addressed. However, Ffowcs Williams[7] revealed that noise can be generated when flow turbulence interacts with the sound absorbent linears. This result was further substantiated by Quinn and Howe.[8] Therefore, it is important to evaluate how this self-noise generation will reduce the performance of a dissipative silencer or wall lining.[9]

Flow turbulence is hard to model analytically. However, the low Mach number flow inside the ductwork makes it possible to treat the turbulence as discrete vortices which can be handled by the potential theory.[10] Though the vortices are very drastic simplifications of real flow turbulence, this semi-analytical analysis of vortex sound has attracted the attention of many researchers as it is expected to be able to provide insights into the much more complicated turbulence-solid surfaces aeroacoustics. Typical examples of such analysis include those of Cannell and Ffowcs Williams,[11] Obermeier and Zhu,[12] and more recently of Howe.[13] In addition, Crighton[14] and Obermeier[15] investigated the sound generation when the vortex interacts with an infinite half plane and a circular cylinder, respectively, by using the method of matched asymptotic expansions.

Most of the above studies are concerned with the sound generation by the unsteady vortex motion in the presence of rigid boundaries. Recently, Tang[16] studied theoretically the dynamics of an inviscid vortex in the presence of porous boundaries and ascribed that the change in the vortex velocity was the major source of self-generated noise in the dissipative silencer. This was later confirmed by the numerical results of an inviscid vortex engaging a flow impedance discontinuity at the edge of a wedge and on a flat surface.[17] Recently, Lau and Tang[18] studied the vortex sound in the presence of a porous cylinder and found that the sound field consisted of dipoles having strengths which can be greater than those in the rigid cylinder cases.

In the present investigation, the sound generation by an inviscid vortex in the presence of a porous material with finite length and thickness on an otherwise rigid plane is studied. Such configuration analogues to the situation of a wall lining in the air conditioning ductwork. It will be shown later that the present analytical method is less straightforward than those in the previous works of the authors. It is hoped that the present study will deepen the general understanding of the vortex sound generation and provide clues for the development of a more appropriate aeroacoustic model for a dissipative silencer or a lined duct.

II. THEORETICAL DEVELOPMENT

An inviscid vortex with circulation Γ initially located at $z_0(t)=(x_0(t),y_0(t))$ very far upstream of the piecewise porous

[a]Present address Ove Arup & Partners, Level 5, Festival Walk, Kowloon, Hong Kong, P.R. China.
[b]Author to whom correspondence should be addressed. Electronic mail: besktang@polyu.edu.hk

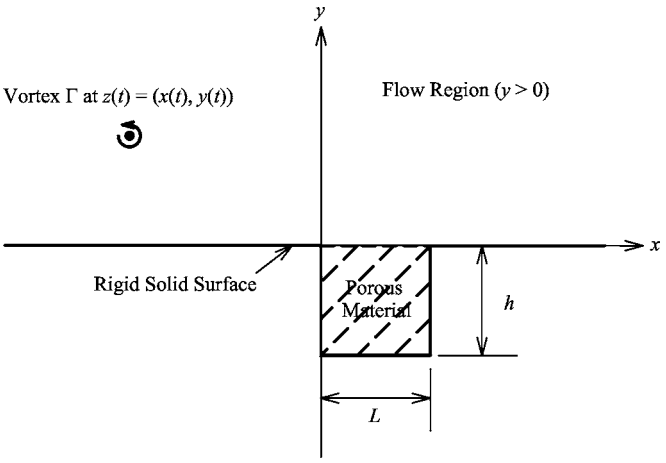


FIG. 1. Schematics for the vortex-porous material system.

material with length L and thickness h as shown in Fig. 1 is considered in the present study. The properties of the piecewise porous material are characterized by the effective fluid density ρ_e and the flow resistance R_f inside its lattice[19] while the density in the flow region ($y \geqslant 0$) is denoted by ρ. All the length scales in the present study are normalized by the initial vortex height $y_0(0)$ above the x axis. Also, the time scale t, the velocity of the vortex, and the flow resistance R_f are normalized by $y_0(0)^2/\Gamma$, $\Gamma/y_0(0)$, and $\rho\Gamma/y_0(0)^2$, respectively. It should be noted that the method of conformal mapping, which has been adopted in the previous works of the authors,[16–18] cannot be applied in the present case as an appropriate transformation function cannot be sought.

Denoting the streamfunctions in the fluid and in the porous regions by ψ and ψ_p, respectively, one finds inside the fluid region ($y \geqslant 0$)

$$\nabla^2 \psi = -\delta(x - x_0)\,\delta(y - y_0) \tag{1}$$

and within the porous material ($-h \leqslant y \leqslant 0$, $0 \leqslant x \leqslant L$)

$$\nabla^2 \psi_p = 0 \tag{2}$$

where ∇^2 and δ are the Laplacian operator and delta function, respectively. As the normal fluid velocities at the interfaces between the porous material and the rigid wall vanish, one can conclude that

$$\left.\frac{\partial \phi_p}{\partial x}\right|_{x=0} = \left.\frac{\partial \psi_p}{\partial y}\right|_{x=0} = \left.\frac{\partial \phi_p}{\partial x}\right|_{x=L} = \left.\frac{\partial \psi_p}{\partial y}\right|_{x=L} = -\left.\frac{\partial \psi_p}{\partial x}\right|_{y=-h} = 0, \tag{3}$$

where ϕ_p is the flow potential in the porous region. From Eqs. (2) and (3), it is straightforward to find that

$$\psi_p = \sum_{n=1}^{\infty} A_n e^{\alpha_n h} \sin(\alpha_n x)\sinh[\alpha_n(h + y)], \tag{4}$$

where n is a nonzero integer, $\alpha_n = n\pi/L$, and A_n the mode magnitude.

Taking the Fourier transform of Eq. (1) with respect to x as in Tang,[16] one obtains

$$\Psi = \int_{-\infty}^{\infty} \psi e^{ikx}\, dk = A e^{-|k|y} + B e^{|k|y}, \tag{5}$$

where A, B, are functions of k. The continuity of Ψ and the vorticity jump at $y = y_0$ lead to $B = (1/2|k|)e^{ikx_0 - |k|y_0}$.[16] On the porous boundary at $y = 0$ and $0 \leqslant x \leqslant L$, the continuity of normal fluid velocity gives

$$-\left.\frac{\partial \psi}{\partial x}\right|_{y=0} = -\left.\frac{\partial \psi_p}{\partial x}\right|_{y=0} \Rightarrow \Psi|_{y=0} = \Psi_p|_{y=0}, \tag{6}$$

where Ψ_p is the Fourier transform of ψ_p with respect to x. The relationship between A and A_n can be found by combining Eqs. (4)–(6)

$$A = \sum_{n=1}^{\infty} \alpha_n A_n e^{\alpha_n h} \sinh(\alpha_n h)\frac{(-1)^n e^{ikL} - 1}{k^2 - \alpha_n^2} - \frac{1}{2|k|} e^{ikx_0 - |k|y_0}. \tag{7}$$

The continuity of pressure on the porous boundary $y = 0$, $0 \leqslant x \leqslant L$ gives

$$\frac{\partial}{\partial t}\left(\left.\frac{\partial \psi}{\partial y}\right|_{y=0}\right) = \eta \frac{\partial}{\partial t}\left(\left.\frac{\partial \psi_p}{\partial y}\right|_{y=0}\right) + R_f \left.\frac{\partial \psi_p}{\partial y}\right|_{y=0}, \tag{8}$$

where $\eta = \rho_e/\rho$. The application of inverse Fourier transform to Eq. (8) suggests that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} |k|(\dot{B} - \dot{A}) e^{-ikx}\, dk = \sum_{n=1}^{\infty} (\eta \dot{A}_n + R_f A_n)\alpha_n e^{\alpha_n h} \times \sin(\alpha_n x)\cosh(\alpha_n h), \tag{9}$$

where $\cdot$ denotes differentiation with respect to time. After some algebra, Eq. (9) can be expressed as

$$-\frac{1}{\pi}\sum_{n=1}^{\infty} \alpha_n \dot{A}_n e^{\alpha_n h}\sinh(\alpha_n h)\int_0^L I_n(x)\sin(\alpha_m x)dx - \frac{\dot{x}_0}{\pi}\int_0^L \frac{2y_0(x_0 - x)\sin(\alpha_m x)}{[y_0^2 + (x_0 - x)^2]^2}dx + \frac{\dot{y}_0}{\pi}\int_0^L \frac{(x_0 - x)^2 - y_0^2}{[y_0^2 + (x_0 - x)^2]^2}\sin(\alpha_m x)dx = \frac{1}{2}(\eta \dot{A}_m + R_f A_m)\alpha_m e^{\alpha_m h}\cosh(\alpha_m h), \tag{10}$$

where $I_n = \cos(\alpha_n x)[ci(\alpha_n x) - ci(\alpha_n(L - x))] + \sin(\alpha_n x)[si(\alpha_n x) + si(\alpha_n(L - x)) + \pi]$ and m is a nonzero integer. si and ci represent the sine and cosine integrals, respectively. Details and properties of these integrals can be found in standard references, such as Abramowitz and Stegun,[20] and thus are not described here. One can also find that

$$\int_0^L I_n(x)\sin(\alpha_m x)dx$$

$$= \begin{cases} \dfrac{2\alpha_m}{\alpha_m^2-\alpha_n^2}[ci(m\pi)-ci(n\pi)] & \text{if } m+n \text{ is even}, \quad m\neq n \\ \dfrac{\pi L}{2}+\displaystyle\int_0^L si(\alpha_n x)dx & \text{if } m=n \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

On the other hand, the longitudinal and transverse velocities of the vortex are, respectively,[16]

$$\dot{x}_0 = \left.\frac{\partial(\psi - Be^{|k|y})}{\partial y}\right|_{x=x_0,y=y_0}$$

$$= \frac{1}{\pi}\sum_{n=1}^{\infty}\alpha_n A_n e^{\alpha_n h}\sinh(\alpha_n h)\int_0^L \frac{(x_0-x)\cos(\alpha_n x)}{(x_0-x)^2+y_0^2}dx + \frac{1}{4\pi y_0} \tag{12}$$

and

$$\dot{y}_0 = -\left.\frac{\partial(\psi - Be^{|k|y})}{\partial x}\right|_{x=x_0,y=y_0}$$

$$= \frac{1}{\pi}\sum_{n=1}^{\infty}\alpha_n A_n e^{\alpha_n h}\sinh(\alpha_n h)\int_0^L \frac{y_0\cos(\alpha_n x)}{(x_0-x)^2+y_0^2}dx. \tag{13}$$

The integrals in Eqs. (12) and (13) correspond to the flow field induced by the normal velocity at $y=0$, $0\leqslant x\leqslant L$. The time variations of the vortex position (x_0,y_0) and the modes magnitude A_n can be estimated by using Eqs. (10), (12), and (13) with the standard fourth-order Runge–Kutta integration method. Initially, $A_n\equiv 0$ for all n.

The streamfunction ψ can be derived by the inverse Fourier transform of Eq. (5) together with the help of Eq. (7)

$$\psi = \frac{1}{2\pi}\sum_{n=1}^{\infty}\alpha_n A_n e^{\alpha_n h}\sinh(\alpha_n h) \times\int_{-\infty}^{\infty}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}e^{-|k|y}e^{-ikx}\,dk - \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{1}{|k|}e^{-ik(x-x_0)-|k|(y+y_0)}\,dk + \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{1}{|k|}e^{-ik(x-x_0)+|k|(y-y_0)}\,dk. \tag{14}$$

The flow potential can then be evaluated through the use of the Cauchy–Rieman principle:[21]

$$\phi = \frac{1}{2\pi}\sum_{n=1}^{\infty}\alpha_n A_n e^{\alpha_n h}\sinh(\alpha_n h) \times\int_{-\infty}^{\infty}\frac{1}{ik}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}|k|e^{-|k|y}e^{-ikx}\,dk - \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{1}{ik}e^{-ik(x-x_0)-|k|(y+y_0)}\,dk - \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{1}{ik}e^{ik(x-x_0)-|k|(y-y_0)}\,dk + C, \tag{15}$$

where C is the integration constant. Since the flow potential vanishes when $|z|\to\infty$, $C=0$. Using the formula tabulated in Gradshteyn *et al.*,[22] the flow potential becomes

$$\phi = \frac{1}{2\pi}\sum_{n=1}^{\infty}\alpha_n A_n e^{\alpha_n h}\sinh(\alpha_n h)\int_{-\infty}^{\infty}\frac{1}{ik}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2} \times|k|e^{-|k|y}e^{-ikx}\,dk + \frac{1}{2\pi}\tan^{-1}\frac{y-y_0}{x-x_0} - \frac{1}{2\pi}\tan^{-1}\frac{y+y_0}{x-x_0} \tag{16}$$

The integral term in Eq. (16) represents the flow potential induced by the porous material while the other is the flow potential induced by the infinite rigid plane. The far-field inner potential at large $|z|$ is

$$\phi = \left[-\frac{2}{\pi}\sum_{n=1,3,5,\ldots}^{\infty}\frac{1}{\alpha_n}A_n e^{\alpha_n h}\sinh(\alpha_n h) - \frac{1}{\pi}r_0\sin\theta_0\right]\frac{\cos\theta}{r}+O(r^{-1}), \tag{17}$$

where (r_0,θ_0) is the polar coordinates of the inviscid vortex and (r,θ) is a point in the flow field. Thus, following the work of Lau and Tang,[18] the far-field sound pressure p at large distance r is

$$p = \frac{1}{\pi}\sqrt{\frac{1}{2cr}}\frac{\partial}{\partial t}\int_{-\infty}^{t_0-r/c}\frac{\partial}{\partial\tau}\left[\frac{2}{\pi}\sum_{n=1,3,5,\ldots}^{\infty}\frac{1}{\alpha_n}A_n e^{\alpha_n h} \times\sinh(\alpha_n h)+\frac{1}{\pi}r_0\sin\theta_0\right]\frac{d\tau}{\sqrt{t_o-\tau-r/c}}\cos\theta, \tag{18}$$

where c and t_o are the speed of sound and the far-field observer time, respectively. The far-field sound pressure is normalized by $\rho\Gamma/y_0(0)^2$. It can be shown from Eq. (18) that the far-field sound pressure consists of longitudinal dipoles. The overall acoustical energy radiated per unit spanwise length, E, is estimated by an integration of p^2 over a half circle of radius r centered at $(L/2,0)$

$$E = \frac{1}{c}\int_{-\pi}^{\pi}p^2 r\,d\theta. \tag{19}$$

In the present study, the effect of a mean flow in the flow region is not considered. However, it has been shown by Ffowcs Williams and Lovely[23] and recently further con-

firmed by Tang *et al.*[24] that the low Mach number mean flow tends to amplify the overall sound power radiation and the amount of amplification can be inferred from their results. The present study can be regarded as the fundamental case.

III. RESULTS AND DISCUSSION

In the present investigation, the far-field sound pressure is obtained at large r, and the directivity of the sound pressure is dipole like. The vortex dynamics are obtained from Eqs. (10), (12), and (13) with the appropriate number of mode A_n and the far-field sound pressure p obtained from Eq. (18). Ten terms for A_n are enough for a reliable solution (not shown here).[25] Thus, only results obtained with the first ten A_n terms are presented in this paper.

For the case of an infinite rigid plane, the longitudinal and transverse velocities of the vortex from Eqs. (12) and (13) converge to $\dot{x}_0=1/4\pi y_0$ and $\dot{y}_0=0$. The inviscid vortex propagates with a constant velocity in the longitudinal direction and generates no sound. p tends to zero at large R_f or large η (not shown here). The density ratio η in practice is usually below 5.[19] A detailed discussion on the practice range of R_f is given in Lau and Tang[18] and thus is not repeated here.

A. Perfectly inviscid fluid

For a perfectly inviscid fluid, the flow resistance vanishes ($R_f=0$). Figure 2(a) shows the effects of L and η for $h=2$ on the vortex path. The inviscid vortex experiences the pressure-releasing effect due to the porous material[16] and bends toward the porous material for $-2<x<L/2$. It gradually propagates back to its original height afterward due to the pressure-supporting effect of the rigid plane at $y=0$, $x \geqslant L$. The vortex path is basically symmetrical about the vertical $x=0.5L$ plane. The higher the η, the less pressure-releasing effect from the porous material, resulting in less severe bending towards the porous material. The effects of η on the vortex path vanish for large η.

When the length of the porous material is increased from $L=1$ to $L=2$ at a fixed η, the degree of vortex path bending towards the porous material increases. This situation becomes more serious when the length of the porous material is further increased to $L=10/3$. It is due to the longer duration of the interaction between the inviscid vortex and the porous material. The vortex undergoes a substantially larger rate of change of velocity under such a prolonged interaction as shown in Figs. 2(b)–2(e). In these figures, t_a denotes the time at which the vortex passes across the vertical plane containing the upstream edge of the porous material ($x=0$). The larger the value of η, the smaller the magnitudes of $\dot{x}_0$ and $\dot{y}_0$. The magnitudes of $\dot{x}_0$ and $\dot{y}_0$ increase with L [Figs. 2(b) and 2(c)]. One can observe that the vortex accelerations are increased by either decreasing η or increasing L [Figs. 2(d) and 2(e)].

Figure 2(f) shows the time variations of the longitudinal dipole pressure p at $h=2$ (with the directivity factor $\cos\theta$ removed). The dipole magnitude increases as η decreases due to the amplification of the vortex accelerations [Figs. 2(d) and 2(e)], and the dipole magnitude increases with increasing L. A longer duration of active sound generation is also observed at increased L. It is due to the earlier movement of the inviscid vortex and longer duration under the influence of the porous material.

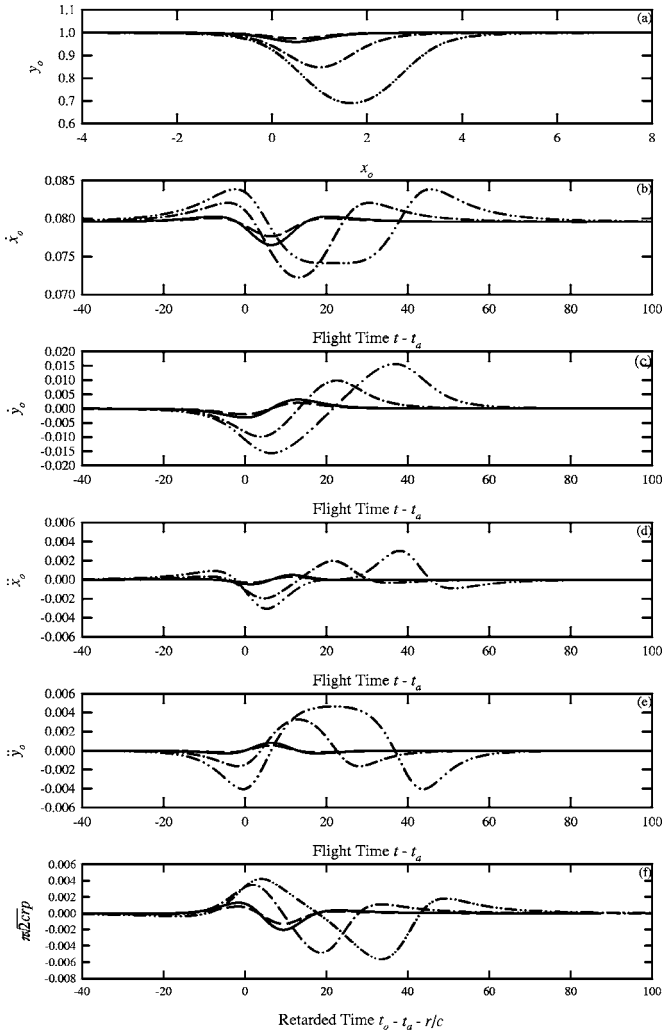


FIG. 2. Effect of pressure-releasing surface on the vortex motion, velocity, acceleration, and the sound generation at a fixed $h=2$: (a) vortex path; (b) longitudinal velocity; (c) transverse velocity; (d) longitudinal acceleration; (e) transverse acceleration; and (f) dipole sound pressure; (—) $L=1$, $\eta=3$; (---) $L=1$, $\eta=5$; (-·-) $L=2$, $\eta=3$; (-···-) $L=10/3$, $\eta=3$.

At a fixed L and η, the inviscid vortex propagates in a path closer to the porous material because of the stronger pressure-releasing effect when a thicker porous material is installed [Fig. 3(a)]. The larger the value of h, the more serious bending towards the porous material, but such an effect becomes insignificant when h is increased beyond 0.8 for $L=1$ and $\eta=3$. One can expect that this critical h value increases with increasing L and/or η (shown later). Again, the magnitudes of vortex velocities and accelerations increase at increased h (not shown here). An amplification of p is observed as h increases [Fig. 3(b)] because of the more rigorous vortex accelerations resulted from the stronger pressure-releasing effect of the thicker porous material. However, the amplification of p reaches its maximum at $h \approx 1$ as shown in Fig. 4(a).

Figures 4(a) and 4(b) summarize the combined effects of L, h, and η on the dipole magnitude and the radiated acous-

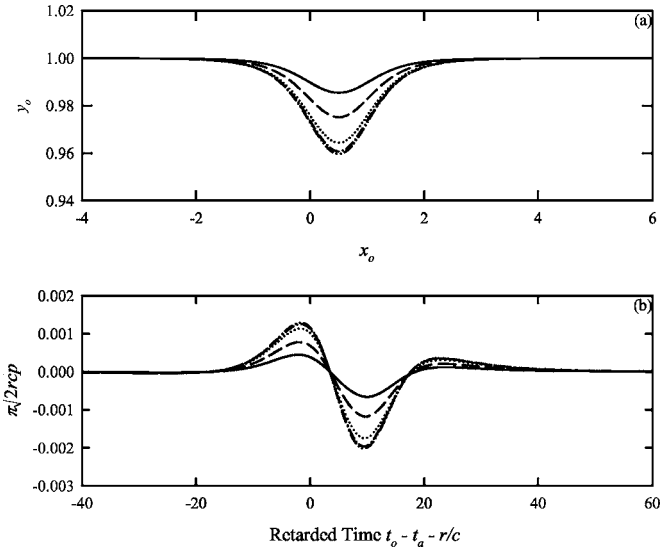


FIG. 3. Effect of porous material thickness on the vortex motion and the sound generation: (a) vortex path; and (b) dipole sound pressure; (—) $h=0.1$; (---) $h=0.2$; (.....) $h=0.4$; (–·–) $h=0.6$; (-·-) $h=0.8$; $L=1$ and $\eta=3$.

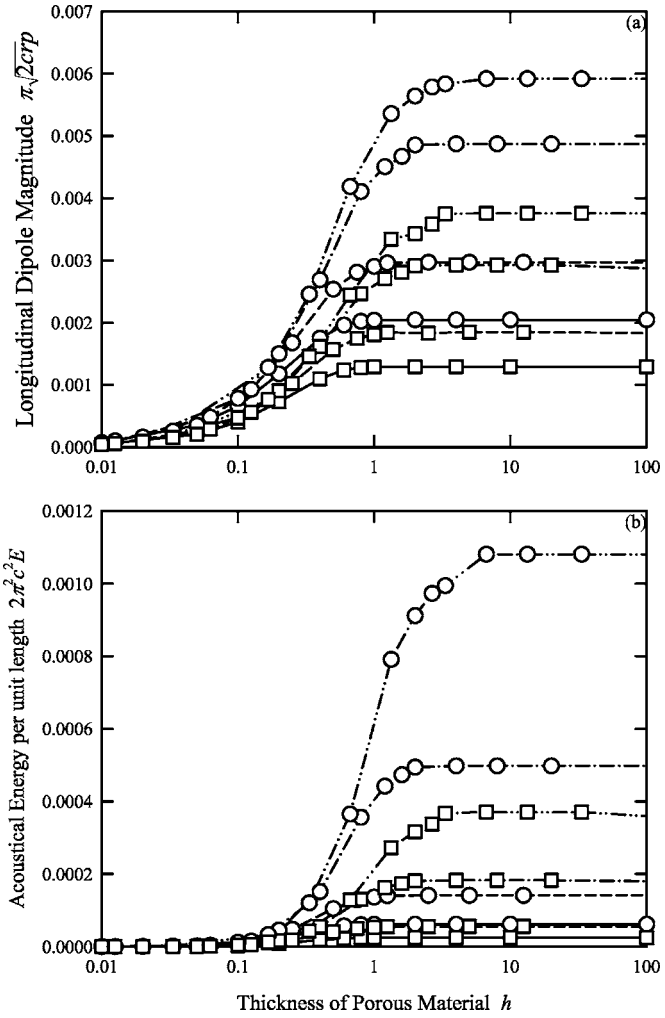


FIG. 4. Combined effects of L and h on vortex sound radiation in a perfectly inviscid fluid: (a) longitudinal dipole magnitude; and (b) acoustical energy per unit spanwise length; (—) $L=1$; (--) $L=1.25$; (–·–) $L=2$; (– · –) $L=10/3$; (○) $\eta=3$; (□) $\eta=5$.

tical energy per unit spanwise length for a perfectly inviscid fluid, respectively. In general, the dipole magnitude increases when L or h increases. When h is small, for instance when $h=0.01$, the influence of the porous material on the vortex accelerations is very weak, resulting in a lower dipole magnitude. Increase in the dipole magnitude is concentrated within the range of h from 0.1 to 1 [Fig. 4(a)]. The dipole magnitude is not affected much by further increasing h. It is probably due to the gradually diminishing effect from the rigid surface at the bottom of the porous material at $y=-h$. One can notice that there is still a slight increase in dipole magnitude for $h>1$ when $L=2$ and $L=10/3$ [Fig. 4(a)]. The overall trend is that the critical h tends to increase with L, but such increase is insignificant in the range of L studied. It can also be observed that a higher dipole magnitude is obtained under a stronger pressure-releasing effect at smaller η. However, the value of the critical h appears not to be affected by a change of η from 3 to 5. The acoustical energy radiated in a perfectly inviscid fluid exhibits a pattern similar to the dipole magnitude [Fig. 4(b)]. The longer duration of active sound generation and the higher dipole magnitude result in a large increase of the overall acoustical energy radiated at increased L or reduced η.

B. Combined effects of η and R_f

When the flow resistance R_f is finite, the effective fluid density ρ_e and the flow resistance R_f inside the lattice of the porous material will produce pressure-releasing and pressure-supporting effects, respectively.[16] Figure 5(a) shows their combined effects on the vortex motions for $L=1$, $h=1$, and $\eta=3$ under various R_f. The situation for the perfectly inviscid fluid ($R_f=0$) is also shown in the figure for the sake of comparison. For $R_f=0.05$, the vortex bends toward the porous material for $-2<x<0.5$ because of the pressure-releasing effect, and then moves upward for $0.5<x<1.2$ due to the presence of the rigid wall. For $x>1.2$, the inviscid vortex propagates in a path lower than those in the cases for the rigid wall condition and the perfectly inviscid fluid. A similar phenomenon was also observed in Lau and Tang[18] where the inviscid vortex will not resume its original height after interacting with the porous half cylinder mounted on an otherwise rigid plane when R_f is small. In addition, the vortex path is not symmetrical about $x=0.5$. When $R_f=0.1$, the vortex propagates more closely to its original height for $-2<x<1.2$ but for $x>1.2$, it moves closer to the horizontal axis when compared with that for $R_f=0.05$. This situation of the vortex path for $x>1.2$ is reversed once the flow resistance is increased to above $R_f=1$. The vortex gradually rises back to its initial height as R_f is further increased, and soon recovers to that under the rigid wall condition for $R_f=10$. At large R_f, the pressure-supporting effect is so strong that it overcomes the pressure-releasing effect, A_n tends to zero, and becomes less influential to the vortex velocity [Eqs. (12) and (13)] such that the path of the inviscid vortex matches that for the rigid wall case.

Figures 5(b)–5(e) show the corresponding time variations of vortex velocities and accelerations for $L=1$, $h=1$, and $\eta=3$ at different R_f. The magnitude of $\dot{x}_0$ increases as R_f

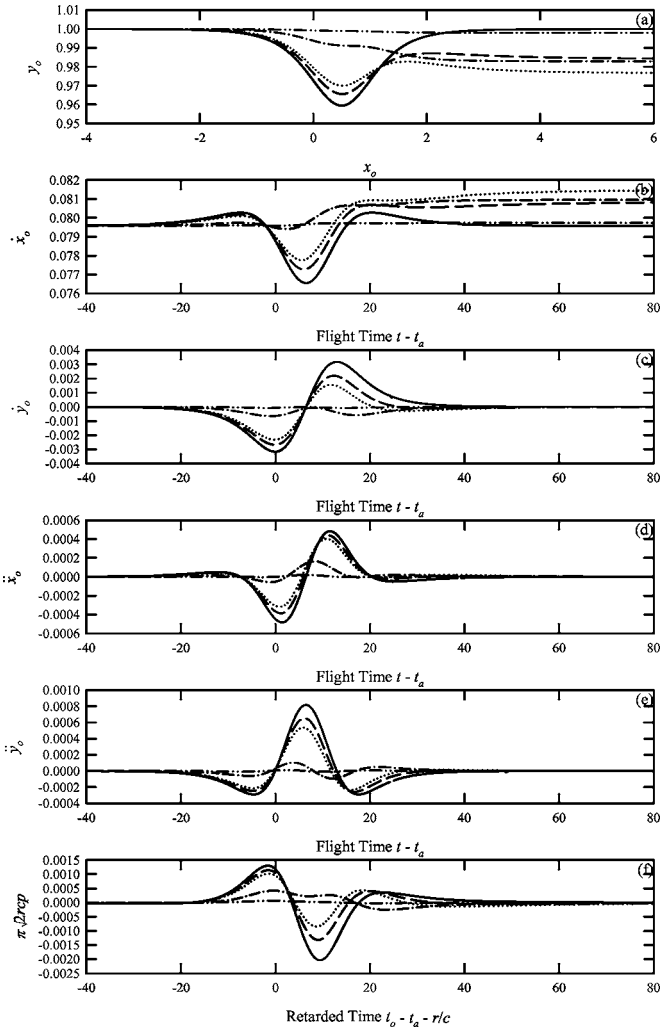


FIG. 5. Combined effects of effective fluid density and flow resistance on the vortex motion, velocity, acceleration, and the sound generation at a fixed $L=1$, $h=1$ and $\eta=3$: (a) vortex path; (b) longitudinal velocity; (c) transverse velocity; (d) longitudinal acceleration; (e) transverse acceleration; and (f) dipole sound pressure; (—) $R_f=0$; (---) $R_f=0.05$; (····) $R_f=0.1$; (-·-) $R_f=1$; (–· · –) $R_f=10$.

increases from 0 to 0.1 for $t-t_a>20$ [Fig. 5(b)] (that is, for $x_0>1.2$ when the vortex is downstream of the porous material). However, $\dot{x}_0$ decreases with increasing R_f for $0.1<R_f<10$ and is close to that with the rigid wall condition when $R_f=10$. The magnitude of $\dot{y}_0$ decreases with increasing R_f [Fig. 5(c)]. The vortex acceleration increases with decreasing R_f [Figs. 5(d) and 5(e)]. Figure 5(f) shows some examples of the time variation of p at $L=1$, $h=1$, and $\eta=3$ at various R_f. The dipole magnitude is the highest for the perfectly inviscid fluid case. It decreases with increasing R_f. The magnitudes of the first p crest and trough decrease while those of the second crest increase as R_f increases from 0 to 0.1. For $R_f>1$, the dipole magnitude decreases as the vortex accelerations become weaker. The properties of the porous material are dominated by the pressure-supporting effect at increased R_f which tends to weaken the unsteady motion of the vortex. The sound radiation becomes insignificant when R_f is increased to 10.

Figure 6(a) shows the vortex dynamics for various combinations of h and η for $L=1$ and $R_f=1$. The smaller the

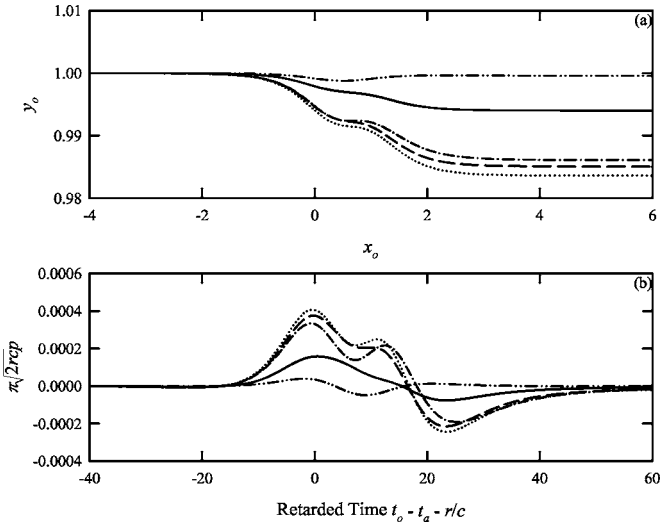


FIG. 6. Effects of h and η on the vortex path and the sound generation for nonvanishing R_f: (a) vortex motion; and (b) sound pressure; $L=1$ and $R_f=1$; (---) $h=0.1$, $\eta=3$; (——) $h=0.4$, $\eta=3$; (·····) $h=0.6$, $\eta=3$; (–·–) $h=0.6$, $\eta=5$; (–··–) $h=0.6$, $\eta=100$.

value of η, the greater the bending towards the porous material for a fixed h. This serious bending in the vortex path is due to the strong pressure-releasing effect at small η. The vortex path converges to that for the infinite rigid plane condition for $\eta=100$. Similar to the perfectly inviscid fluid case, the degree of bending increases with increasing h, and the vortex motion is not affected much by further increasing h. The effect of increasing h is similar to that of decreasing η (increasing the pressure-releasing effect). The effects of varying h and η with $R_f=1$ on the vortex dynamics are essentially similar to those presented in the perfectly inviscid fluid case [Figs. 2(a) and 3(a)] and thus are not presented. Figure 6(b) shows the time variations of p at different combinations of h and η for $L=1$ and $R_f=1$. The dipole magnitude decreases as h decreases or η increases. When h increases, a longer duration of a compression wave radiation is observed.

Figure 7(a) shows the effects of various combinations of L and R_f on the vortex dynamics at $h=2$ and $\eta=5$. When the flow resistance is fixed, the vortex propagates in a path closer to the porous material with a longer L, resulting in a higher vortex acceleration (especially the longitudinal component) when the vortex leaves the porous material at $x\sim L$. For $L\leqslant 2$ at a fixed R_f, the vortex dynamics [Fig. 7(a)] and the time variations of the vortex velocities [Figs. 7(b) and 7(c)] resemble those shown in Figs. 5(a)–5(c), respectively. However for a longer L, the magnitude of the longitudinal vortex acceleration in this case increases as R_f increases from 0 to 0.45 and then decreases upon further increase in R_f, while that of the transverse vortex acceleration decreases monotonically with increasing R_f [Figs. 7(d) and 7(e)]. The effects of increasing L and R_f thus appear to be counteracting each other. When the porous material is sufficiently long, the vortex is allowed to move closer to the porous material and can keep a weak longitudinal acceleration if the porous material is not very pressure releasing because of the induction effect. However, a strong R_f tends to prohibit the downward vortex

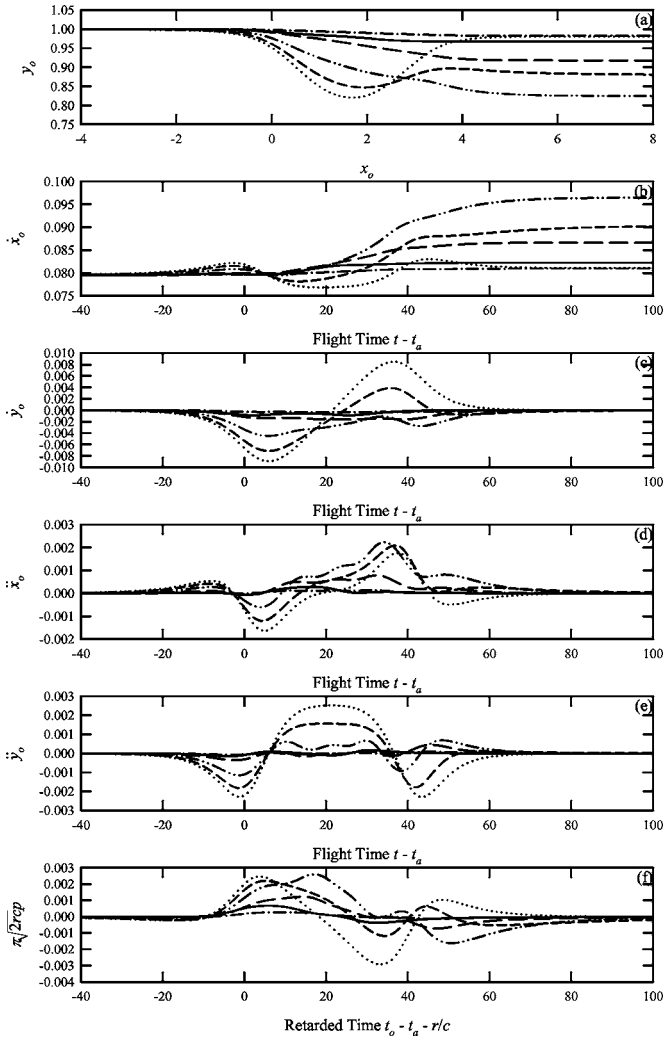


FIG. 7. Vortex motion and time variation of vortex velocity, acceleration, and sound generation of the inviscid vortex at a fixed h and η; $h=2$ and $\eta=5$; (a) vortex path; (b) longitudinal velocity; (c) transverse velocity; (d) longitudinal acceleration; (e) transverse acceleration; and (f) dipole sound pressure; (—) $L=2$, $R_f=2$; (·····) $L=10/3$, $R_f=0.01$; (----) $L=10/3$, $R_f=0.09$; (–·–·) $L=10/3$, $R_f=0.36$; (——) $L=10/3$, $R_f=2$; (–··–) $L=10/3$, $R_f=9$.

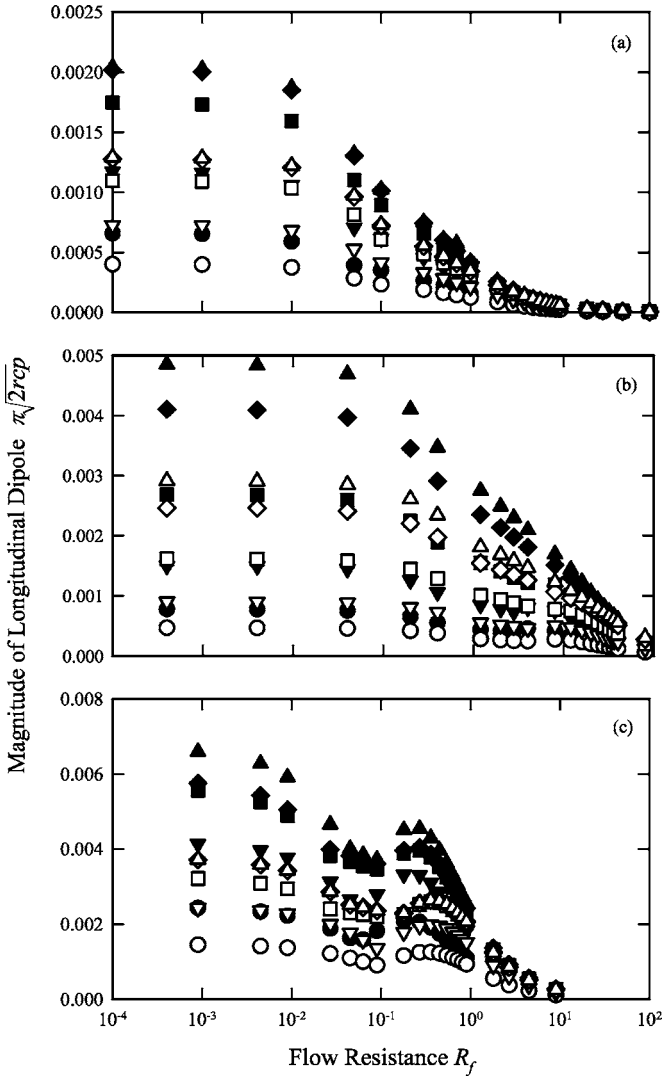


FIG. 8. Variations of longitudinal dipole magnitude with different L, h, η, and R_f: (a) $L=1$; (b) $L=2$; (●) $h=0.1$; (▼) $h=0.2$; (■) $h=0.4$; (♦) $h=0.8$; (▲) $h=2$; (c) $L=10/3$; (●) $h=1/3$; (▼) $h=2/3$; (■) $h=2$; (♦) $h=10/3$; (▲) $h=20/3$; closed symbols for $\eta=3$, open symbols for $\eta=5$.

movement and thus reduces the vortex acceleration when the vortex is at $x\sim L$. Since the dipole magnitude is closely related to the vortex acceleration, one can expect that the general trend is that the dipole magnitude decreases with increasing R_f but decreasing L. However, the decrease in dipole magnitude is not monotonic [Figs. 7(f) and 8] when L is sufficiently long.

The dipole magnitudes at different L, h, η, and R_f are summarized in Fig. 8. As shown in Fig. 8(a), the increase in the thickness h of the porous material increases the dipole magnitude for $L=1$, but the dipole magnitude is not affected by further increasing h beyond 1. The thickness h does not affect the vortex dynamics [Fig. 7(a)] and the vortex accelerations [Figs. 7(d) and 7(e)] very much. The dipole magnitude is higher for $\eta=3$ than for $\eta=5$, and it decreases more than two orders as R_f increases from 0 to 100. A more rapid decrease in the dipole magnitude is observed from $R_f=0$ to $R_f=10$ when L is small.

The dipole magnitudes for $L=2$ are presented in Fig. 8(b). The dipole magnitude increases when L is increased from 1 to 2 for a fixed η and R_f. The vortex propagates with high acceleration under the pressure-releasing effect of the porous material, resulting in an amplification of the dipole. This is the consequence of a longer duration of interaction between the inviscid vortex and the porous material. Comparing the effects of thickness of the porous material at $L=1$ and $L=2$, an increase of the dipole magnitude is observed for h deeper than 1 at the longer L case.

The corresponding information for the dipole magnitude with $L=10/3$ is given in Fig. 8(c). At this L, local maxima are observed at $R_f\sim 0.25$ for $\eta=3$ and at $R_f\sim 0.45$ for $\eta=5$ with the reason given during the discussion of Figs. 7(a)–7(e) three paragraphs before. The corresponding acoustical energy radiation is summarized in Fig. 9. The energy variation pattern is similar to those presented in Fig. 8 in general. However, one can observe a slight plateau or a weak local maximum of the acoustical energy E for $1<R_f<10$ for $L=2$ [Fig. 9(b)]. Such phenomenon is not observed for the dipole magnitude variations shown in Fig. 8(b). The increase

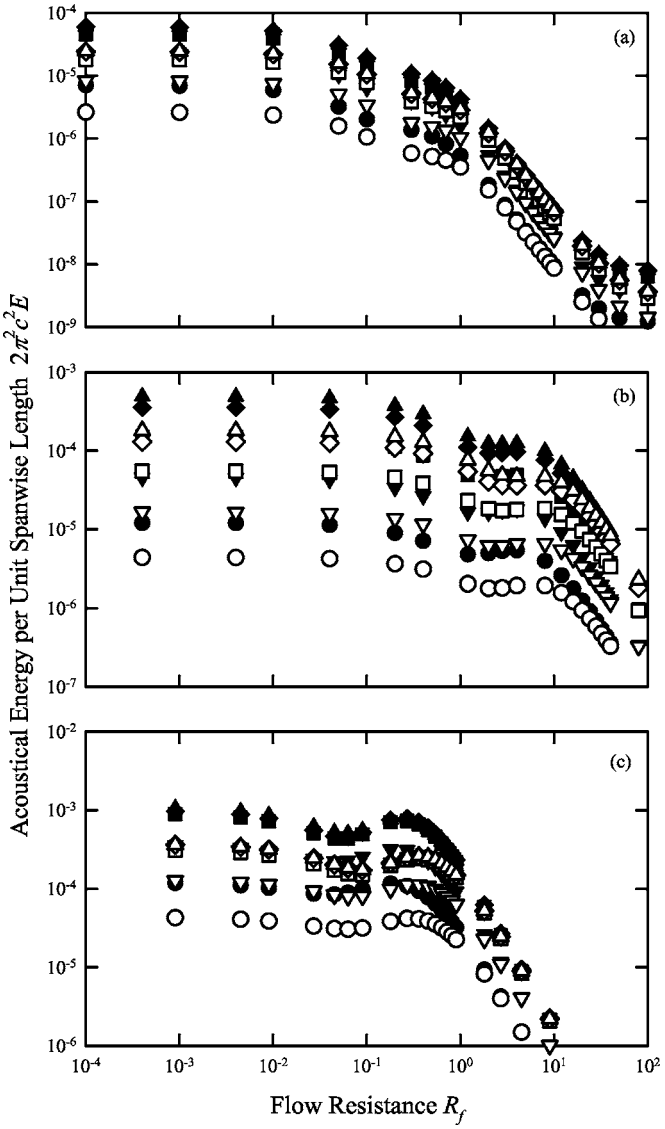


FIG. 9. Acoustical energy radiated with different L, h, η, and R_f: (a) $L=1$; (b) $L=2$; and (c) $L=10/3$; Legends: same as those in Fig. 8.

in L also prolongs the active interaction between the vortex and the porous material. At some combinations of R_f, h, and η, the effect of such an increase in the sound radiation period on the overall energy radiation does compensate that from dipole magnitude reduction due to stronger R_f.

IV. CONCLUSIONS

The vortex sound generation in the presence of a piecewise porous material on an otherwise infinite rigid plane has been studied in a configuration analogous to the internal boundary of a lined duct. The streamfunctions inside the fluid medium and the porous material were derived and the standard fourth-order Runge–Kutta method was used to solve the coupled vortex dynamic equations. The far-field sound pressure was evaluated by the method of matched asymptotic expansions.

When an inviscid vortex engages a finite length porous material, the sound pressure radiated is a longitudinal dipole. The sound pressure magnitude increases as the effective fluid density decreases because of the strong pressure-releasing effect of the porous boundary. The sound magnitude basically decreases with stronger flow resistance inside the porous material, but it does not decrease monotonically with increasing flow resistance when the length of the porous is long enough. There appears to be some combinations of parameters where a slightly stronger sound energy radiation is observed with increasing R_f before the energy drops again as the latter is further increased. This critical R_f range shifts higher when the material length or the effective fluid density inside the porous material are increased, but the range is much less sensitive to the change of material thickness provided that the porous material is not too thin.

It is also observed that the effect of thickening the porous material is in general similar to that of lengthening it. A thicker porous material produces a stronger sound, but the sound magnitude will eventually reach its upper bound when the material is so thick that the vortex motion is no longer affected by further increasing its thickness for a fixed porous material length. This critical material thickness increases with the material length but is not sensitive to the change of the effective fluid density within the practical range of the latter.

ACKNOWLEDGMENT

This study is supported by a grant from the Research Grant Council, The Hong Kong Special Administration Region, People's Republic of China (Project No. PolyU5030/00E).

[1]M. J. Lighthill, "On sound generated aerodynamically I. General theory," Proc. R. Soc. London, Ser. A **A211**, 564–587 (1952).
[2]H. G. Davies and J. E. Ffowcs Williams, "Aerodynamic sound generation in a pipe," J. Fluid Mech. **32**, 765–778 (1968).
[3]N. Curle, "The influence of solid boundaries upon aerodynamic sound," Proc. R. Soc. London, Ser. A **A231**, 505–514 (1955).
[4]C. M. Harris, *Handbook of Noise Control* (McGraw-Hill, New York, 1979).
[5]A. Cummings and I.-J. Chang, "Sound attenuation of a finite length dissipative flow duct silencer with internal mean flow in the absorbent," J. Sound Vib. **127**, 1–17 (1988).
[6]K. S. Peat and K. L. Rathi, "A finite element analysis of the convected acoustic wave motion in dissipative silencers," J. Sound Vib. **184**, 529–545 (1995).
[7]J. E. Ffowcs Williams, "The acoustics of turbulence near sound-absorbent liners," J. Fluid Mech. **51**, 737–749 (1972).
[8]M. C. Quinn and M. S. Howe, "On the production and absorption of sound by lossless liners in the presence of mean flow," J. Sound Vib. **97**, 1–9 (1984).
[9]L. L. Beranek and I. L. Vér, *Noise and Vibration Control Engineering. Principles and Applications* (Wiley, New York, 1992).
[10]L. M. Milne-Thomson, *Theoretical Hydrodynamics* (The University Press, Macmillan, Glasgow, 1968).
[11]P. Cannell and J. E. Ffowcs Williams, "Radiation from line vortex filaments exhausting from a two-dimensional semi-infinite duct," J. Fluid Mech. **58**, 65–80 (1973).
[12]F. Obermeier and K. Q. Zhu, "Sound generation by rotor-vortex interaction in low Mach number flow," J. Aircr. **30**, 81–87 (1993).
[13]M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, Cambridge, 2003).
[14]D. G. Crighton, "Radiation from vortex filament motion near a half plane," J. Fluid Mech. **51**, 357–362 (1972).
[15]F. Obermeier, "The influence of solid bodies on low Mach number vortex sound," J. Sound Vib. **72**, 39–49 (1980).
[16]S. K. Tang, "Effects of porous boundaries on the dynamics of an inviscid vortex filament," Q. J. Mech. Appl. Math. **54**, 65–84 (2001).
[17]S. K. Tang and C. K. Lau, "Vortex sound in the presence of a wedge with

inhomogeneous surface flow impedance," J. Sound Vib. **281**, 1077–1091 (2005).
[18]C. K. Lau and S. K. Tang, "Sound generated by vortices in the presence of a porous half-cylinder mounted on a rigid plane," J. Acoust. Soc. Am. **119**, 2084–2095 (2006).
[19]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[20]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables* (Wiley, New York, 1972).
[21]R. V. Churchill and J. W. Brown, *Complex Variables and Applications* (McGraw-Hill, New York, 1990).
[22]I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffery, *Tables of Integrals, Series and Products* (Academic, Boston, 1994).
[23]J. E. Ffowcs Williams and D. J. Lovely, "Sound radiation into uniformly flowing fluid by compact surface vibration," J. Sound Vib. **71**, 689–700 (1975).
[24]S. K. Tang, R. C. K. Leung, and R. M. C. So, "Vortex sound due to a flexible boundary backed by a cavity in a low Mach number mean flow," J. Acoust. Soc. Am. **121**, 1345–1352 (2007).
[25]C. K. Lau, "Vortex sound generation in the presence of porous materials with an application to dissipative silencers and lined ducts," Ph.D. thesis, The Hong Kong Polytechnic University, Hong Kong, 2006.

Small-slope simulation of acoustic backscatter from a physical model of an elastic ocean bottom

Raymond J. Soukup[a)]
United States Naval Research Laboratory, 4555 Overlook Avenue, Washington, DC 20375-5350

Gaetano Canepa
NATO Undersea Research Centre Viale San Bartolomeo 400, 1916 La Spezia, Italy

Harry J. Simpson, Jason E. Summers, and Robert F. Gragg
United States Naval Research Laboratory, 4555 Overlook Avenue, Washington, DC 20375-5350

(Received 20 December 2006; revised 1 August 2007; accepted 10 August 2007)

An underwater acoustic experiment with a two-dimensional rough interface, milled from a slab of PVC, was performed at a tank facility. The purpose was to verify the predictions of numerical models of acoustic rough surface scattering, using a manufactured physical model of an ocean bottom that featured shear effects, nonhomogeneous roughness statistics, and root-mean-square roughness amplitude on the order of the acoustic wavelength. Predictions of the received time series and interface scattering strength in the 100–300 kHz band were obtained from the Bottom Reverberation from Inhomogeneities and Surfaces–Small-Slope Approximation (BORIS-SSA) numerical scattering model. The predictions were made using direct measurements of scattering model inputs—specifically, the geoacoustic properties from laboratory analysis of material samples and the grid of surface heights from a touch-trigger probe. BORIS-SSA predictions for the amplitude of the received time series were shown to be accurate with a root-mean-square residual error of about 1 dB, while errors for the scattering strength prediction were higher (2–3.5 dB). The work is part of an ongoing effort to use physical models to examine a variety of acoustic scattering and propagation phenomena involving the ocean bottom.
[DOI: 10.1121/1.2783116]

PACS number(s): 43.30.Hw, 43.30.Gv [KGF] Pages: 2551–2559

I. INTRODUCTION

The study of acoustic interactions with ocean boundaries utilizes theories of scattering from rough surfaces to characterize the interaction of sound with the ocean surface and bottom (the air-water and water-bottom interfaces). For characterizing the acoustic interaction of an underwater transmitted signal with the ocean boundaries, a number of theories have been used to make predictions of the received time series and/or the scattering strength. These theories have been applied over a wide range of frequencies and the full range of scattering geometries in three dimensions. Some theories (e.g., perturbation and small slope[1,2]) require limitations on the roughness, while others (e.g., wedge assemblage[3,4]) place limitations on the sound speed of the ocean bottom. The problem reaches a high degree of complexity when one is dealing with an ocean bottom for which spatially varying geoacoustic parameters (density, compressional wave-speed, shear-wave speed, compressional-wave attenuation, and shear-wave attenuation) and/or discrete inclusions in the sediment play a significant role. In the latter case, the scattering from the interface is just one component of the acoustic interaction with the ocean bottom.

For elastic ocean bottoms such as limestone, mudstone, or basalt, the parameters of the rough surface (e.g., the exponent and strength of the roughness spectrum) may be the most important predictors of the scattering strength as a function of the incident/scattering angles and frequency. In cases where inclusions or layering do not contribute significantly to the scattered field, it is possible to predict the scattering behavior at subcritical grazing angles for an elastic bottom using a numerical interface scattering model.[5] Elastic interfaces may result from the removal of sediments by geological processes involving rapid ocean currents producing exposed rock, as seen off the Carolina Coast,[6] or may reflect an overall diversity of bottom types that are encountered in the littoral, as seen in the North Atlantic.[7] For the elastic bottom problem, one can employ small-slope or perturbation models that include the shear-wave speed and attenuation as inputs. For example, when one performs small-slope backscatter predictions for a limestone bottom with a power-law roughness spectrum, the overall level depends heavily on the parameters of the roughness spectrum, while some individual details depend on the shear speed. In the general case where the roughness is heterogeneous (i.e., has spatially varying statistics) or does not follow a power-law behavior over the wave numbers of interest, one may elect to use a numerical model for which the rough surface itself, rather than its spectral parameters, is the input. Such a situation is addressed in this paper. The motivation for considering this level of complexity is the nature of the littoral ocean environment for sonar operations in the 10–100 kHz range. The bottom relief may be a combination of power-law roughness and periodic

[a)]Electronic mail: soukup@abyss.nrl.navy.mil

ripples, subject to changes over a period of hours due to biological processes.[8] Therefore, we require a prediction tool that encompasses a relatively high level of complexity for interface scattering, including shear properties, ability to model rough surfaces with heterogeneous statistical behavior, and small-slope calculations that are accurate for roughness on the order of the acoustic wavelength.

Given that the predicted dependence of backscattering on roughness parameters needs to be accurately verified, the use of manufactured physical models is a natural approach. In this paper, the term "physical model" will refer to a fabricated object with a rough surface that serves as an analog (usually with comparable geoacoustic parameters but smaller length scales) to an ocean bottom, while "numerical model," "scattering model," or simply "model," will indicate a physics-based scattering prediction algorithm that is invoked in practice by executing a computer program. The physical models are placed in tank facilities where the positions of the source and receiver are robotically controlled to produce a variety of experimental geometries. In this way one can examine a number of rough surfaces that differ in their roughness properties but not in their geoacoustic properties—a degree of experimental control that presumably cannot be achieved in the real ocean. The methodology for creating mathematical realizations of rough surfaces and producing physical manifestations of these realizations using milling equipment is described in Summers *et al.*[9] The physical model can be constructed from a material without layers or inclusions that might produce volume scattering. It can also be made of sufficient thickness to minimize returns from the back side and thus essentially simulate an infinite half-space. Physical models with specific power-law roughness spectral parameters are currently being fabricated from polyvinyl chloride (PVC) slabs at the Naval Research Laboratory (NRL).

Physical models with rough surfaces have been employed in a number of applications. Horton *et al.*[10] constructed polystyrene models of rough ocean bottoms in the 1960s, based on the autocorrelation properties of land topography, to test scattering theory for a pressure release surface. Chu and Stanton[11] used a Styrofoam™ model to simulate a pressure-release surface and test a numerical model of scattering from three-dimensional protuberances distributed across the rough surface. Williams *et al.*[12] used acrylic scale models of ice ridges to address a similar scattering problem from protuberances. Mellema[13] milled a power-law rough surface on a Plexiglas™ sample to investigate subcritical angle penetration, using inverted geoacoustic properties of the material to account for the arrivals of scattered returns from the interface, lateral waves traveling along the interface, and returns from the flat rear face of the physical model. Sessarego[14] added an oil layer to a physical model with a rough surface to obtain scattering strengths and test the feasibility of detecting spilled oil on the sea floor.

In this paper, we describe a water tank experiment designed to address the problem of interface scattering. The experimental results are used to verify the predictions of quasimonostatic backscatter by a small-slope scattering model in the 100–300 kHz band. The manufactured physical model is a PVC slab with a milled rough surface. This physical model is a "found object" from the Allied Geophysical Laboratory/University of Houston whose roughness spectral properties were *a priori* unknown. The physical model works well as a test case for a rough-surface scattering model with deterministic profilometry as input, as its roughness can be treated (to sub-wavelength scales) as fluctuations of height relative to a flat surface. The numerical scattering model we use for deterministic rough surfaces is the Bottom Reverberation from Inhomogeneities and Surface – Small-Slope Approximation (BORIS-SSA) code developed at the NATO Undersea Research Centre.[15] The code produces a predicted time series of sound pressure level from a small-slope calculation. This time series can then be processed to produce scattering strength by calculating the ensonified area and assuming spherical spreading for calculation of the transmission loss. BORIS-SSA calculations of scattering strength have been compared with integral-equation and Kirchhoff approximations for benchmark surfaces that have been mathematically generated.[16] These benchmark surfaces were significantly smoother than the one analyzed in this paper.

The following sections of the paper describe an experimental verification of predictions of the BORIS-SSA model, using the measured profilometry of the physical model as input. Section II describes the design and execution of the tank experiment, Section III provides more detail on the BORIS-SSA model, and Sec. IV presents the experimental results with discussion.

II. TANK EXPERIMENT

The acoustic experiment, called GPM RSS2 (Geoacoustic Physical Modeling Rough Surface Scattering experiment 2), was performed at a water tank facility (the NRL Shallow Water Acoustic Laboratory) in May 2004. The acoustic data and documentation of the experimental procedures were included in a Naval Research Laboratory memorandum report in CD format by Soukup *et al.*[17] The rough surface is 122 $\times$ 122 cm in size, with a mean thickness of 6.2 cm. Close examination at the length scale of the acoustic wavelength of the experiment (5–15 mm) indicates that roughness statistics of the physical model are on the whole heterogeneous. The root-mean-square roughness was 9 mm, which is on the order of the wavelength. The two-dimensional grid of heights, obtained using a touch-trigger probe with a spacing of 1 mm, is shown in Fig. 1. Details of the profilometry measurements for the rough surface are given in a memorandum report,[17] and an overall description of a touch-trigger probe method for the general surface is given in Summers *et al.*[18] A close-up photograph of the physical model is given in Fig. 2, which shows artifacts (grooves) that result from the manufacturing process. These grooves are 0.02–0.05 mm peak-to-trough and are separated by a nominal distance of 1 mm.

The transmitted signal was a shaped pulse with an approximate length of 20 μs and an approximately flat spectrum in the band 100–300 kHz. This frequency band is high enough to produce minimal returns from the flat rear face of the block (opposite to the rough surface) because of the attenuation in the material. It is also low enough to ensure that

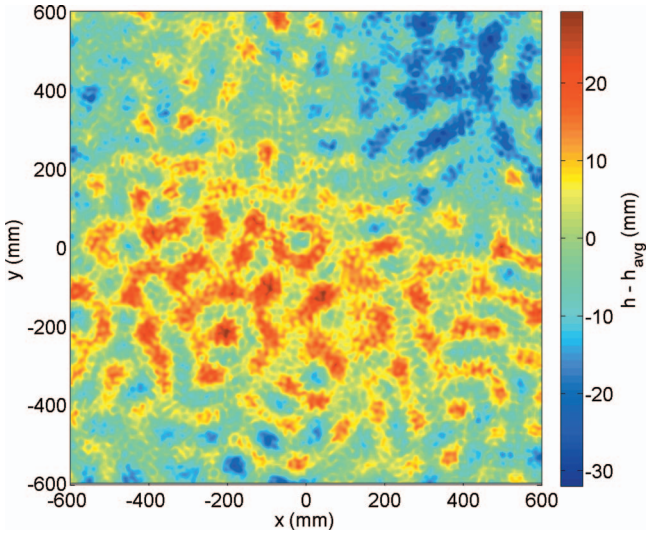


FIG. 1. Heights (relative to the overall average height) on the University of Houston physical model as a function of position. The sampling interval is 1 mm in both directions for a total of 1.44×10^6 sample points. The nine dots represent "target positions" that are described in the discussion of the experimental setup.

the spacing of the profilometry grid is subwavelength. Figures 3 and 4 show the temporal and spectral characteristics of the transmitted signal, obtained from a reference measurement with a hydrophone at a distance of 1 m from the source. The source and receiver were both RESON 4034 transducers. Their omnidirectionality was verified to within ±1 dB in a small tank designed for measurement of the source level as a function of angle.

Compressional/shear speeds and attenuations were measured by Dr. De-Hua Han at the Houston Advanced Research Center (HARC) on a 35 mm×50.25 mm×52.29 mm sample of the PVC material from the same manufacturer.[19] Table I is a summary of the estimated results at 300 kHz.[20] The error values are derived from the following percentage errors for these measurements: 0.5% for the speeds, 2% for the compressional-wave attenuation, and 4% for the shear-wave attenuation.

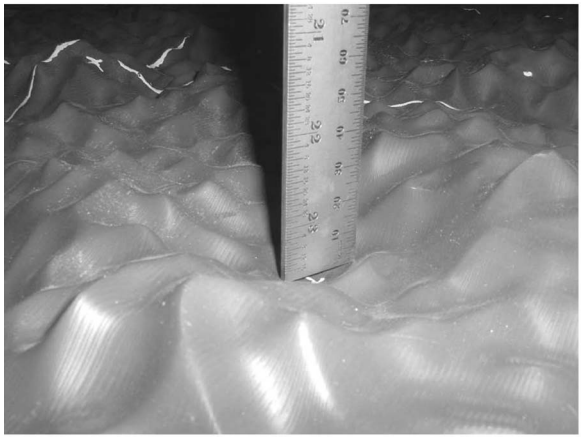


FIG. 2. Close-up photo of part of the University of Houston physical model. The ruler is graduated in inches on the left and millimeters on the right. The parallel grooves along the surface are artifacts of the milling process.

TABLE I. Material properties of the University of Houston PVC block at 300 kHz.

Parameter	Value
Density	1700 kg/m^3
Compressional speed	2381±11 m/s
Compressional attenuation	0.25±0.005 dB/m/kHz
Shear speed	1117±6 m/s
Shear attenuation	0.56±0.02 dB/m/kHz

The milled PVC slab was suspended vertically in a water tank. The water in the tank was 4.0 m in depth, and the tank walls were 10.7 m apart in the direction perpendicular to the slab and 7.6 m apart in the horizontal direction parallel to the slab. The large size of the tank, compared with the distances of the experimental geometry, removes the problem of reflections from the walls, air-water interface, and tank bottom. The sound speed and density of the tank water were 1482 m/s and 1.00 kg/m^3, respectively. The temperature may have varied a maximum of 1° around 20 °C, which would imply a range of values of roughly 1481–1483 m/s.

For the results shown in the remainder of this paper, except where otherwise specified, the source and receiver were positioned at distances of 51.4±0.1 cm and 43.4±0.1 cm from the flat back of the sample, respectively, as shown in Fig. 5. The source and receiver were placed on a common horizontal axis that intersected the surface of the sample at a series of nine points, which will be referred to as "target positions" (the dots superimposed on the surface relief in Fig. 1). The convention used here is that a negative value for the target position relative to center means lowering the PVC slab deeper into the water. Time series of the measured pressure at the receiver, consisting of 8192 points taken at a sampling interval of 0.3333 μs (i.e., 3 million samples/s) were obtained and a time series for subsequent analysis was obtained by averaging over 100 time series acquired at each source/receiver position.

III. BORIS-SSA SCATTERING MODEL

BORIS-SSA predicts the received sound pressure level for acoustic transmissions that have been scattered from the sea surface or sea floor. The model generates predictions of the received time series using a second- or fourth-order small-slope theory. The fourth-order theory involves a more sophisticated treatment of the slope of the rough surface, involving higher-order derivatives and other multiplicative factors that are advantageous for characterizing low-grazing angle scattering for surfaces of sufficient complexity. The formalism for describing the order of a small-slope calculation is given by Thorsos and Broschat.[21] BORIS-SSA uses a two-dimensional matrix of surface heights as a model input. The code can generate these surface heights internally according to a particular mathematical formulation (e.g. power-law, Pierson-Moskowitz). Alternatively, the surface heights may be input from measured profilometry data, such as that obtained, for example, from stereophotographic[22] or conductivity probe[23] measurements, or (as in this paper) from a touch-trigger probe. In addition, a ripple field or other deter-

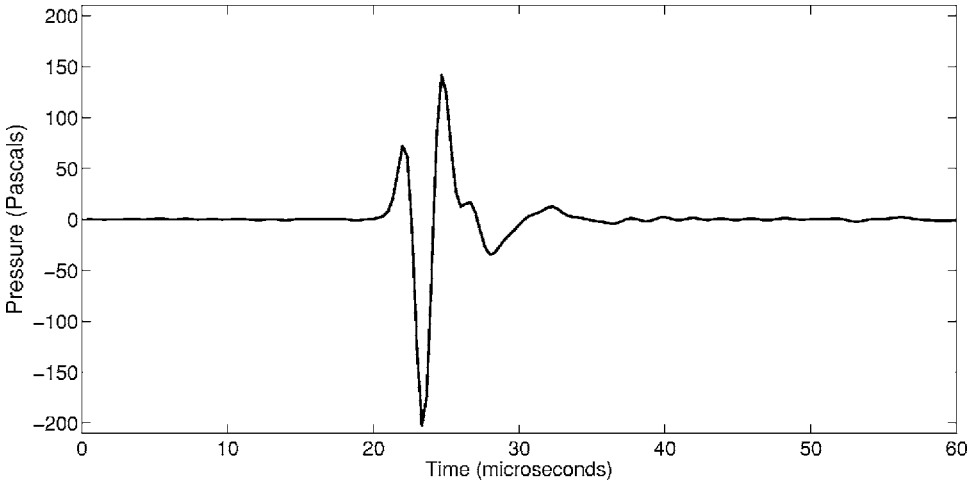


FIG. 3. Received signal (time domain) from the reference measurement, with the source and receiver separated by 1 m. This signal was retransmitted throughout the experiment.

ministic features can be superimposed on the profilometry measurements. For the University of Houston physical model, this capability was used to estimate the contribution to the scattering prediction from the artifacts of the manufacturing process (i.e., the small grooves in Fig. 2). The simulations indicated that these artifacts have a negligible effect on the received time series. This was an expected result, as the heights of the grooves are (at most) 1/100th of the smallest wavelength used in the experiment.

The geoacoustic inputs to BORIS-SSA are density, compressional speed and attenuation, and shear speed of the sea floor.[24] For predictions of the interface scattering in this experiment, the values of these inputs are taken from Table I. BORIS-SSA can also perform calculations of scattering due to slab volume inhomogeneity, using root-mean-square deviations and vertical and horizontal correlation lengths of the density and sound speed as input. X-ray analysis of samples of the Type I Gray PVC used in this experiment showed no inclusions larger than 1/100th of the wavelength.[25] Given this degree of slab volume homogeneity, the volume scattering component was assumed to be negligible. The results of the experiment indicate no consistent underprediction, as might result from omitting significant noninterface scattering. Additionally, the laboratory analysis of material samples was performed for each of three orthogonal axes, producing no variability with respect to measurement axis that would have resulted from the effects of layering.

BORIS-SSA can be used to perform scattering calculations for a full range of bistatic geometries, taking into account the transmission/reception characteristics of the source and receiver (source level, pulse shape, and directivity patterns) and their positions relative to the rough surface. In this paper, only the quasimonostatic geometries of Fig. 5 are considered, with a relatively small difference (maximum of 4°) between the transmitted and scattered angles. To simplify the presentation of results, a single "grazing angle" is calculated by averaging the transmitted and scattered angles.

In the results presented in the following, the measured scattering strength is calculated from the predicted time series by computing the ensonified area (for this experiment, an annulus) for each point in the time series, taking into account the transmitted signal, and calculating the spherical spreading loss. For all simulations, the transmitted signal of Fig. 3, with the amplitude corrected for the 1 m distance in the reference measurement, was used as input. The grazing angle is computed for each point on the time series and it is this angle that is used to plot the scattering strengths.

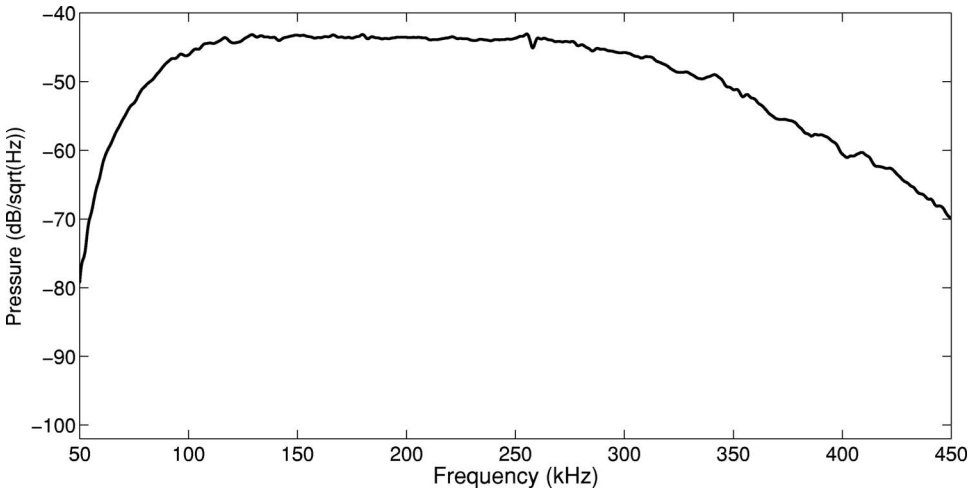


FIG. 4. Received signal (frequency domain) from the reference measurement, with the source and receiver separated by 1 m. This is the squared magnitude of the Fourier transform of the time series of Fig. 3.

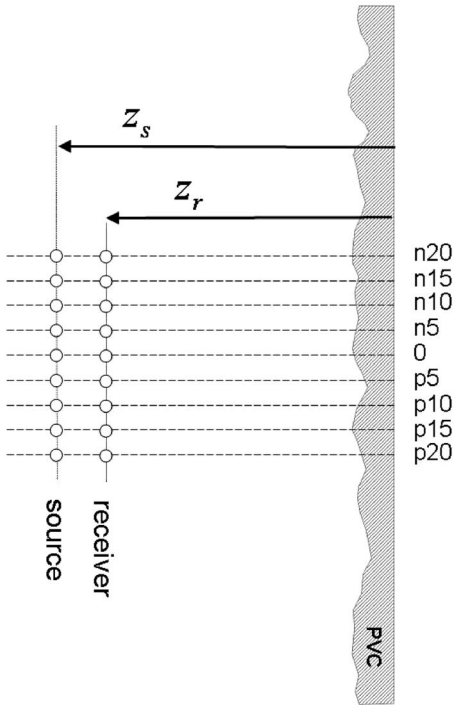


FIG. 5. Diagram (side view, not to scale) of the positions of the source and receiver relative to the physical model. The target positions are n20, n15, n10, n5, 0, p5, p10, p15, p20, going from top to bottom. In the actual experiment, the physical model, rather than the source and receiver, was moved to change the target position. The letters p and n refer to positive (upward) and negative (downward) moves of the physical model resulting in the target positions shown in the diagram. The number after n or p indicates the vertical displacement in centimeters relative to the sample center, e.g., n20 means moving the sample downward 20 cm. For the results shown in this paper, z_s=51.4 cm and z_r=43.4 cm.

IV. RESULTS AND DISCUSSION

A. Experimental results for target position averages

Representative pressure amplitudes (versus time) and scattering strengths for the whole surface were obtained by computing averages of these quantities over the nine target positions shown in Fig. 5. The averaging was performed on the values of the squared magnitude of the received pressure, and the resulting average values were converted to decibels for plotting. The same averaging was performed for BORIS-SSA predictions with second-order and fourth-order small-slope scattering calculations. The results are shown in Fig. 6, where the x-axis values represent the elapsed time after the beginning of the transmitted signal. The received level before the arrival of the scattered returns, as seen in the red curve from 500 to 525 μs, indicates the noise level.[26] The noise level can be obtained for the time frame corresponding to the scattering from the physical model by performing a measurement with the source and receiver farther away from the sample. The noise maintains an amplitude of about 5 dB in the time interval shown in Fig. 6, so some of the values at the very end of the time series (>750 μs) can be affected. The second-order BORIS-SSA calculation underpredicts the measurement as time progresses (and grazing angle decreases). The fourth-order prediction agrees well with the overall measurement and its fluctuations over time. For the comparison of the measured data and fourth-order prediction, the root-mean-square value of the residuals (an estimate of the standard error, given the small number of curves in the average) is 1.25 dB, using the time interval 550–800 μs. The limitation of the time series to 800 ms after transmit is due to the finite size of the physical model.

Scattering strengths were obtained from the time series from each position and the values of these curves were averaged and converted to decibels for plotting. This resulted in the data/model comparison as a function of grazing angle shown in Fig. 7, which includes both second- and fourth-order predictions. The lower limit of the x axis represents the smallest possible grazing angle that can be obtained, due to the finite extent of the physical model. The fourth-order prediction is clearly superior—a deviation of the second-order

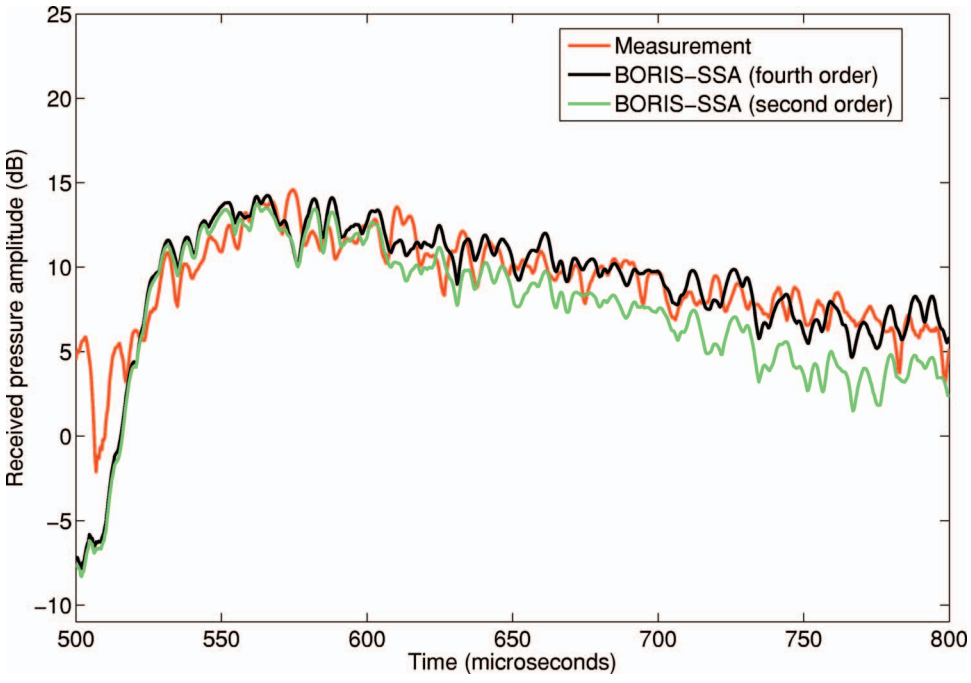


FIG. 6. Squared amplitude of the measured pressure (red) and BORIS-SSA simulations (green for second order; black for fourth order). The time series for the nine target positions were averaged as described in the text. The origin of the x axis is the start of the transmitted signal. The first acoustic interactions with the PVC model occurred between 500 and 550 μs after transmit.

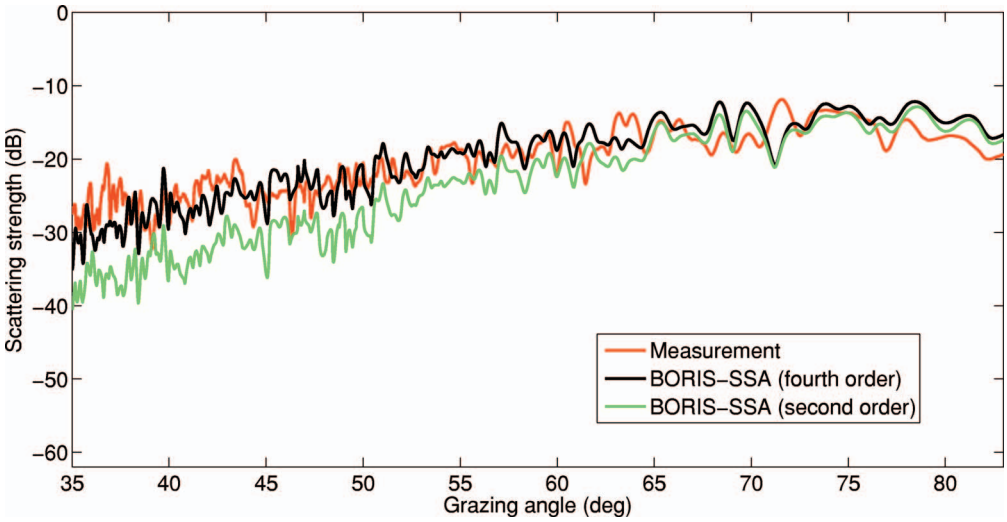


FIG. 7. Scattering strengths as a function of grazing angle, averaged over the nine target positions of Fig. 5 as described in the text. For grazing angles below 40°, the received level is on order of the noise level and there is the potential for underprediction by the numerical model.

small-slope prediction from the data for grazing angles less than about 50° is typical of this surface. (For the remaining graphs in the paper, only the fourth-order BORIS-SSA prediction is shown.) From the water tank speed and the material properties given in Table I, a compressional wave critical grazing angle of 51° is obtained (there is no shear wave critical angle, the shear speed being less than the water speed). For both the measured data and BORIS-SSA simulation, there seem to be no special features in the scattering curve that are associated with the compressional-wave critical angle. The root-mean-square value of the residuals for the comparison in Fig. 7 of the data with the fourth-order prediction is 2.3 dB.

The treatment of the reflection coefficient in BORIS-SSA can present problems when the signal is wideband, and it is suggested for general use that the model be applied for signals with a bandwidth that is less than 10% of the center frequency. To address this issue in our experiment, additional results were obtained by using the measured pulse in the simulation and passing the data and the predicted time series through a second-order Butterworth filter. To prevent phase distortion, the filtered sequence was reversed and run back through the filter. Center frequencies of 150, 200, and 250 kHz and a filter bandwidth of 60 kHz were used. Since the measured data and BORIS-SSA predictions were very similar for all three frequencies, only the bandpass filter results for 200 kHz (i.e., 170–230 kHz) are presented in this paper. These results are shown in Figs. 8 and 9. The root-mean-square value of the residuals for the comparison in Fig. 8, using the time interval 500–800 μs, is 0.8 dB. For Fig. 9, the root-mean-square value of the residuals is 1.8 dB. The bandpass filtered results are visually clearer, with a slight improvement in the error statistics, but essentially the analysis can be performed successfully with either the unfiltered or filtered signal.

B. Experimental results for individual target positions

Fitting the results from a single target position is somewhat problematical because the transmitted signal contains a number of zero crossings, producing a received time series with occasional groups of outliers (typically 3 consecutive outliers per 100 time points, with a typical duration of 1 μs), each including the minimum value allowed by the digitizer

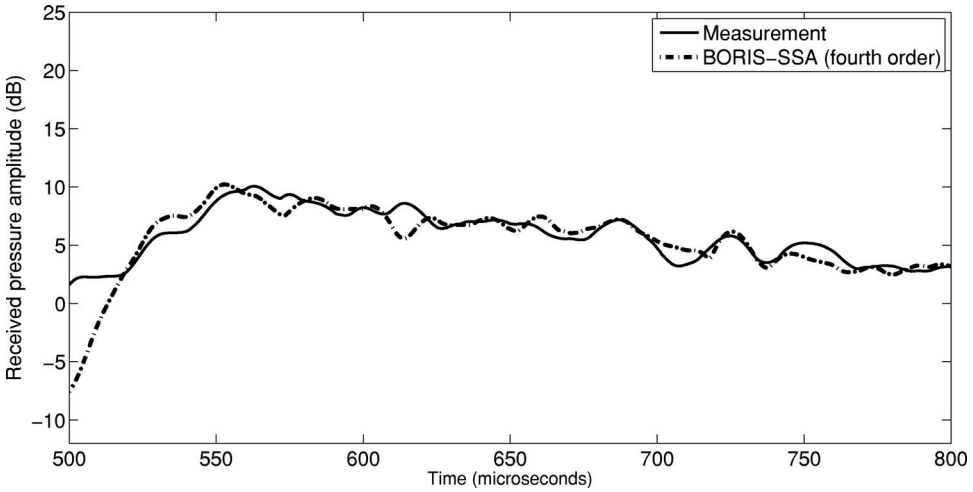


FIG. 8. Squared amplitude of the measured pressure (solid line) as a function of time for the bandpass filtered signal (170–230 kHz) with the BORIS-SSA fourth-order simulation. The time series for the nine target positions were averaged as described in the text. The first acoustic interactions with the PVC model occurred between 500 and 550 μs after transmit. In comparison with Fig. 6, there is a decrease in the overall level due to the amplitude response of the bandpass filter.

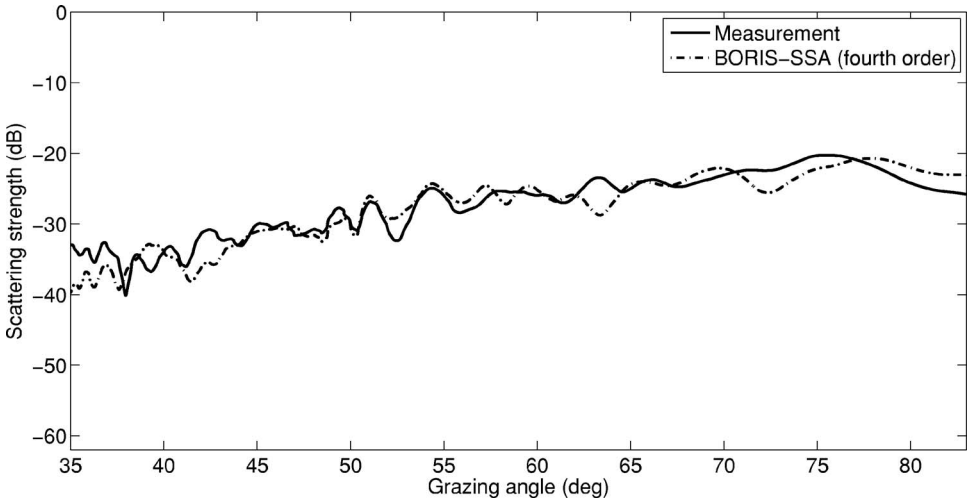


FIG. 9. Scattering strengths as a function of grazing angle derived from the measured and modeled time series for the bandpass filtered signal (170–230 kHz) and averaged over the nine target positions as described in the text. In comparison with Fig. 7, there is a decrease in the overall level due to the amplitude response of the bandpass filter.

as an extreme value. These extreme features are not modeled accurately by BORIS-SSA. These outliers can be removed by applying a five-point (1.666 μs) moving average to the data and model curves. The resulting root-mean-square errors of the residuals for the comparison of the measured and modeled time series for the nine individual target positions range from 2 to 3.5 dB. To achieve smaller standard error, a small ensemble of positions can be used, as shown in Figs. 5–9.

C. Sources of experimental error

The errors in the time series and scattering strength simulations, either from single target positions or as averages over an ensemble of positions, do not reflect a consistent underprediction or overprediction. Neither are there consistent errors in the timing of the returns, nor a significant departure from the overall trend of the data (given that the fourth-order rather than the second-order calculation is performed). While there are some uncertainties in the geoacoustic parameters (Table I), for this particular material varying the geoacoustic parameters will typically just shift the overall level or modify the trend of the scattering curve, instead of producing changes in its detailed structure. Therefore, the model inputs for the predictions presented here are simply taken directly from Table I. Positioning of the source and receiver was accurate to within 1 mm, and varying the position of the source and/or receiver in the BORIS-SSA predictions produced relatively small changes in the results. Data were also collected with the receiver positioned 2 mm closer and farther away. This positioning introduced the expected systematic changes in the timing of the returns but no introduction of random error. Measurements with the source and receiver at larger distances (71 and 61 cm, respectively) were used to obtain the received time series in the absence of the model and estimate the noise. Removing this noise component from the scattered field produced a negligible effect, except at grazing angles below 40°.

The BORIS-SSA model itself was used to estimate scattering contributions from the grooves in Fig. 2, since these toolpathing artifacts constitute a simple periodic rough surface for which BORIS-SSA has been shown to be an effective model. These contributions had a negligible effect on the predicted time series. The model was also used to estimate the effects of tilt in the PVC slab on the predictions. The tilt was estimated from underwater cameras as being on the order of tenths of a degree in the worst case.

Other sources of error are more consistent with the observed random errors. One possibility is that errors in the scattering from certain features could be due to errors in the profilometry (expected to be <1 mm in the x, y, and z directions, the latter representing the surface height) or to the presence of local regions of the physical model where the small-slope approximation fails. The profilometry was analyzed to locate any concentrations of especially large slopes in areas of the physical model. Such features were found to be scattered sparsely and rather uniformly over the surface.

Additionally, the possibility of returns from the back of the sample in the thinner regions of the physical model cannot be discounted. For a 200 kHz transmission, a nominal 6 dB loss would be encountered for a normal-incident wave traveling through the mean thickness of the model (6 cm) without considering any scattering effects. However, there are locations on the physical model where the slab thickness is as thin as 3 cm, and the returns from the back of the model could lead to intermittent errors in the predictions. The fact that an interface scattering code models the frequency results are modeled just as well as high frequency results indicates that returns from the back of the model are limited in their effect on the overall scattering.

Given these various analyses, it is probable that the errors in the predictions are not due to any particular problem with the experiment or model, with the exception of the intermittent errors arising from the thinner parts of the sample and the outliers described in Sec. IV B.

D. General comments

The use of a fabricated physical model in testing a rough surface scattering model provides significant advantages. The use of a homogeneous slab with no apparent inclusions or layering structure provides an opportunity to examine scattering from the rough interface in isolation. Additionally, the geoacoustic parameters and surface heights can be mea-

sured accurately and used as inputs to the scattering model. Therefore, the ability of the scattering model to perform an accurate prediction, without having to maximize data/model agreement by inverting for unknown parameters, can be assessed. In this case, the overall scattering behavior of this heterogeneous surface, with a root-mean-square roughness on the same order as the wavelength, could be modeled using a fourth-order small-slope calculation.

One interesting result was the absence of any features in the scattering curve associated with the critical angle. Both the data and the BORIS-SSA prediction exhibit an overall monotonic increase with no special behavior at near-critical angles. Given that frequencies, surface heights, speeds, and attenuations (as a function of wavelength) are scalable quantities, we expect that these high-frequency results here can be extrapolated to lower frequencies, and that the BORIS-SSA model would work similarly well at lower frequencies for an appropriately scaled rough surface. It is also of interest, as shown by analysis not presented in this paper, that the shear speed of the material did not have a significant influence on the predicted backscatter, although including it slightly increased the accuracy of the overall predictions. In terms of addressing the issues of manufacturing artifacts, frequency dependence, heterogeneity of the roughness profile, and uncertainty in source/receiver position, the BORIS-SSA model proved to be a versatile physics-based tool that allowed us to perform a detailed analysis of the problem.

Further experiments in this area will focus on expanding the range of grazing angles by introducing directional sensors. This should make shadowing and diffraction effects more important and thus provide a more serious challenge for small-slope models. Current efforts (including a tank experiment performed in July 2007) involve experiments of physical models that will exhibit homogeneous power-law roughness spectra and can be used for the validation of both stochastic and deterministic models. There are several issues that can be addressed by this effort, such as the behavior of the scattering strength curve in the vicinity of the critical angle. NRL has also performed propagation experiments, using PVC slabs to create simulated waveguides to compare measured transmission loss with predictions from an elastic PE model.[27] Given the success in using this approach to testing models for interface scattering, it is evident that investigations into the full range of scattering mechanisms encountered in the littoral ocean bottom—fractal surfaces, rippled surfaces, subsurface layers and volume inhomogeneities—can be addressed through a physical modeling approach.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research and the NATO Undersea Research Centre. The authors wish to acknowledge the contributions of Dr. Robert Wiley of the Allied Geophysical Laboratory/University of Houston for providing the PVC physical model, Dr. De-Hua Han of the University of Houston for carrying out the laboratory analysis of geoacoustic properties of multiple samples of PVC material, Dr. Allen Reed of the Naval Research Laboratory for performing an x-ray analysis of a material sample, and Dr. Eric Pouliquen for his assistance in developing the NRL/NURC collaboration.

[1]A. G. Voronovich, *Wave Scattering from Rough Surfaces*, 2nd ed. (Springer, Berlin, 1999).
[2]R. F. Gragg, D. Wurmser, and R. C. Gauss, "Small-slope scattering from rough elastic ocean floors: General theory and computational algorithm," J. Acoust. Soc. Am. **110**, 2878–2901 (2001).
[3]W. A. Kinney, C. S. Clay, and G. A. Sandness, "Scattering from a corrugated surface. Comparison between experiment, Helmholtz-Kirchhoff theory, and the facet-ensemble method," J. Acoust. Soc. Am. **73**, 183–194 (1983).
[4]R. S. Keiffer and J. C. Novarini, A wedge assemblage method for 3-D acoustic scattering from sea surfaces: Comparison with a Helmholtz-Kirchhoff method," in *Computational Acoustics*, edited by D. Lee, A. Cakmak, and R. Vichnevetsky (Elsevier, Amsterdam, 1990), pp. 67–81.
[5]R. J. Soukup and R. F. Gragg, "Backscatter from a limestone seafloor at 2–3.5 kHz: Measurements and modeling," J. Acoust. Soc. Am. **113**, 2501–2514 (2003).
[6]R. J. Soukup, "Bottom backscattering measured off the Carolina coast during the littoral warfare advanced development system concept Validation Experiment 97 (LWAD SCV 97)," NRL Formal Rep. No. 7140-98-9885, Naval Research Laboratory, Washington, DC, 1998.
[7]E. L. Kunz and R. C. Gauss, "Bottom backscattering strengths measured at 2–5 kHz in shallow water west of Scotland," *Proceedings of the MTS/IEEE Oceans 2005 Conference*, Washington, DC, 2005.
[8]E. Pouliquen, G. Canepa, L. Pautet, and A. P. Lyons, "Temporal variability of sea floor roughness and its impact on acoustic scattering," in *Proceedings of the Seventh European Conference on Underwater Acoustics, ECUA 2004*, edited by D. G. Simons, Delft, The Netherlands, Vol. 1, pp. 583–588.
[9]J. E. Summers, R. J. Soukup, and R. F. Gragg, "Mathematical modeling and computer-aided manufacturing of rough surfaces for experimental study of seafloor scattering," IEEE J. Ocean. Eng., in press (2007).
[10]C. W. Horton, Sr., S. K. Mitchell, and G. R. Barnard, "Model studies on the scattering of acoustic waves from a rough surface," J. Acoust. Soc. Am. **41**, 635–643 (1967).
[11]D. Chu and T. K. Stanton, "Application of Twersky's boss scattering theory to laboratory measurements of sound scattered by a rough surface," J. Acoust. Soc. Am. **87**, 1557–1568 (1990).
[12]H. E. F. Williams, T. C. Yang, and R. M. Kille, "Laboratory-scale measurement of scattering from modeled Arctic ice ridges," Ultrasonics **32**, 441–445 (1994).
[13]G. R. Mellema, "Subcritical acoustic scattering across a rough fluid-solid interface," Ph. D. dissertation, University of Washington, Seattle, 1999.
[14]J.-P. Sessarego, L. Guillon, and P. Sanchez, "Oil detection of acoustic means: Tank experiments," in *Proceedings of the Seventh European Conference on Underwater Acoustics, ECUA 2004*, edited by D. G. Simons, Delft, The Netherlands, Vol. 1, pp. 1077–1082 (2004).
[15]G. Canepa, L. Pautet, and E. Pouliquen, "Manual of BORIS-SSA: Bottom response from inhomogeneities and surface using small slope approximation," NATO Undersea Research Centre, Memorandum M-0152, 2005.
[16]G. Canepa, E. Pouliquen, L. Pautet, and N. G. Pace, "Bistatic scattering from the seabed at high frequency," in *Proceedings of the Seventh European Conference on Underwater Acoustics, ECUA 2004*, edited by D. G. Simons, Delft, The Netherlands, Vol. 1, pp. 595–600.
[17]R. J. Soukup, H. J. Simpson, E. C. Porse, J. E. Summers, and R. F. Gragg, "Geoacoustic Physical Modeling Rough Surface Scattering 2 (GPM RSS2) experiment: Measurement report, acoustic data, and profilometry," NRL Memorandum Rep. No. NRL/MR-MM/7140-04-8827, Naval Research Laboratory, Washington, DC, 2004.
[18]J. E. Summers, R. F. Gragg, and R. J. Soukup, "Topography measurement of scale-model representations of the rough ocean bottom by touch-trigger probe and its implications for spectral characterization," *Proceedings of the MTS/IEEE Oceans 2006 Conference*, Boston, MA, 2006.
[19]Our tests of various PVC samples have shown that there is no significant variation in the acoustic parameters relevant to this paper, *provided* the samples all come from the same manufacturer.
[20]The frequency band of the material analysis (300 kHz to 1.5 MHz) is higher than that of the acoustic tank experiment (100 to 300 kHz). The material analysis was restricted to a higher band because of the difficulty of finding an acoustic source that could produce the needed sound level

without ringing. The higher frequency results, showing a linear dependence of attenuation on frequency, allow for plausible extrapolation into the lower band.
[21]E. I. Thorsos and S. L. Broschat, "An investigation of the small slope approximation for scattering from rough surfaces. I. Theory," J. Acoust. Soc. Am. **97**, 2082–2093 (1995).
[22]K. B. Briggs, D. Tang, and K. L. Williams, "Characterization of interface roughness of rippled sand off Fort Walton Beach, Florida," IEEE J. Ocean. Eng. **27**, 505–514 (2002).
[23]D. Tang, "Fine scale measurements of sediment roughness and subbottom variability," IEEE J. Ocean. Eng. **29**, 929–939 (2004).
[24]While shear attenuation is not a direct input parameter to BORIS-SSA, its effect can be simulated by using a table of reflection coefficients, calculated using shear attenuation, as input. In our case shear attenuation did not change the reflection coefficients.
[25]Analysis performed by Dr. Allen Reed, Naval Research Laboratory, Stennis Space Center, MS, using a HD-500 microfocus x-ray computed tomography system.
[26]It is highly probable that the noise levels in this experiment result from the electronics of the data acquisition system, rather than acoustic phenomena in the tank.
[27]J. M. Collis, W. L. Siegmann, M. D. Collins, H. J. Simpson, and R. J. Soukup, "Comparison of simulations and data from a seismo-acoustic tank experiment," J. Acoust. Soc. Am. **122**, 1987–1993 (2007).

Spatial field shifts in ocean acoustic environmental sensitivity analysis

Stan E. Dosso and Michael G. Morley
School of Earth and Ocean Sciences, University of Victoria, Victoria British Columbia, V8W 3P6 Canada

Peter M. Giles and Gary H. Brooke
General Dynamics Canada, 3785 Richmond Road, Ottawa Ontario, K2H 5B7 Canada

Diana F. McCammon
McCammon Acoustical Consulting, 475 Baseline Rd., Waterville Nova Scotia, B0P 1V0 Canada

Sean Pecknold and Paul C. Hines
Defence Research and Development Canada–Atlantic, Dartmouth Nova Scotia, B2Y 3Z7 Canada

(Received 22 May 2007; revised 13 August 2007; accepted 14 August 2007)

This article examines the effects of spatial field shifts in ocean acoustic environmental sensitivity analysis. Acoustic sensitivity studies are typically based on comparing acoustic fields computed for a reference environmental model and for a perturbed model in which one or more parameters have been changed. The perturbation to the acoustic field due to the perturbed environment generally includes a component representing a spatial shift of the field (i.e., local field structure remains coherent, but shifts in range and/or depth) and a component representing a change to the shifted field. Standard sensitivity measures based on acoustic perturbations at a fixed point can indicate high sensitivity in cases where the field structure changes very little, but is simply shifted by a small spatial offset; this can conflict with an intuitive understanding of sensitivity. This article defines and compares fixed-point and field-shift corrected sensitivity measures. The approaches are illustrated with examples of deterministic sensitivity (i.e., sensitivity to a specific environmental change) and stochastic sensitivity (sensitivity to environmental uncertainty) in range-independent and range-dependent environments. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783123]

PACS number(s): 43.30.Pc [AIT] Pages: 2560–2570

I. INTRODUCTION

Sensitivity studies are often applied in ocean acoustics to examine the influence of various environmental parameters on propagation in a particular environment. Such studies are typically based on observing the change in acoustic fields that results from a given environmental change; for instance, by comparing fields computed for a reference environmental model and a perturbed model. Although this comparison is often carried out qualitatively, quantitative sensitivity measures can be defined.[1,2] The most sensitive parameters are those with the greatest effect on the fields; these parameters are often considered the most important, since errors or uncertainties in these parameters are potential sources of errors in model predictions.

It is well known in matched-field processing (MFP) that errors in environmental parameters (referred to as environmental mismatch) can lead to systematic errors in source localization.[3–5] This results from the fact that an error in some environmental parameters, particularly water depth, can lead to a systematic spatial shift in the acoustic field (i.e., the local field structure remains essentially coherent, but shifts in range and/or depth). This concept is also relevant to acoustic sensitivity studies, since changes in the acoustic field between reference and perturbed environments will generally include a component representing a spatial shift of the field and a component representing a change to the shifted field. Standard sensitivity measures based on the field perturbation at a fixed point in space can indicate high sensitivity in cases where the field structure changes very little but is shifted by a small spatial offset, which can conflict with an intuitive understanding of sensitivity. In practice, it is useful to examine both fixed-point and field-shift corrected sensitivity measures to understand acoustic sensitivity in a particular environment. This article develops and illustrates a practical approach to account for spatial field shifts in acoustic sensitivity analysis.

II. THEORY

A. Sensitivity measures

Ocean acoustic sensitivity studies are generally based on examining the differences in acoustic fields that result from a given change in the environment (i.e., comparing fields computed for a reference and a perturbed environmental model). This is referred to here as a deterministic sensitivity, as it applies to a specific environmental change. A quantitative deterministic sensitivity measure that is independent of the data and model scales is[1,2]

$$S = \frac{\big||p(m_j)| - |p(m_j + \delta m_j)|\big|}{|p|} = \frac{|\delta p|}{|p|}, \tag{1}$$

where m_j represents the jth model parameter, p is the complex acoustic pressure, and δm_j and δp are model and pressure perturbations. Deterministic sensitivities are specific to

the particular environmental perturbation, but are efficient to compute.

To generalize the previous concept to sensitivities to uncertainties in environmental parameters, estimated environmental uncertainties can be assumed to represent the standard deviation of a Gaussian probability distribution centered on the assumed parameter value. Determining the acoustic sensitivity then requires transferring these environmental uncertainties through the forward problem (propagation modeling), and quantifying the resulting variability in the acoustic field. Monte Carlo methods[6] provide a general approach to this problem, based on applying a large number of random Gaussian-distributed perturbations to the environmental parameters, and computing the corresponding perturbed fields. A stochastic sensitivity measure can then be defined to quantify the statistical variability of the ensemble of perturbed fields, such as the coefficient of variation (standard deviation divided by the mean),[2,7]

$$S = \frac{[\langle |\delta p|^2 \rangle]^{1/2}}{|p|}, \tag{2}$$

where $\langle \cdot \rangle$ represents an ensemble average.

The Monte Carlo approach characterizes the full nonlinearity of the sensitivity. However, for sufficiently small environmental perturbations or perturbations to relatively insensitive parameters, a linear approximation to the nonlinear stochastic sensitivity can be valid[2]

$$S = \frac{|\delta p(\delta m_j = \sigma_j)|}{|p|}, \tag{3}$$

where $\delta p(\delta m_j = \sigma_j)$ represents the acoustic perturbation resulting from a perturbation to model parameter m_j equal to its standard deviation σ_j. Hence, for a perturbation to an environmental parameter equivalent to its assumed uncertainty, the deterministic sensitivity represents a linear approximation to the nonlinear stochastic uncertainty.

B. Analytic spatial field shifting

The sensitivity measures defined in the previous section consider changes in the acoustic field at a fixed point in space. However, the acoustic-field perturbation due to an environmental perturbation generally includes a component representing a spatial shift of the field in addition to a change to the shifted field. In some applications, it may be desirable to examine these components of sensitivity separately. To motivate this work, it is instructive to first consider analytic results for a simple model of acoustic propagation due to a source at depth z_s in a Pekeris waveguide of depth D, sound speed c_1, and density ρ_1, over a seabed of sound speed c_2 and density ρ_2. Weston[8] and Buckingham[9] showed that propagation in such an environment is approximated by the normal mode solution for pressure-release boundaries at both the sea surface and seafloor employing an equivalent water depth $\tilde{D}$. In particular, the far-field acoustic pressure at angular frequency ω can be expressed as

$$p(r,z) = a \sum_{m=1}^{M} \sin(k_{zm} z_s) \sin(k_{zm} z) \frac{\exp(ik_{rm} r)}{\sqrt{k_{rm} r}}, \tag{4}$$

where the mth vertical and radial wave numbers are given by

$$k_{zm} = \frac{m\pi}{\tilde{D}}, \quad k_{rm} = \sqrt{\frac{\omega^2}{c_1^2} - k_{zm}^2}, \tag{5}$$

respectively, a is a complex constant, M is the number of propagating modes, and the equivalent water depth is given by

$$\tilde{D} = D + \frac{c_1 \rho_2}{\omega \rho_1 \sqrt{1 - (c_1/c_2)^2}}. \tag{6}$$

Now consider the reference pressure field defined for an environment with equivalent depth $\tilde{D}$ and the perturbed field defined for a perturbed environment $\tilde{D} + \delta\tilde{D}$ (the perturbation can involve changes to any of the environmental parameters defining $\tilde{D}$). To estimate the approximate depth shift in the acoustic field due to the environmental perturbation requires finding the value of δz such that the depth dependence of the acoustic field is unaltered by the perturbation, i.e., that satisfies

$$k_{zm}(\tilde{D}) z = k_{zm}(\tilde{D} + \delta\tilde{D})[z + \delta z]. \tag{7}$$

Substituting for k_{zm} leads to the relation

$$\delta z = \frac{\delta\tilde{D}}{\tilde{D}} z. \tag{8}$$

According to Eq. (8), the depth shift δz is of the same sign as the equivalent water-depth perturbation $\delta\tilde{D}$, indicating that an increase (decrease) in equivalent depth leads to a downward (upward) field shift. Equations (6) and (8) indicate that the field shift associated with a change in the water depth D is frequency independent, but the shifts associated with sound speeds and densities decrease with frequency. Further, the field shift is independent of r, increases linearly with $\delta\tilde{D}$ and z, and decreases with $\tilde{D}$.

The range evolution of the acoustic field is determined largely by model interference. Hence, to estimate the approximate range shift requires finding the value for δr such that the modal phase difference is unaltered by the environmental perturbation, i.e., that satisfies

$$[k_{rm}(\tilde{D}) - k_{rn}(\tilde{D})] r = [k_{rm}(\tilde{D} + \delta\tilde{D}) - k_{rn}(\tilde{D} + \delta\tilde{D})][r + \delta r]. \tag{9}$$

Making use of the approximate solution for modal dispersion[4,10]

$$k_{rm} - k_{rn} = \frac{c_1 (n^2 - m^2) \pi^2}{2\omega \tilde{D}^2}, \tag{10}$$

substituting for the radial wave number, and neglecting second-order terms leads to

$$\delta r = 2\frac{\delta\tilde{D}}{\tilde{D}}r. \tag{11}$$

According to Eq. (11), the range shift δr is of the same sign as the perturbation $\delta\tilde{D}$, indicating that an increase (decrease) in equivalent water depth leads to an outward (inward) field shift that increases linearly with $\delta\tilde{D}$ and r, is independent of z, and decreases with $\tilde{D}$. The fractional range shift of Eq. (11) is twice as large as the fractional depth shift of Eq. (8). Similar equations were derived for MFP by Shang and Wang[4] by considering the high resolution mode match estimator.

This analysis can be extended to weakly range-dependent environments using the adiabatic approximation and the concept of equivalent bathymetry. The adiabatic extension of Eq. (4) for an environment with range-dependent equivalent water depth $\tilde{D}(r)$ can be written

$$p(r,z) = a\sum_{m=1}^{M} \sin(k_{zm}[\tilde{D}(0)]z_s)\sin(k_{zm}[\tilde{D}(r)]z) \times \frac{\exp\left(i\int_0^r k_{rm}[\tilde{D}(u)]du\right)}{\sqrt{\int_0^r k_{rm}[\tilde{D}(u)]du}}, \tag{12}$$

where the dependence of the vertical and radial wave numbers on $\tilde{D}(r)$ is indicated explicitly and M represents the minimum number of modes between source and receiver. Zakarauskas *et al.*[11] showed that the adiabatic solution for range-dependent water depth is approximated by a range-independent solution for an appropriate constant water depth, referred to as the equivalent bathymetry. Substituting the equivalent water depth $\bar{D}$ defined by Eq. (6) for the water depth D in the results by Zakarauskas *et al.* leads to

$$p(r,z) = a\sum_{m=1}^{M} \sin(k_{zm}[\tilde{D}(0)]z_s)\sin(k_{zm}[\tilde{D}(r)]z)\frac{\exp(ik_{rm}[\bar{D}]r)}{\sqrt{k_{rm}[\bar{D}]r}}, \tag{13}$$

where the equivalent bathymetry is defined

$$\bar{D} = \left[\frac{1}{r}\int_0^r \frac{du}{\tilde{D}^2(u)}\right]^{-1/2}. \tag{14}$$

The similarity in form of Eqs. (4) and (13) indicate that Eq. (8) still applies for the depth shift, with $\delta\tilde{D}$ evaluated at the range point of interest. However, the range shift for adiabatic range dependence becomes

$$\delta r = 2\frac{\delta\bar{D}}{\bar{D}}r, \tag{15}$$

involving the integral of the effective depth from source to receiver, as per Eq. (14).

C. Numerical spatial field shifting

The analytic analysis of acoustic field shifting in the previous section is approximate, in that it neglects possible changes in the number of modes resulting from the environmental perturbation and second-order effects. These errors will generally increase with the size of the environmental perturbation and with frequency. Further, the analysis is based on a simplisitic model for acoustic propagation. Hence, the approach is not sufficiently accurate and general to encompass sensitivity studies involving realistic ocean environments and arbitrary range dependence. Therefore, a straightforward, general, and robust numerical approach to estimate and account for the spatial field shift component of sensitivity is developed and applied here, with the analytic analysis providing an approximate reference for comparison.

Consider a particular spatial point of interest on the reference acoustic field which is to be compared to the perturbed field. A range-depth window of preselected size is defined for the reference field, centered on this point. The two-dimensional correlation is then computed between the reference field over this window and the perturbed field over a series of windows of identical size which are centered at a grid of points about the point of interest (i.e., a series of windows shifted in space with respect to the point). The window center for the perturbed field that produces the highest correlation with the reference field is taken to define the range and depth shift resulting from the environmental perturbation. (Points near the edges of the range-depth grid can only be shifted to the edges, so may not be fully compensated.) The shift-corrected sensitivity can then be computed using any of Eqs. (1)–(3) by defining δp to be the difference between the reference acoustic field and the perturbed field at the optimally shifted point.

Comparing sensitivity values computed both with and without the field-shifting procedure (referred to here as fixed-point and field-shifted sensitivities, respectively) indicates the relative importance of the two components of sensitivity. By accounting for the spatial shift, the field-shifted sensitivity should generally be lower than the fixed-point sensitivity; however, the field-shifting procedure is based explicitly on matching local field structure, not on minimizing sensitivity.

III. SENSITIVITY ANALYSIS

A. Test case and implementation

This section provides examples of the sensitivity measures and field-shifting analysis outlined in Sec. II. The environmental parameters and uncertainties considered are based on the Malta Plateau, a well-studied shallow-water region of the Mediterranean Sea.[2,12] The environmental model, illustrated in Fig. 1, is comprised of a water column of depth $D=131$ m over a seabed consisting of three homogeneous layers. The ocean sound-speed profile (SSP), measured at the site, includes a strong negative gradient at shallow depths and a weak sound channel with its axis near midwater depth. The three seabed layers are characterized by sound speeds v_1, v_2, v_3, densities ρ_1, ρ_2, ρ_3, and attenuation coefficients $\alpha_1, \alpha_2, \alpha_3$. The upper two sediment layers have thicknesses

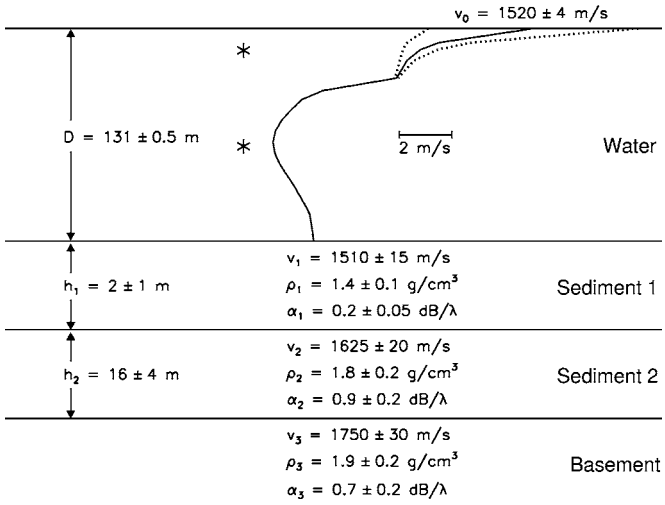


FIG. 1. Schematic diagram of the Malta Plateau environment, including assumed values and standard deviations for the water-column sound-speed profile and geoacoustic parameters. Asterisks indicate source depths considered.

h_1 and h_2; the basement is semiinfinite. Geoacoustic parameter values and uncertainties (standard deviations) that are representative of existing knowledge of the Malta Plateau region[2,12] are given in Fig. 1. The SSP uncertainty is taken to represent oceanographic variability due to surface heating/cooling and wind mixing, with the effects decaying exponentially with depth over the top 30 m, as shown in Fig. 1. This variability is represented in terms of the standard deviation of the surface sound speed v_0.

Numerical propagation modeling is carried out using PECAN,[13] a split-step Padé-expansion parabolic equation model. For consistency, all modeling was based on a computation grid with a range step of 5 m and depth step of 0.031 25 m, although the output grid used for subsequent field-shifting, sensitivity calculations, and plotting used an increased depth step of 1 m. Four Padé terms were included in the computations and a matched boundary condition was applied at the base of the computational grid, just below the deepest seabed interface.

As described in Sec. II C, the acoustic field shift for a particular point in space is based on determining the maximum correlation between the reference acoustic field in a range-depth window about the point and similar windows of the perturbed field centered at a grid of points about the point of interest. After examining a variety of possibilities, an appropriate window size was determined to extend ±50 m in range (10 grid points) and ±5 m in depth (5 grid points) about the point of interest. Correlations were calculated for perturbed-field windows shifted in increments of 5 m in range and 1 m in depth to maximum shifts of ±500 m in range and ±10 m in depth with respect to the reference-field window. (These window size and maximum shift values are used in all examples in this article.) In some cases it might be useful to employ larger field windows at lower frequencies and smaller windows at higher frequencies; however, we found no appreciable advantage to this for the cases considered here.

TABLE I. Mean sensitivities, averaged over range and depth, for cases considered in this article (see the text and figure captions for details).

Figure	Parameter perturbation	Frequency (Hz)	Deterministic (D) or stochastic (S)	Fixed-point sensitivity	Field-shifted sensitivity
3(a) and 3(b)	δD=0.5 m	100	D	0.18	0.02
3(e) and 3(f)	δD=1.0 m	100	D	0.32	0.04
3(i) and 3(j)	δD=2.0 m	100	D	0.52	0.07
4(a) and 4(b)	δD=0.5 m	1200	D	0.32	0.13
4(e) and 4(f)	δD=1.0 m	1200	D	0.45	0.23
4(i) and 4(j)	δD=2.0 m	1200	D	0.55	0.43
5(a) and 5(b)	Upslope	100	D	0.43	0.04
5(e) and 5(f)	Wedge	100	D	0.13	0.01
5(i) and 5(j)	Double wedge	100	D	0.13	0.02
6(a) and 6(b)	δv_1=15 m/s	100	D	0.14	0.06
6(e) and 6(f)	δv_1=30 m/s	100	D	0.26	0.12
6(i) and 6(j)	δv_1=60 m/s	100	D	0.48	0.23
7(a) and 7(b)	δv_1=15 m/s	1200	D	2.2	2.0
7(e) and 7(f)	δv_1=30 m/s	1200	D	4.5	4.2
7(i) and 7(j)	δv_1=60 m/s	1200	D	7.7	7.0
8(a) and 8(b)	δv_2=20 m/s	100	D	0.41	0.25
8(e) and 8(f)	δh_1=1.0 m	100	D	0.44	0.13
8(i) and 8(j)	δv_0=4.0 m/s	100	D	0.19	0.07
9(a) and 9(b)	v_1	100	S	0.15	0.08
9(e) and 9(f)	v_1, h_1	100	S	0.32	0.14
9(i) and 9(j)	All seabed	100	S	0.44	0.33
10(a) and 10(b)	1 pinch-out	100	S	0.20	0.08
10(e) and 10(f)	2 pinch-outs	100	S	0.18	0.09
10(i) and 10(j)	3 pinch-outs	100	S	0.14	0.09

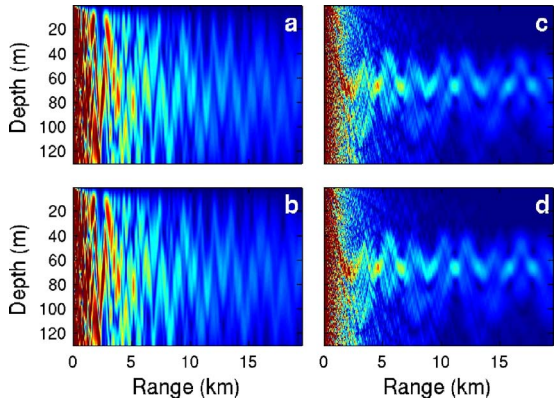


FIG. 2. (Color online) Acoustic fields for the reference environment (Fig. 1) and for a water depth perturbation of $\delta D = 0.5$ m are shown in (a) and (b), respectively, for a 100-Hz source at 65-m depth. Panels (c) and (d) show the same at 1200 Hz.

Finally, propagating acoustic fields are highly variable spatial functions, and hence the sensitivity at a point may not be representative of the overall sensitivity of the local field. Effective sensitivity stabilization is achieved here by spatially averaging results over ±100 m in range and ±2 m in depth.[1,2]

B. Deterministic sensitivity

1. Water depth and bathymetry

Deterministic sensitivities indicate the effect on acoustic propagation of specific environmental perturbations. The first example considers the acoustic sensitivity to water depth D for a 100-Hz source at 65-m depth in the Malta Plateau environment represented in Fig. 1. Figures 2(a) and 2(b) show the acoustic field (magnitude) computed for the reference environment and for a perturbed environment with $\delta D = 0.5$ m (i.e., a water-depth increase of 1 standard deviation). Portrayed in this manner, it is virtually impossible to perceive any difference between the fields. Figure 3(a) shows the fixed-point sensitivity for this depth perturbation, computed according to Eq. (1) as the difference between the reference and perturbed fields divided by the reference field at each point. According to this plot, the acoustic field is highly sensitive to this small change in water depth, with many localized regions of $S > 0.5$, particularly in the near-surface and near-bottom waters away from the sound-channel axis. The mean sensitivity averaged over the range-depth surface is 0.18 (Table I), indicating a substantial change in the acoustic field. However, the major component of this fixed-point sensitivity is due to small spatial field shifts. Accounting for these shifts by applying the numerical field-shifting procedure for every point on the acoustic field surface before computing the sensitivity, as described in Sec. II C, leads to the field-shifted sensitivity results shown in Fig. 3(b). Few field-shifted sensitivity values are higher than 0.1 and the mean sensitivity is 0.02. This represents a tenfold decrease over the fixed-point sensitivity and could represent the difference between classifying the acoustic fields as sensitive or insensitive to this water-depth perturbation.

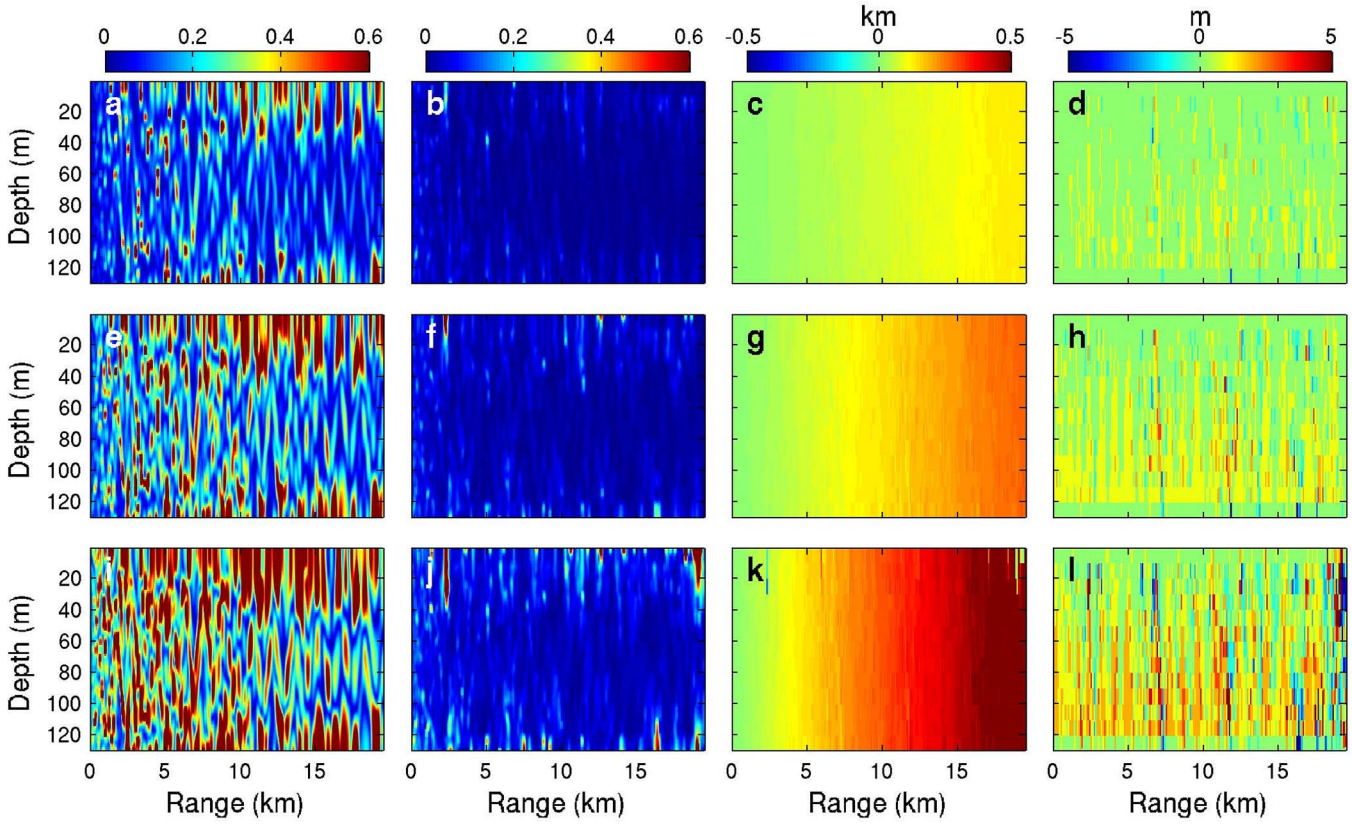


FIG. 3. (Color online) Deterministic sensitivity results for a 100-Hz source at 65-m depth for perturbations to water-depth D of 0.5, 1, and 2 m are shown in rows 1–3, respectively. Columns 1–4 show fixed-point sensitivities, field-shifted sensitivities, range shifts, and depth shifts, respectively.

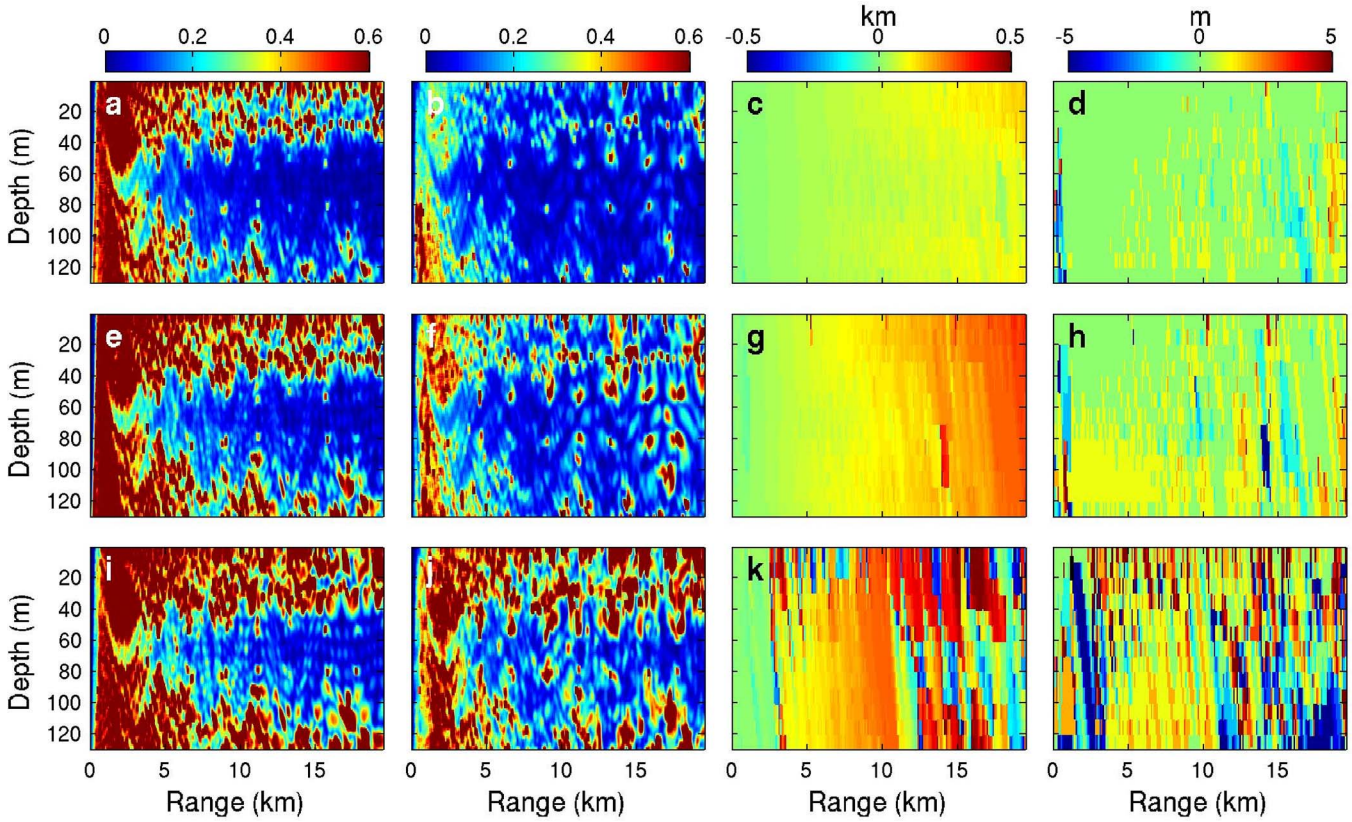


FIG. 4. (Color online) Deterministic sensitivity results for a 1200-Hz source at 65-m depth for perturbations to water-depth D of 0.5, 1, and 2 m are shown in rows 1–3, respectively. Columns 1–4 show fixed-point sensitivities, field-shifted sensitivities, range shifts, and depth shifts, respectively.

The spatial field shifts determined in calculating the field-shifted sensitivity of Fig. 3(b) are shown in Figs. 3(c) and 3(d). Figure 3(c) shows that positive range shifts result from the 0.5-m increase in water depth, and that the range shifts are essentially independent of depth and increase uniformly with range, in agreement with the simple model of Eq. (11). Further, the computed shift at 20-km range of 0.15 km agrees well with the prediction of Eq. (11). The depth shifts shown in Fig. 3(d) fluctuate between 0 and 1 m. This is consistent with the simple model of Eq. (8), which predicts a maximum depth shift of 0.5 m (at the seafloor), whereas the grid employed for the correlation search had a depth increment of 1 m. The depth shifts in Fig. 3(d) generally increase with depth, as predicted by Eq. (8), although this increase is not uniform and range-dependent variations are evident. Note that a zero depth shift is associated with the deepest 10 m of the field (one correlation window) since the fields here cannot be shifted downward into the seabed.

Figures 3(e)–3(h) and 3(i)–3(l) are similar to Figs. 3(a)–3(f) but for water-depth perturbations of $\delta D=1$ and 2 m, respectively. As δD increases, the fixed-point sensitivity increases significantly; however, the increase in the field-shifted sensitivity is considerably smaller. Mean fixed-point sensitivity values are 0.32 and 0.52 for $\delta D=1$ and 2 m, respectively, while the corresponding mean field-shifted sensitivities are 0.04 and 0.07 (Table I). The range shifts increase uniformly with δD, in agreement with Eq. (11). The depth shifts also generally increase with δD, although the dependence is somewhat more complicated than that predicted by Eq. (8).

To consider the dependence of water-depth sensitivity on frequency, Fig. 4 shows fixed-point and field-shifted sensitivities for the same scenario as Fig. 3, but at a higher frequency of 1200 Hz [acoustic fields at 1200 Hz are shown in Figs. 2(c) and 2(d)]. Compared to the corresponding results for 100 Hz in Fig. 3, the fixed-point sensitivities in Fig. 4 are generally higher at short ranges and near the water-column boundaries, but lower at midwater depth as sound-channel propagation with limited bottom interaction is better developed. The field-shifted sensitivities shown in Fig. 4 are lower in each case than the corresponding fixed-point sensitivities; however, the differences are not as great as those in Fig. 3. Further, the difference between the fixed-point and field-shifted sensitivities appears to decrease with increasing depth-perturbation size. These trends are borne out by the mean sensitivity values given in Table I. The range and depth shifts at 1200 Hz in Fig. 4 are generally similar to those at 100 Hz in Fig. 3, although the field-shifting pattern appears to break down with increasing depth perturbation size at the higher frequency. The previous observations are consistent with the fact that the errors in the assumptions of the analytic field-shift analysis in Sec. II B increase with increasing frequency and perturbation size.

The acoustic sensitivity to range-dependent water depth (bathymetry) is considered in Fig. 5 for a 100-Hz source at 65-m depth. The reference environment consists of the range-independent environment of Fig. 1 and three perturbed environments are considered including: a uniform upslope of 4 m over the first 5-km range followed by a flat bottom; a

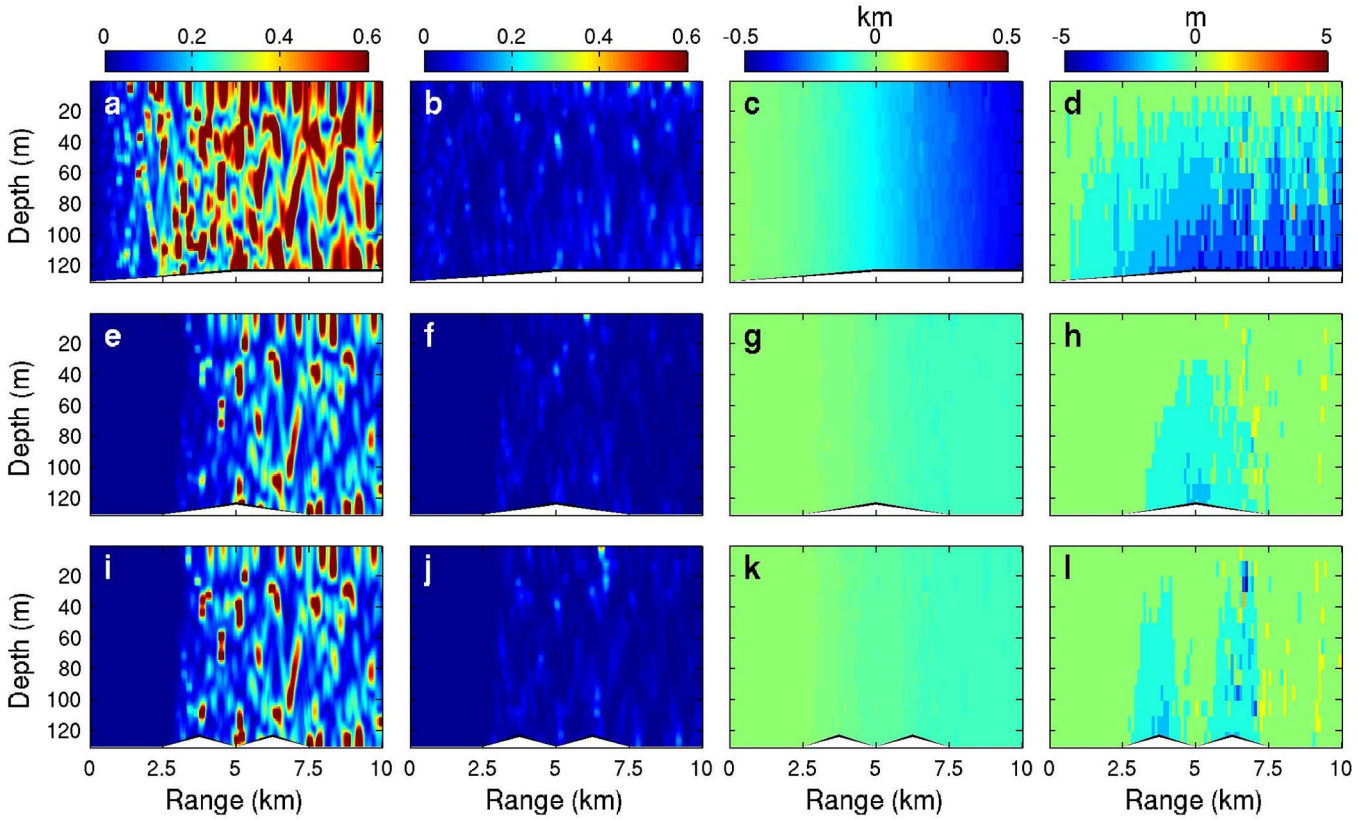


FIG. 5. (Color online) Deterministic sensitivity results for a 100-Hz source at 65-m depth for range-dependent water-depth perturbations consisting of a uniform upslope, a single wedge, and a double wedge are shown in rows 1–3, respectively. Bathymetry of the perturbed environment is indicated in white at the bottom of each panel, scaled by a factor of 2 (see the text for detailed description). Columns 1–4 show fixed-point sensitivities, field-shifted sensitivities, range shifts, and depth shifts, respectively.

wedge consisting of an upslope of 2 m over 2.5–5-km range followed by a downslope of 2 m over 5–7.5-km range; and a double wedge consisting of two wedges of maximum elevation 2 m extending from 2.5–5-km to 5–7.5-km range. Figure 5 and Table I show that the fixed-point sensitivity is significantly higher for the upslope case than for the wedge cases, with similar sensitivities for the one- and two-wedge cases. In each case the field-shifted sensitivity is much lower than the fixed-point sensitivity (Table I indicates a reduction in mean sensitivity of about a factor of ten), indicating that the main component of the acoustic-field perturbation is a spatial shift. Figure 5(c) shows a negative range shift for the upslope case which continues to increase in magnitude with range beyond the end of the slope at 5 km, consistent with the predictions of Eqs. (14) and (15). Figure 5(d) shows a negative depth shift that increases in magnitude over the upslope section, then remains roughly constant (with range-dependent fluctuations). For the one- and two-wedge cases, negative range shifts occur over the initial upslope portions which are not canceled out over the downslope portions, consistent with Eq. (15). The negative depth shifts for the wedge cases mimic the bathymetry, consistent with Eq. (8), although some anomalous shifts are obtained at greater ranges.

2. Geoacoustic parameters

The acoustic sensitivity to seabed geoacoustic parameters is also expected to have a field-shift component, as outlined in Sec. II B. The acoustic sensitivity to perturbations in the upper-sediment sound speed v_1 at 100 Hz is considered in Fig. 6 for a range-independent case. The fixed-point sensitivities increase with perturbation size, with mean values of 0.14, 0.26, and 0.48 for perturbations of $\delta v_1 = 15$, 30, and 60 m/s, respectively (Table I). The field-shifted sensitivities are significantly smaller, with corresponding mean values of 0.06, 0.12, and 0.23. However, the difference between the fixed-point and field-shifted sensitivities is not as large as was the case for water-depth perturbations at this frequency (Figs. 3 and 5 and Table I), indicating that field shifting accounts for a smaller proportion of the acoustic-field perturbation for sediment sound-speed perturbations. Figure 6 indicates negative range shifts, in agreement with Eq. (11) where $\tilde{D}$ decreases with an increase in bottom speed c_2 in Eq. (6). However, the depth shifts are more complicated than predicted in Eq. (8), with positive and negative values and rapid range-dependent fluctuations.

Figure 7 shows sensitivities for the same scenario as Fig. 6, except at a higher frequency of 1200 Hz. The acoustic fields are much more sensitive to the sound-speed perturbation to the 2-m sediment layer at the higher frequency, with values as high as $S \approx 10$. In Fig. 7, the field-shifted sensitivities are not significantly smaller than the fixed-point sensitivity (see also Table I), and the range and depth shifts are poorly determined. These results are consistent with Eq. (6),

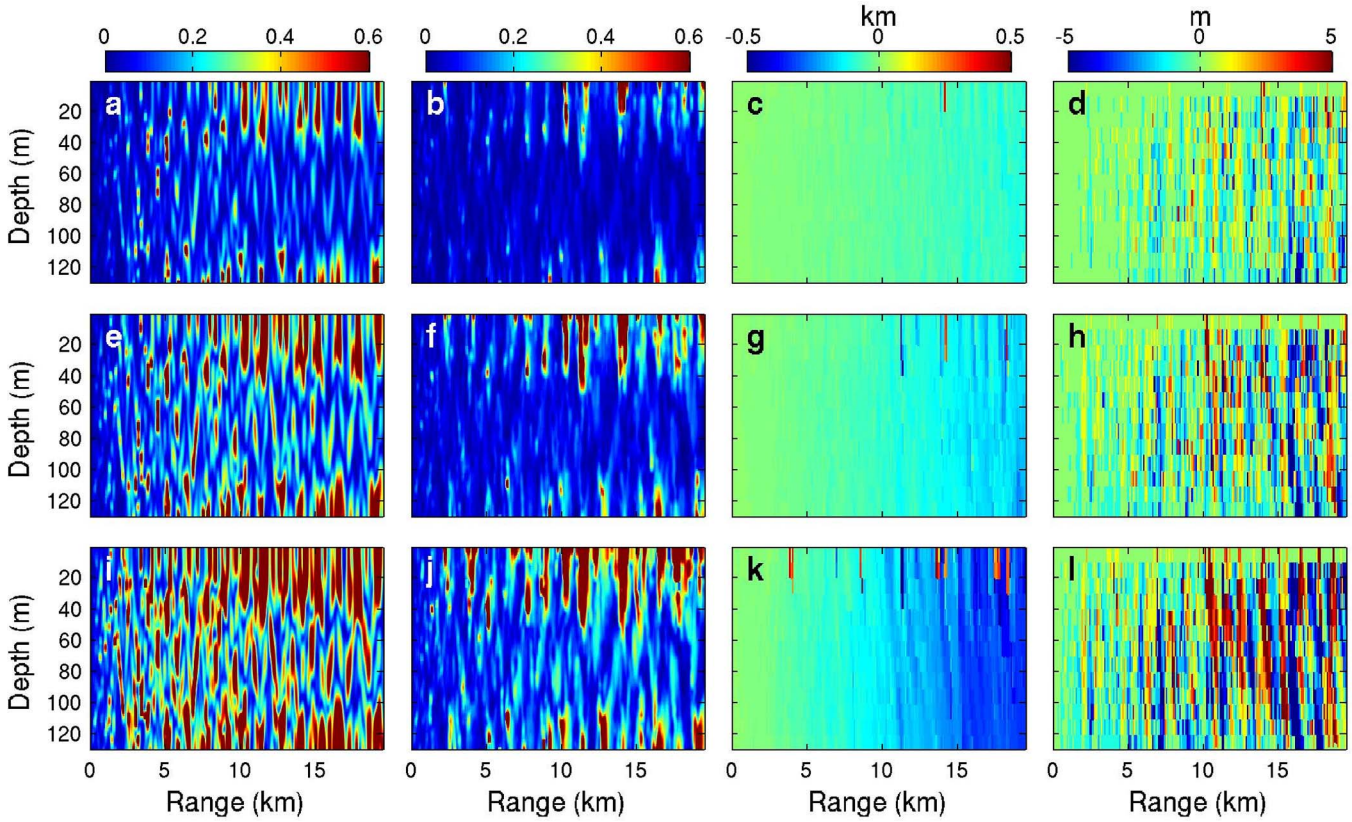


FIG. 6. (Color online) Deterministic sensitivity results for a 100-Hz source at 65-m depth for perturbations to sediment speed v_1 of 15, 30, and 60 m/s are shown in rows 1–3, respectively. Columns 1–4 show corresponding fixed-point sensitivities, field-shifted sensitivities, range shifts, and depth shifts, respectively.

which indicates that the spatial shift component of sensitivity to geoacoustic parameters decreases with increasing frequency.

Deterministic sensitivities for three other environmental parameters at 100 Hz are considered in Fig. 8. Figures 8(a)–8(d) show that the sensitivity results for a perturbation to the second sediment layer sound speed v_2 of 20 m/s (1 standard deviation) are qualitatively similar to the v_1 sensitivity in Fig. 6, with a negative range shift, complicated depth shift, and reasonable reduction between the fixed-point and field-shifted sensitivities. Figures 8(e)–8(h) show that the sensitivity results for a perturbation to the first sediment layer thickness h_1 of 1 m are qualitatively similar to the water-depth sensitivity in Fig. 3, with a positive range shift and substantial reduction between the fixed-point and field-shifted sensitivities, although the depth shifts are somewhat more complicated. Finally, Figs. 8(i)–8(l) show the sensitivity results for a perturbation to the upper SSP parameter v_0 of 4 m/s. In this case the source depth was taken to be 15 m to coincide with the SSP perturbation (Fig. 1); in all other cases the source depth was 65 m. The fixed-point and field-shifted sensitivities indicate that a significant component of the sensitivity is due to spatial field shifts, with the positive SSP perturbation leading to negative range shifts. Mean sensitivities for all cases in Fig. 8 are given in Table I. The sensitivities to the other environmental parameters in Fig. 1 were considered in a similar manner (not shown), with results indicating substantially lower sensitivities than those illustrated here.

C. Stochastic sensitivity

Deterministic sensitivity analysis, as illustrated in Sec. III B, is straightforward to carry out and interpret for specific environmental perturbations. However, stochastic sensitivity analysis based on Monte Carlo sampling is required to investigate acoustic sensitivity to uncertainties in environmental parameters. Computing field-shifted stochastic sensitivity requires that field shifts are computed and taken into account for every range-depth point of each perturbed acoustic field in the ensemble of Eq. (2).

To quantify the acoustic sensitivity to uncertainty in sediment sound speed v_1, Figs. 9(a) and 9(b) show fixed-point and field-shifted stochastic sensitivities computed from 100 Gaussian-distributed random perturbations to v_1 with standard deviation 15 m/s representing the uncertainty in this parameter (Fig. 1). Repeating the Monte Carlo sampling (with different random perturbation sequences) yielded essentially identical results, indicating that the process converged to stable sensitivity estimates. The stochastic sensitivities in Figs. 9(a) and 9(b) are similar to the deterministic sensitivities in Figs. 6(a) and 6(b) for a single perturbation to v_1 of 15 m/s (1 standard deviation), although the structure in Fig. 9 is somewhat smoother and broader. The mean sensitivities are also similar (Table I). This general agreement indicates that the linear approximation of Eq. (3) applies reasonably well in this case.

Stochastic analysis can be applied to compute the acoustic sensitivity to uncertainties in multiple environmental pa-

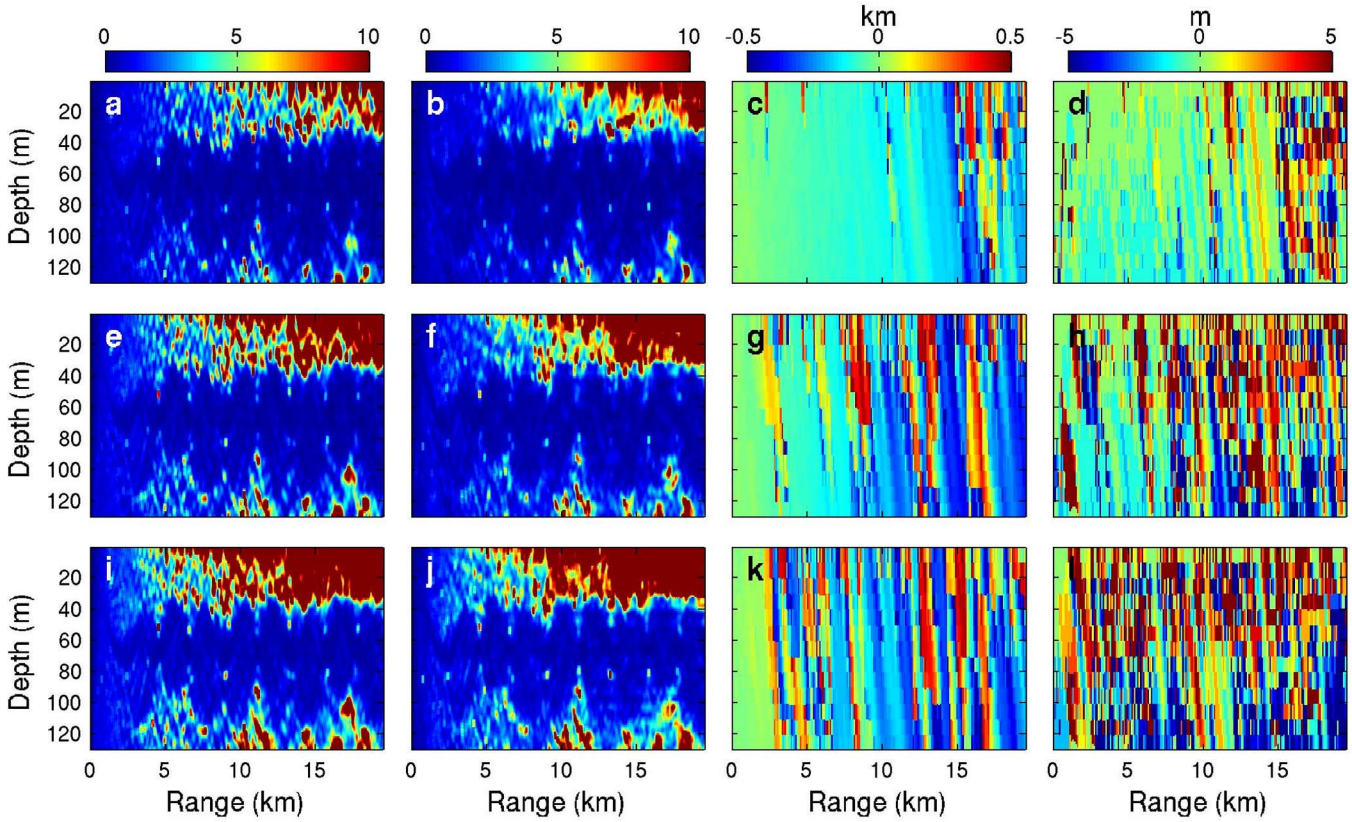


FIG. 7. (Color online) Deterministic sensitivity results for a 1200-Hz source at 65-m depth for perturbations to sediment speed v_1 of 15, 30, and 60 m/s are shown in rows 1–3, respectively. Columns 1–4 show corresponding fixed-point sensitivities, field-shifted sensitivities, range shifts, and depth shifts, respectively. Note that the sensitivity scales differ from other figures in this article.

rameters, which is not straightforward with a linearized deterministic approach. This is illustrated in Figs. 9(c) and 9(d), which show fixed-point and field-shifted stochastic sensitivities computed from 200 perturbed environments, each involving a Gaussian-distributed random perturbation to both sediment speed v_1 and sediment thickness h_1 of standard deviations 15 m/s and 1 m, respectively. Figures 9(c) and 9(d) show that simultaneous uncertainty in both v_1 and h_1 leads to a significant increase in sensitivity over the single-parameter case in Figs. 9(a) and 9(b). Table I indicates that the mean sensitivity for the two-parameter case is approximately twice that of the single-parameter case, and that the field-shifted sensitivity is approximately half the fixed-point sensitivity in each case. To extend the analysis to all geoacoustic parameters, Figs. 9(e) and 9(f) show fixed-point and field-shifted stochastic sensitivities computed from 300 perturbed environments, with each environment involving Gaussian-distributed random perturbations to all parameters with standard deviations given in Fig. 1. Figures 9(c) and 9(e) indicate that the fixed-point sensitivity increases only a moderate amount in going from uncertainties in v_1 and h_1 to all seabed parameters (mean sensitivity increases from 0.32 to 0.44). However, Figs. 9(d) and 9(f) indicate a larger increase in field-shifted sensitivities (mean sensitivity increases from 0.14 to 0.33). This suggests that uncertainties in multiple, diverse parameters cannot be accounted for to the same extent by simple spatial field shifts.

A final example of the stochastic analysis considers the acoustic sensitivity to range-dependent environmental uncertainties consisting of random "pinch-outs" of the upper sediment layer in Fig. 1 (i.e., regions where the upper layer is absent). Figures 10(a) and 10(b) show fixed-point and field-shifted sensitivities computed from 100 perturbed environments in which the water depth is constant but the upper sediment layer is absent over an interval of random length from 2.4 to 9.6 km which is positioned randomly with starting point between 1 and 19 km range. Similarly, Figs. 10(c) and 10(d) show sensitivities for two randomly positioned pinch-outs, each of random length from 1.2 to 4.8 km, with the constraint that the pinch-outs do not overlap. Finally, Figs. 10(e) and 10(f) show sensitivities for three nonoverlapping pinch-outs, each of random length from 0.8 to 3.2 km. (Note that, on average, the total length of the pinch-out segment is the same for the three cases.) Figure 10 shows comparable fixed-point sensitivities for the three cases (mean values 0.14–0.20 in Table I), suggesting that the number of pinch-outs does not strongly influence these results. The field-shifted sensitivities (means 0.08–0.09) indicate that spatial field shifts represent a significant component of the sensitivity.

IV. SUMMARY AND DISCUSSION

This paper considered the role of spatial field shifts in ocean acoustic environmental sensitivity studies, including both deterministic sensitivity (i.e., sensitivity to a specific environmental change) and stochastic sensitivity (sensitivity to environmental uncertainty). Simple analytic expressions

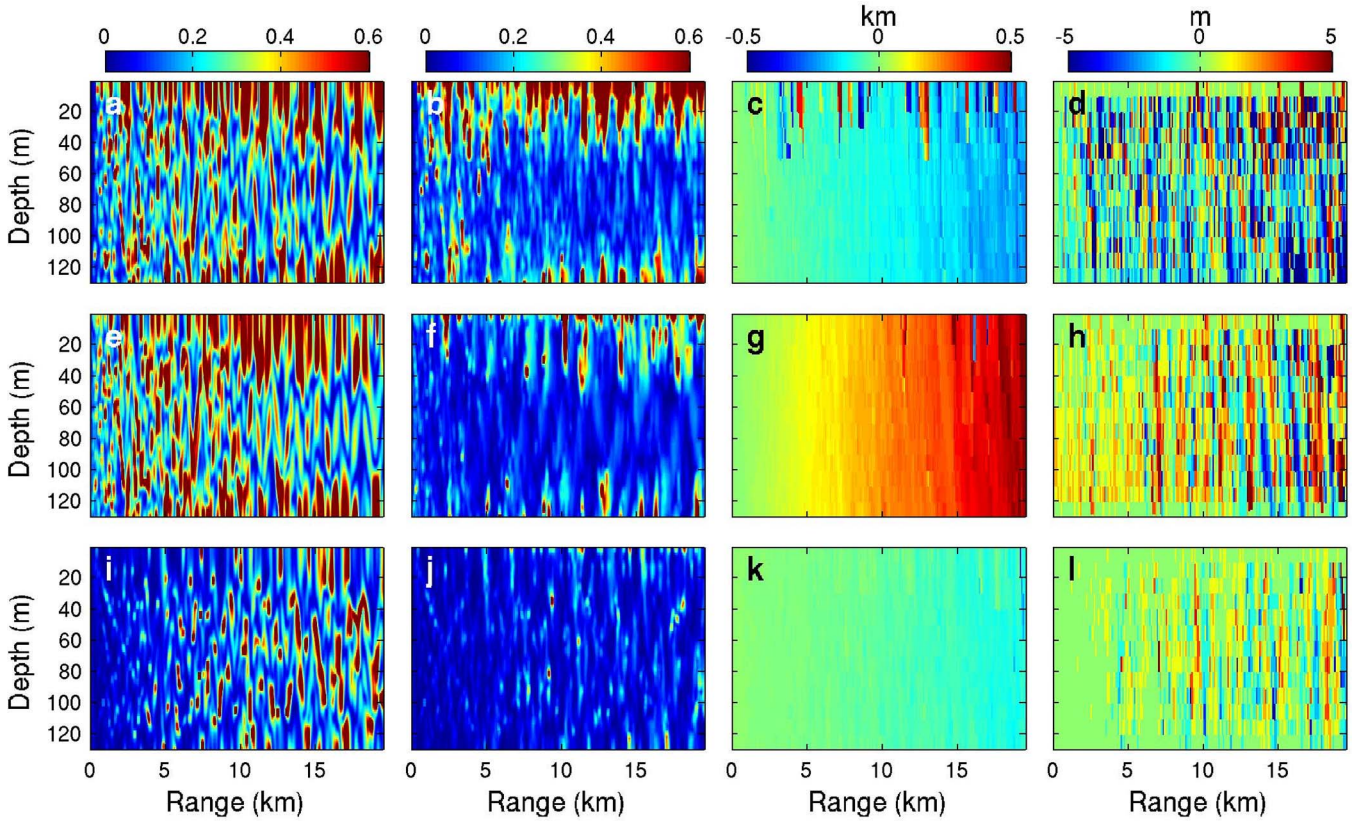


FIG. 8. (Color online) Deterministic sensitivity results for a 100-Hz source for one standard deviation perturbations to geoacoustic parameters v_2, h_1, and v_0 are shown in rows 1–3, respectively. Columns 1–4 show corresponding fixed-point sensitivities, field-shifted sensitivities, range shifts, and depth shifts, respectively. The source depth is 65 m for rows 1 and 2, and 15 m for row 3.

were derived for spatial field shifts based an equivalent-depth representation of a Pekeris waveguide and an equivalent-bathymetry representation of adiabatic range dependence. While the approximations and assumptions inherent in the analytic results render them inadequate for sensitivity studies in realistic environments, they provide a theoretical framework for understanding field shifting and yield predictions that compare reasonably well (qualitatively) with a more general numerical approach. The numeri-

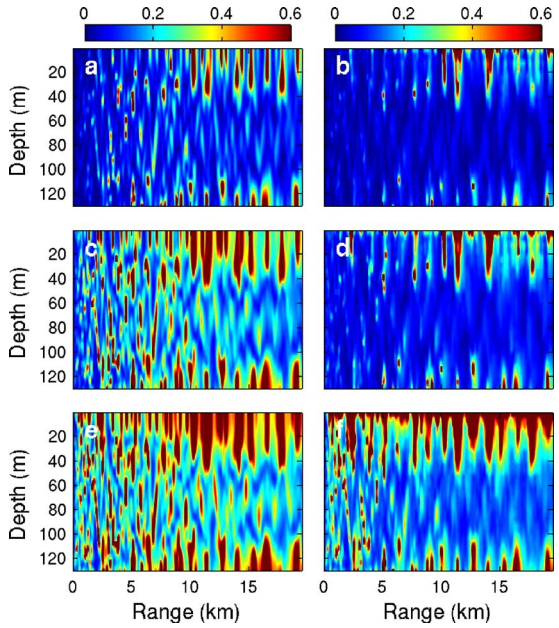


FIG. 9. (Color online) Stochastic sensitivity results for a 100-Hz source at 65-m depth from Monte Carlo perturbations to v_1, v_1, and h_1, and all seabed geoacoustic parameters are shown in rows 1–3, respectively. Columns 1 and 2 show fixed-point and field-shifted sensitivities.

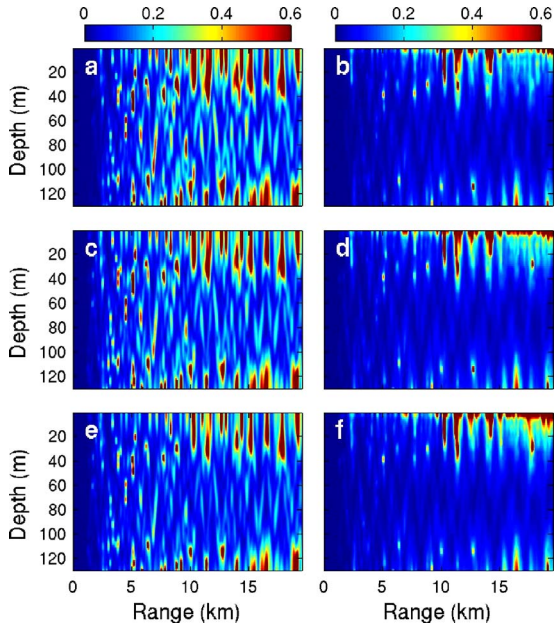


FIG. 10. (Color online) Stochastic sensitivity results for a 100-Hz source at 65-m depth from Monte Carlo perturbations representing one, two, or three random pinch-outs of the upper sediment layer (see the text) are shown in rows 1–3, respectively. Columns 1 and 2 show fixed-point and field-shifted sensitivities, respectively.

cal approach was based on applying a grid search for the shift that maximizes the correlation between range-depth windows of the reference and perturbed acoustic fields.

A number of examples were presented of deterministic sensitivity analysis for an environment based on the Malta Plateau. It was found that spatial field shifts represent the dominant component of sensitivity to water depth (range independent and range dependent) at low frequency (100 Hz), and was a substantial component at high frequency (1200 Hz). The field-shift component of sensitivity to sediment sound speed decreased dramatically with frequency (as predicted analytically). Sensitivity to sediment layer thickness was similar to water-depth sensitivity. Sensitivity to the SSP included a significant field-shift component. Other environmental parameters were found to be relatively insensitive for the cases considered.

Stochastic sensitivities to environmental uncertainties were considered for single and multiple geoacoustic parameters. Field shifting was found to be of reduced importance when uncertainties on all geoacoustic parameters were considered simultaneously. Finally, stochastic sensitivity to range-dependent environmental uncertainty was illustrated for the case of pinch-outs of the upper sediment layer. The sensitivity was found to be similar for one, two, or three pinch-outs (extending over the same average range), with field-shifting representing a significant component.

The significance of field shifting in sensitivity analysis is, of course, problem dependent, and the results could differ significantly for different environments. The theoretical treatment presented here is limited to Pekeris-like waveguides, and the numerical examples, while representative of realistic shallow-water environments, are by no means exhaustive. However, the numerical field-shifting approach is generally applicable, and can be used to investigate this issue in any environment of interest.

[1]R. T. Kessel, "A mode-based measure of field sensitivity to geoacoustic parameters in weakly range-dependent environments," J. Acoust. Soc. Am. **105**, 122–129 (1999).

[2]S. E. Dosso, P. M. Giles, G. H. Brooke, D. McCammon, S. P. Pecknold, and P. C. Hines, "Linear and nonlinear measures of ocean acoustic environmental sensitivity," J. Acoust. Soc. Am. **121**, 42–45 (2007).

[3]D. R. Del Balzo, C. Feuillade, and M. R. Rowe, "Effects of water-depth mismatch on matched-field localization in shallow water," J. Acoust. Soc. Am. **83**, 2180–2185 (1988).

[4]E. C. Shang and Y. Y. Wang, "Environmental mismatching effects on source localization processing in mode space," J. Acoust. Soc. Am. **89**, 2285–2290 (1991).

[5]G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Both, and P. W. Schey, "Mirages in shallow-water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).

[6]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986).

[7]R. E. Walpole, *Introduction to Statistics* (Macmillan, New York, 1982).

[8]D. E. Weston, "A Moiré fringe analog of sound propagation in shallow water," J. Acoust. Soc. Am. **31**, 647–654 (1959).

[9]M. J. Buckingham, "Array gain of a broadside vertical line array in shallow water," J. Acoust. Soc. Am. **65**, 148–161 (1979).

[10]T. C. Wang and E. C. Shang, *Underwater Acoustics* (Scientific Press, Beijing, 1981), Chap. 4, p. 196.

[11]P. Zakarauskas, S. E. Dosso, and J. A. Fawcett, "Matched-field inversion for source localization and optimal equivalent bathymetry," J. Acoust. Soc. Am. **100**, 1493–1500 (1996).

[12]S. Pecknold, P. C. Hines, D. D. Ellis, S. E. Dosso, D. F. McCammon, P. M. Giles, and G. H. Brooke, "Effect of environmental variability on measured transmission loss during the Base '04 experiment," Proceedings of the 8th European Conference on Underwater Acoustics, edited by S. M. Jesus and O. C. Rodriguez (University of Algarve, Carvoeiro, Portugal, 2006), pp. 433–438.

[13]G. H. Brooke, D. J. Thomson, and G. R. Ebbeson, "PECan: A Canadian parabolic equation model for underwater sound propagation," J. Comput. Acoust. **9**, 69–100 (2001).

Statistical estimation of transmission loss from geoacoustic inversion using a towed array

Yong Han Goh,[a] Peter Gerstoft, and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238, USA

Chen-Fen Huang
Department of Marine Environmental Informatics, National Taiwan Ocean University, Keelung, Taiwan

(Received 4 May 2007; revised 9 August 2007; accepted 9 August 2007)

Geoacoustic inversion estimates environmental parameters from measured acoustic fields (e.g., received on a towed array). The inversion results have some uncertainty due to noise in the data and modeling errors. Based on the posterior probability density of environmental parameters obtained from inversion, a statistical estimation of transmission loss (TL) can be performed and a credibility level envelope or uncertainty band for the TL generated. This uncertainty band accounts for the inherent variability of the environment not usually contained in sonar performance prediction model inputs. The approach follows [Gerstoft *et al.* IEEE J. Ocean. Eng. **31**, 299–307 (2006)] and is demonstrated with data obtained from the MAPEX2000 experiment conducted by the NATO Undersea Research Center using a towed array and a moored source in the Mediterranean Sea in November 2000. Based on the geoacoustic inversion results, the TL and its variability are estimated and compared with the measured TL. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2782915]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 2571–2579

I. INTRODUCTION

Geoacoustic inversion using matched-field processing is a model-based technique that has been applied successfully to derive environmental and seabed parameters for propagation prediction.[1–4] Computer simulations are used to model the acoustic response to different sea-bed types (forward models), and efficient search algorithms used to find the environment giving an optimal match between the modeled and measured data. It should be noted, however, that inverse problems are usually under-determined, and solutions may not be unique. For example, if the results of the inversion are only required for sonar performance prediction, it is only the resulting acoustic field in the water that matters, often at long range and within a restricted range of frequencies. In this context, a precise description of the seabed is not necessary, and it usually is sufficient to describe a simpler "effective" seabed model having a similar acoustic effect on the underwater sound field within the range-frequency domain of interest.[5]

By far, a sound source in combination with a receiving vertical line array (VLA) is the most common configuration for collection of acoustic data for geoacoustic inversion. It is a sensible choice as the propagating acoustic field is received at almost all angles if the VLA spans a large portion of the water column. The use of horizontal line arrays (HLAs) has been gaining in popularity as it offers several advantages over a VLA. This includes the ease of deployment from a ship, and the ability to cover large areas of interest as the ship travels, together with either a separate towed source,[2,6,7] or using the ship self-noise for inversion.[8,9] Others have used bottom-moored horizontal line arrays and a towed source.[3,10] In this paper, a towed array with a fixed source was used for geoacoustic inversion, which to the authors' knowledge, is a setup that has not been presented in current literature on geoacoustic inversion.

There are uncertainties associated with the seabed parameters obtained from geoacoustic inversion, and in this paper, the mapping of these uncertainties to the transmission loss domain is also presented. This has been performed using results from a VLA with a moving source.[11,12] In this paper, we present results obtained using a HLA with a stationary source. Sonar performance prediction using a probability density function based on environmental variability has also been discussed in Refs. 13–15.

II. PARAMETER ESTIMATION USING GENETIC ALGORITHMS AND INVERSION PROCEDURE

A. Base line environment model

The base line environmental model established for the North Elba site from Ref. 16 was used. Figure 1 illustrates the base line environmental model, comprising the water column, a constant thickness sediment layer with depth dependent sound speed increasing with depth, and a bottom half-space layer.

B. Inversion procedure and cost function

The inversion is performed as follows:
(1) Record the acoustic field at the site of interest.

[a]Author to whom correspondence should be addressed. Now with DSO National Laboratories, 20 Science Park Drive, Singapore 118230. Electronic mail: gyonghan@dso.org.sg.

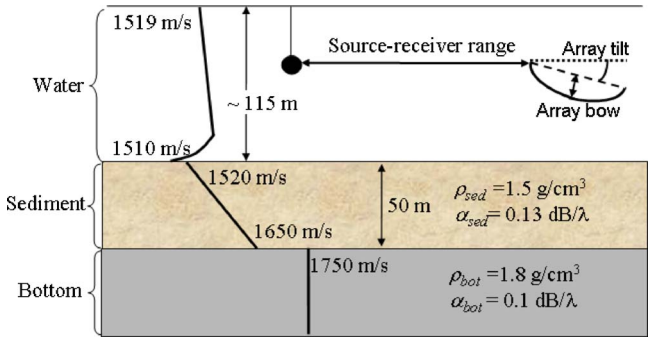


FIG. 1. (Color online) The base line environmental model for the North Elba experiment site (see Ref. 16).

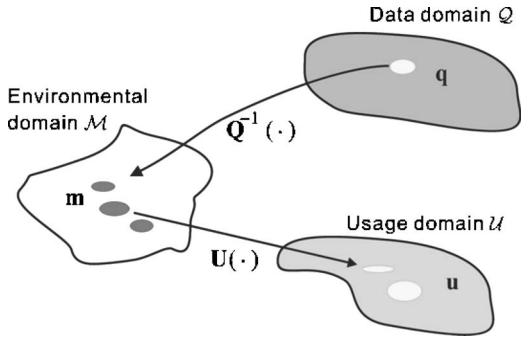


FIG. 2. An observation $\mathbf{q}$ ($\in \mathcal{Q}$) is mapped into a distribution of environmental parameters $\mathbf{m}$ ($\in \mathcal{M}$) that potentially could have generated it. These environmental parameters are then mapped into the usage domain $\mathcal{U}$.

(2) Choose a suitable propagation model. In this paper, the NURC SNAP normal-mode propagation model,[17] which is robust, fast, and suitable for low frequencies, was used.

(3) Choose a suitable cost function to minimize. The phone-coherent method traditionally has been used. However, recent analysis[2,7,18] has shown that frequency-coherent matched-field inversion works well for multi-frequency HLA data when the source spectrum is known, where matched-field correlations between data and replica frequency vectors are performed coherently over frequency but incoherently over range (hydrophones). The frequency-coherent cost function is derived using the maximum likelihood method in Sec. III A and is given by

$$C_F = \left[\prod_{i=1}^{N_H} (1 - \phi_{F_i}) \right]^{-N_H}, \tag{1}$$

where

$$\phi_{F_i} = \frac{|\sum_{j=1}^{N_F} [s(\omega_j) a_{ij}]^* q_{ij}|^2}{\sum_{j=1}^{N_F} |s(\omega_j) a_{ij}|^2 \sum_{j=1}^{N_F} |q_{ij}|^2}. \tag{2}$$

In Eqs. (1) and (2), N_F is the number of frequency components, N_H is the number of hydrophones, a_{ij} corresponds to the modeled Green's function (* denotes the complex conjugate), q_{ij} is the measured complex pressure, and $s(\omega_j)$ is the complex source term at the jth frequency. In Eq. (2), the correlation is between the measured and modeled acoustic pressure vectors. The cost function takes on a value of 0 for two identical signals and 1 for completely uncorrelated signals.

Equation (2) may be expressed equivalently as

$$\phi_{F_i} = \frac{\left| \sum_{j=1}^{N_F} a_{ij}^* \frac{q_{ij}}{s(\omega_j)} \right|^2}{\sum_{j=1}^{N_F} |a_{ij}|^2 \sum_{j=1}^{N_F} \left| \frac{q_{ij}}{s(\omega_j)} \right|^2}. \tag{3}$$

In Eq. (3), the correlation is between the measured and modeled Green's function, where the measured Green's function is obtained by dividing the measured acoustic data q_{ij} by the known source spectrum over the frequency band. Equation (2) is preferred as Eq. (3) may suffer from numerical instabilities when $s(\omega_j)$ is close to zero. Note that the source term needs to be explicitly included in the frequency-coherent cost function in both Eqs. (2) and (3).

The frequency-coherent cost function was used in this paper, as we found the phone-coherent method yields worse results, as in Refs. 2, 7, and 18.

(4) An efficient algorithm is needed to navigate the enormous search space and find the global minimum of the cost function. In this paper, a genetic algorithm search is used with the propagation model SNAP as implemented in the inversion package, SAGA,[19] 40,000 forward models were used in the inversion searches. Reference 20 provides a detailed description of GA and their application to geoacoustic parameter estimation.

III. MAPPING GEOACOUSTIC PARAMETER UNCERTAINTIES TO TRANSMISSION LOSS DOMAIN

The mapping of geoacoustic parameter uncertainties to the transmission loss (TL) domain has been described in Refs. 11 and 12 and is summarized in this section for completeness.

Figure 2 summarizes the estimation of TL (usage domain $\mathcal{U}$) from ocean acoustic data observed on a vertical or horizontal array (data domain $\mathcal{Q}$).[11] Based on the ocean acoustic data, we statistically characterize TL, the usage domain $\mathcal{U}$. The vector $\mathbf{q}_i$ represents the acoustic data observed at the ith hydrophone and the vector $\mathbf{u}$ represents TL at several ranges and depths. This is mapped via a set of environmental parameters $\mathbf{m}$ in the environmental domain $\mathcal{M}$. Both the experimental data $\mathbf{q}_i$ and the usage domain quantity $\mathbf{u}$ are related to $\mathbf{m}$ via forward models $\mathbf{Q}(\mathbf{m})$ and $\mathbf{U}(\mathbf{m})$. The geoacoustic inverse problem is solved as an intermediate step to obtain the posterior distribution of environmental parameters $p(\mathbf{m}|\mathbf{q}_i)$. We are not just interested in the environment itself, but also a statistical estimation of the TL field. Based on the posterior distribution $p(\mathbf{m}|\mathbf{q}_i)$, the probability distribution of the TL$p(\mathbf{u}|\mathbf{q}_i)$ is obtained via Monte Carlo integration. From this TL probability distribution, all relevant statistics of TL such as the median, percentiles and correlation coefficients can be obtained. The vector $\mathbf{u}$ is used to denote the transmission loss as a K-dimensional vector at discrete (r_k, z_k) positions, where $u_k = u(r_k, z_k)$.

A. Frequency-coherent likelihood and cost functions

The likelihood function for the frequency-coherent model [Eq. (4)] is derived following a similar approach as in Refs. 21 and 22 which focused on the phone-coherent model.

In the case of the coherent Bartlett processor, the source strength is known from frequency to frequency in a relative sense, but the absolute amplitude and phase are unknown. This is often the case, for example, when a Linear Frequency Modulation (LFM) signal is transmitted, where the absolute amplitude and phase are unknown. At a single phone, the relation between the observed complex-valued data vector $\mathbf{q}_i=[q_i(\omega_1)\ldots q_i(\omega_{N_F})]^T$ and the modeled data may be described by the model[23]

$$\mathbf{q}_i = \alpha \mathbf{S}\mathbf{a}_i(\mathbf{m}) + \mathbf{e}. \tag{4}$$

Here $\mathbf{S}$ is the diagonal matrix with its diagonal being the source vector at zero phase,

$$\mathbf{S} = \begin{bmatrix} s(\omega_1) & 0 & \ldots & 0 \\ 0 & s(\omega_2) & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & s(\omega_{N_F}) \end{bmatrix},$$

and $\alpha=Ae^{i\theta}$ represents the unknown amplitude scaling and phase shift which are independent of frequency. The Green's function $\mathbf{a}_i(\mathbf{m})$ at the corresponding N_F frequencies is obtained using an acoustic propagation model and an environmental model $\mathbf{m}$. The error term is represented by $\mathbf{e}$.

In the Bayesian inference framework, the solution to the inverse problem is given by

$$p(\mathbf{m}|\mathbf{q}_i) = \frac{p(\mathbf{q}_i|\mathbf{m})p(\mathbf{m})}{p(\mathbf{q}_i)} \propto \mathcal{L}(\mathbf{m})p(\mathbf{m}), \tag{5}$$

where $\mathcal{L}(\mathbf{m})$ is used to denote the likelihood function $p(\mathbf{q}_i|\mathbf{m})$.

Following the derivation in Ref. 11 where the error vector $\mathbf{e}$ is assumed to be Gaussian distributed with zero mean and covariance $\mathbf{C}_e=\nu\mathbf{I}$, the likelihood function is given by

$$\mathcal{L}(\mathbf{m},\nu,\mathbf{S},\alpha) = \frac{1}{\pi^{N_F}\nu^{N_F}} \exp\left(-\frac{\|\mathbf{q}_i - \alpha\mathbf{S}\mathbf{a}_i(\mathbf{m})\|^2}{\nu}\right). \tag{6}$$

Here, ν is assumed to be constant over the frequency band.

The maximum-likelihood estimate of α is obtained by solving $\frac{\partial \log \mathcal{L}}{\partial \alpha}=0$, giving

$$\alpha_{\mathrm{ML}} = \frac{[\mathbf{S}\mathbf{a}_i(\mathbf{m})]^\dagger \mathbf{q}_i}{\|\mathbf{S}\mathbf{a}_i(\mathbf{m})\|^2}. \tag{7}$$

Substituting α_{ML} back into Eq. (6), the likelihood function is then

$$\mathcal{L}(\mathbf{m},\nu) = \frac{1}{\pi^N \nu^N} \exp\left(-\frac{\phi_{F_i}(\mathbf{m})}{\nu}\right), \tag{8}$$

where

$$\phi_{F_i}(\mathbf{m}) = \|\mathbf{q}_i\|^2 - \frac{|[\mathbf{S}\mathbf{a}_i(\mathbf{m})]^\dagger \mathbf{q}_i|^2}{\|[\mathbf{S}\mathbf{a}_i(\mathbf{m})]\|^2} \tag{9}$$

is the cost function for the ith hydrophone. For a single receiver, the cost function in Eq. (1) can be obtained by normalizing Eq. (9) with $\|\mathbf{q}_i\|^2$. For N_H hydrophones, the likelihood function may be written as

$$\mathcal{L}(\mathbf{m},\boldsymbol{\nu}) = \prod_{i=1}^{N_H} (\pi\nu_i)^{-N_F} \exp\left(-\frac{\phi_{F_i}(\mathbf{m})}{\nu_i}\right), \tag{10}$$

where the noise variance ν_i is assumed to be different at different hydrophones.

For an HLA, the noise variance at the first hydrophone could be different from the last hydrophone due to distance from the towing vessel and different flow noise. The error variance $\boldsymbol{\nu}$ may be treated as a nuisance parameter and eliminated via integrating Eq. (10) weighted by a noninformative prior of $\boldsymbol{\nu}[p(\nu_i)\propto 1/\nu_i]$ over its entire range[24]

$$\mathcal{L}(\mathbf{m}) = \int_0^\infty \cdots \int_0^\infty \mathcal{L}(\mathbf{m},\boldsymbol{\nu})p(\nu_1)\ldots p(\nu_{N_H})d\nu_1\ldots d\nu_{N_H}. \tag{11}$$

Integrating out $\boldsymbol{\nu}$, the likelihood function for the frequency-coherent cost function case is then

$$\mathcal{L}(\mathbf{m}) = \frac{1}{\pi^{N_F}} \prod_{i=1}^{N_H} \frac{(N_F-1)!}{\phi_{F_i}(\mathbf{m})^{N_F}} \tag{12}$$

or,

$$\mathcal{L}(\mathbf{m}) \propto \left[\frac{1}{\phi^g_{F_i}(\mathbf{m})}\right]^{N_F N_H} = \left[\prod_{i=1}^{N_H} \phi_{F_i}(\mathbf{m})\right]^{-N_F}, \tag{13}$$

where $\bar{\phi}^g_{F_i}(\mathbf{m}) = \sqrt[N_H]{\prod_{i=1}^{N_H} \phi_{F_i}(\mathbf{m})}$ is the geometric mean of the cost function over N_H hydrophones. The cost function defined in Eq. (2) is thus derived from the likelihood function $\mathcal{L}(\mathbf{m})$.

The above derivation assumes that the errors are independent across both spatial samples of the acoustic field and frequencies. In practice, these errors may be strongly correlated, for example, when the dominant source of errors is due to frequency-dependent modeling mismatch, the modeling errors may not be independent across the frequencies used. In this case, the full data error covariance matrix $\mathbf{C}_e$, which would be nondiagonal in general, should be used.[25]

B. Prediction in the TL domain

Probability density functions that describe yet unobserved events are referred to as predictive distributions. The posterior predictive distribution of $\mathbf{u}$ for a set of discrete ranges and depths given the observed acoustic data $\mathbf{q}_i$ may be obtained by integrating the values of the TL with respect to the posterior distribution of the model parameters[11]

$$p(\mathbf{u}|\mathbf{q}_i) = \int_{\mathcal{M}} \delta(\mathbf{U}(\mathbf{m}) - \mathbf{u}) p(\mathbf{m}|\mathbf{q}_i) d\mathbf{m}. \quad (14)$$

It is possible to implement Eq. (14) directly using Markov chain Monte Carlo (MCMC) methods described in the next subsection. The posterior distribution $p(\mathbf{u}|\mathbf{q}_i)$ carries all the information about TL in the presence of the geoacoustic inversion uncertainties. As the predictive distributions are not necessarily Gaussian, it is preferable to characterize the distributions with medians and distance between the 5th and 95th percentiles instead of means and standard deviations. The βth percentile of the TL distribution at a given position, denoted by u^{β}, is computed by finding the TL value that satisfies

$$\int_{-\infty}^{u^{\beta}} p(u|\mathbf{q}_i) du = \beta/100. \quad (15)$$

The present formulation in the usage domain, Eq. (14), has the advantage that it is easy to incorporate additional *independent* information about the environment $\mathbf{m}$, as carried out in Eqs. (16)–(18). If the additional parameters $\mathbf{m}_a$ with probability density $p(\mathbf{m}_a)$ are independent of $\mathbf{m}$, then

$$p(\mathbf{u}|\mathbf{q}_i) = \int_{\mathcal{M}_a} \delta(\mathbf{U}(\mathbf{m},\mathbf{m}_a) - \mathbf{u}) p(\mathbf{m}|\mathbf{q}_i) p(\mathbf{m}_a) d\mathbf{m} d\mathbf{m}_a, \quad (16)$$

where $\mathcal{M}_a$ is the environmental domain spanned by $\mathbf{m}$ and $\mathbf{m}_a$. Equation (16) is used in the computation of the influence of sound speed variability on TL prediction in Sec. VI.

In addition, suppose the posterior distribution $p(\mathbf{m}_i|\mathbf{q}_i)$ of a parameter m_i from $\mathbf{m}$ is not the correct distribution to be used in mapping to the usage domain (e.g., assumption of constant source depth throughout the trial). Under the assumption that this parameter is independent of the other parameters and the observed data vector $\mathbf{q}_i$, $p(m_i|\mathbf{q}_i)$ can be replaced with an alternative distribution $p(m_i)$. Then

$$p(\mathbf{u}|\mathbf{q}_i) = \int_{\mathcal{M}} \delta(\mathbf{U}(m_i,\mathbf{m}_{-i}) - \mathbf{u}) p(\mathbf{m}_{-i}|\mathbf{q}_i) p(m_i) dm_i d\mathbf{m}_{-i}, \quad (17)$$

where $\mathbf{m}_{-i}$ is the parameter vector $\mathbf{m}$ with its ith component removed. This is used for including uncertainty in the source depth using statistics from the matched-field estimated source depths in the TL prediction in Sec. VI.

Finally, not all parameters in $\mathbf{m}$ are required for input in the forward mapping $\mathbf{u}=\mathbf{U}(\mathbf{m})$. Only a subset $\mathbf{m}'$ is required, e.g., receiver geometric parameters such as receiver range, receiver depth, array bow, and array tilt are not used in the forward mapping. Then $\mathbf{u}=\mathbf{U}(\mathbf{m})=\mathbf{U}(\mathbf{m}')$ and

$$p(\mathbf{u}|\mathbf{q}_i) = \int_{\mathcal{M}'} \delta(\mathbf{U}(\mathbf{m}') - \mathbf{u}) p(\mathbf{m}'|\mathbf{q}_i) d\mathbf{m}', \quad (18)$$

where $\mathcal{M}'$ is the environmental domain spanned by $\mathbf{m}'$ and $p(\mathbf{m}'|\mathbf{q}_i)$ can be obtained from $p(\mathbf{m}|\mathbf{q}_i)$ by integrating out the parameters which are not required.

C. Markov chain Monte Carlo methods

Markov chain Monte Carlo (MCMC) is essentially Monte Carlo integration using Markov chains. MCMC methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its stationary distribution. In the Bayesian framework, there is often a need to integrate over high-dimensional probability distributions to make inference about model parameters or to make predictions. MCMC methods are able to evaluate integrals in high-dimensional space efficiently[26,27] and have been found to be well suited for problems of Bayesian inference. They are extensively used in various fields of inverse problems, including ocean acoustics.[9,28] The commonly used MCMC methods are the Metropolis-Hastings algorithm,[29,30] and Gibbs sampling.[31]

The integral in Eq. (14) is the expectation of function $\delta(\mathbf{U}(\mathbf{m})-\mathbf{u})$ with respect to the posterior distribution of the model parameters. It can be approximated by using the MCMC samples $\{\mathbf{m}^{(t)}\}$ drawn from the posterior distribution of model parameters $p(\mathbf{m}|\mathbf{q}_i)$,

$$p(\mathbf{u}|\mathbf{q}_i) = \frac{1}{T}\sum_{t=1}^{T} \delta(\mathbf{U}(\mathbf{m}^{(t)}) - \mathbf{u}), \quad (19)$$

where the superscript t is used to label the sequence of states in a Markov chain and T denotes the total length of the sequence. Equation (19) is implemented using a numerical approximation by binning the calculated TL values. The bin width should be selected small enough to have negligible effect on the distribution. In this paper, a 1 dB bin width is used.

Using all samples from MCMC runs would consume a large amount of computation time to compute $p(\mathbf{u}|\mathbf{q}_i)$ and memory storage to save all the samples $\{\mathbf{m}^{(t)}\}$.[32] Statistical literature has suggested that inferences could be based on a subsampling of each sequence, with a subsampling factor high enough that successive draws of $\mathbf{m}$ are approximately independent.[27,32,33] The subsampling reduces the number of samples needed to calculate $p(\mathbf{m}|\mathbf{q}_i)$, translating into a large saving in computer time for calculating $p(\mathbf{u}|\mathbf{q}_i)$. The subsampled model parameter vectors then are used to compute $p(\mathbf{u}|\mathbf{q}_i)$.[11]

All results presented in this paper are generated by SAGA,[19] which implements the MCMC method using the Metropolis-Hastings algorithm described in Ref. 34.

IV. EXPERIMENTAL SETUP

The experiment was conducted as part of the MAPEX2000 experiment by the NATO Undersea Research Center (NURC) on 28 November 2000 in a shallow water area north of Elba island, off the Italian west coast (see Fig. 3).

This area is characterized by a flat bottom covered with clay and sand-clay sediments. The bathymetry was measured to be between 110 and 120 m along the track, (42.928° N, 10.145° E) to (42.928° N, 10.260° E). The HLA was towed by NRV Alliance at approximately 4 knots with the first hydrophone approximately 350 m behind the ship's stern. The

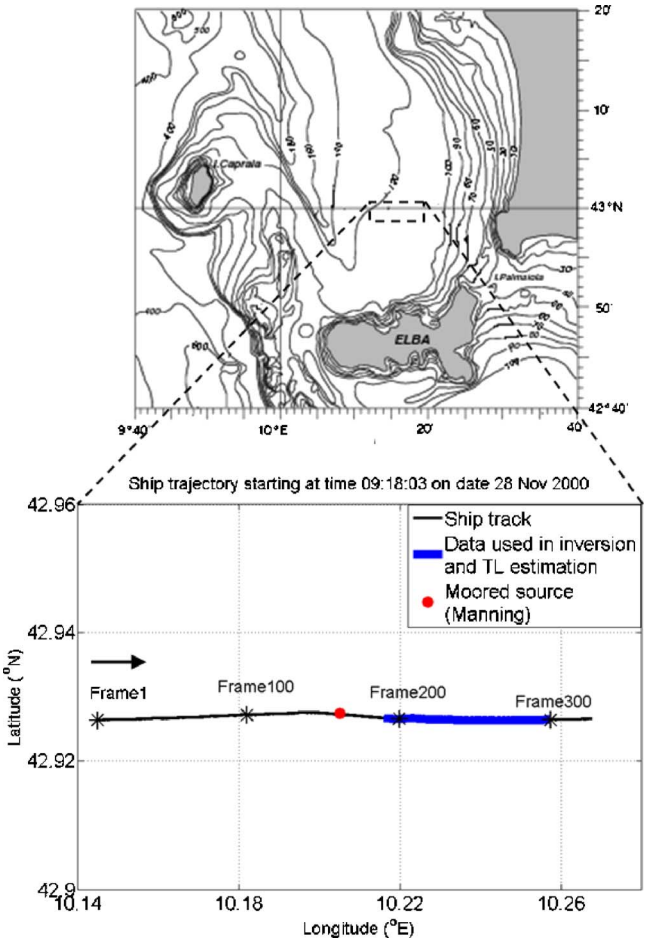


FIG. 3. (Color online) Bathymetry of experimental location, and track of NRV Alliance during the MAPEX2000 experiment. All times are Coordinated Universal Time (UTC). Each frame represents a 15 s increment.

acoustic source deployed from NRV Manning which was moored at (42.926° N, 10.206° E). The tow depth of the HLA was 55–65 m during the trial. The HLA is 254 m in length, and data recorded along the entire length was used (128 hydrophones spaced at 2 m). A sequence of 2 s LFM sweeps from 150 to 500 Hz was transmitted every 15 s. The received time series was converted to the frequency domain using a fast Fourier transform with a frequency bin width of 0.09 Hz. Frequency bins corresponding to 300–500 Hz in 10 Hz increments were used in the inversion for comparison with modeled results.

The sound speed profile was measured before the experiment and shown in Fig. 4. The profile exhibits a slight positive gradient for most of the water column, except near the bottom, where there is a sharp decrease in sound speed.

V. HLA SEABED CHARACTERIZATION: INVERSION RESULTS

During the experiment, a 2 s LFM signal (150–500 Hz) was transmitted from the fixed source at 15 s intervals. This was recorded by the HLA towed by NRV Alliance.

Before inversion, the received acoustic data were deconvolved by dividing by the ideal 2 s LFM source spectrum over the 150–500 Hz frequency band.[7] The frequency bins used for inversion correspond to the frequencies Doppler

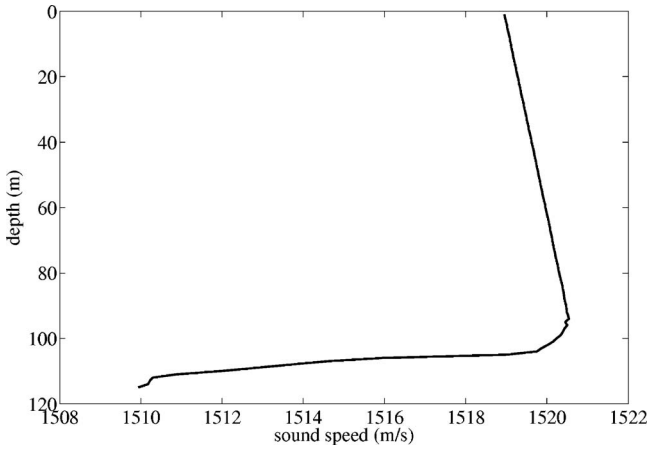


FIG. 4. Sound speed profile taken from a Conductivity Temperature Depth (CTD) cast at position (42.943° N 10.127° E) on 28 November 2000.

shifted according to the Doppler factor of $(1+v/c)$, where c is the sound speed, and v is the ship velocity (positive when moving towards the source, and negative when moving away). With the ship speed at 2 m/s, this corresponds to a maximum shift of about 8 bins at 500 Hz.

Figure 5 shows the one-dimensional posterior probability distribution plots of the geometric and sediment parameters for an inversion performed at frame 195 where the range between the source and the first hydrophone is 750 m. The plots indicate the uncertainty existing in the parameters obtained from the geoacoustic inversion results.

VI. TRANSMISSION LOSS (TL) ESTIMATION FROM INVERSION RESULTS

In this section, the mapping of the uncertainties in the estimation of the seabed parameters resulting from the geoacoustic inversion to uncertainties in the transmission loss domain is discussed.

A. Predictive distribution of transmission loss

The MCMC method using the Metropolis-Hastings algorithm is applied to the geoacoustic inversion results to compute the TL uncertainty as a result of uncertainty in the environmental parameters. The results were obtained based on the posterior probability distributions of the parameters obtained for frame 195 (Fig. 5). The posterior predictive distribution of TL for the position (r_i, z_i) is obtained by integrating the predictions of TL with respect to the posterior distribution of the model parameters, using Eq. (14). Figure 6(a) shows the TL uncertainty plot at a receiver depth of 60 m over range at a frequency of 300 Hz. Figures 6(b) and 6(c) show the uncertainty spread in TL at a region of destructive and constructive interference at 1030 and 1510 m, respectively. The TL uncertainty band is about 5 dB around the region of constructive interference and widens to approximately 15 dB around the region of destructive interference, i.e., regions of destructive interference are predicted with much more uncertainty than regions of constructive interference. Figure 6(d) summarizes the predictive distribution by

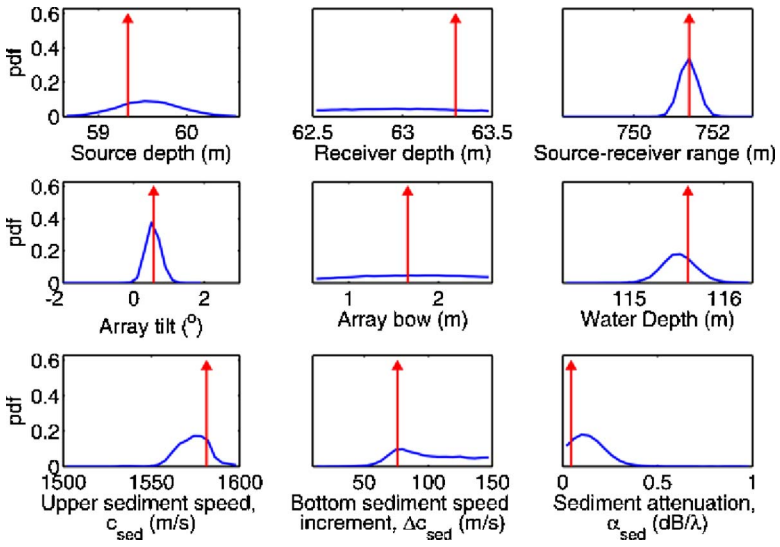


FIG. 5. (Color online) Posterior probability distributions for geometric parameters and sediment parameters, obtained from inversion for frame 195. The arrow indicates the MAP value.

the median (heavy line) and the 90% credibility interval (gray area). This is a practical way to convey the uncertainty in TL.

B. Experimental comparisons

The predictive TL distributions obtained in the preceding subsection are now compared with the actual TL observations. Bayesian inference provides the posterior distribution of the full parameter vector. To estimate the statistical properties of TL, only the posterior probability distribution of the input parameters to the forward model are required. These parameters include the geoacoustic parameters and water depth (water depth is included as it affects the number of propagating modes in the ocean waveguide). Source depth is also an important parameter for predicting TL fields accurately. In this data set, the matched-field estimated source depths varied with a mean of 52.3 m and standard deviation of 3.2 m. Other geometric parameters such as range, receiver depth, array bow and tilt are not used as input parameters to the forward model, and are removed from the parameter vector **m** [see Eq. (18)]. The initial prediction was performed at three frequencies, 300, 400, and 500 Hz, with the source depth fixed at the value obtained at the time of the inversion of frame 195. Figure 7(a) compares the observed TL (crosses) with the initial TL prediction statistics (solid line with gray area) for the frequencies 300, 400, and 500 Hz and for an array depth of 55 m. The initial prediction was not satisfactory as uncertainties in parameters such as the source depth were not modeled, resulting in 30–50% of the observed TL values falling outside the 90% credibility interval (CI). As frequency increases, larger spreads in TL predictions also are observed. This is most pronounced near regions of destructive interference.

Next, the uncertainty in the source depth was modeled using a Gaussian model with a mean of 52.3 m and standard deviation of 3.2 using the statistics from the matched field estimated source depths, and introduced into the parameter vector **m** (see also Sec. III B). The TL prediction process was then repeated with the inclusion of the source depth uncertainties. There is now a marked improvement with >80% of the observed TL values falling within the predicted 90% CI. The TL prediction plot is shown in Fig. 7(b).

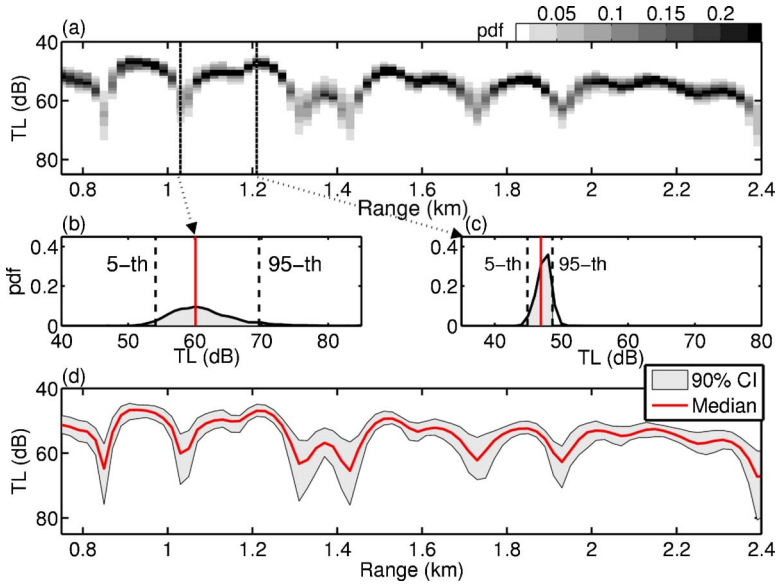


FIG. 6. (Color online) Posterior distribution of TL versus range for 300 Hz at 60 m array depth. (a) Contour of posterior distribution for TL versus range. (b) Region of destructive interference. (c) Region of constructive interference. (d) Statistics of the predicted TL versus range. The heavy line and the surrounding gray area represent the median and the 90% credibility interval, respectively.

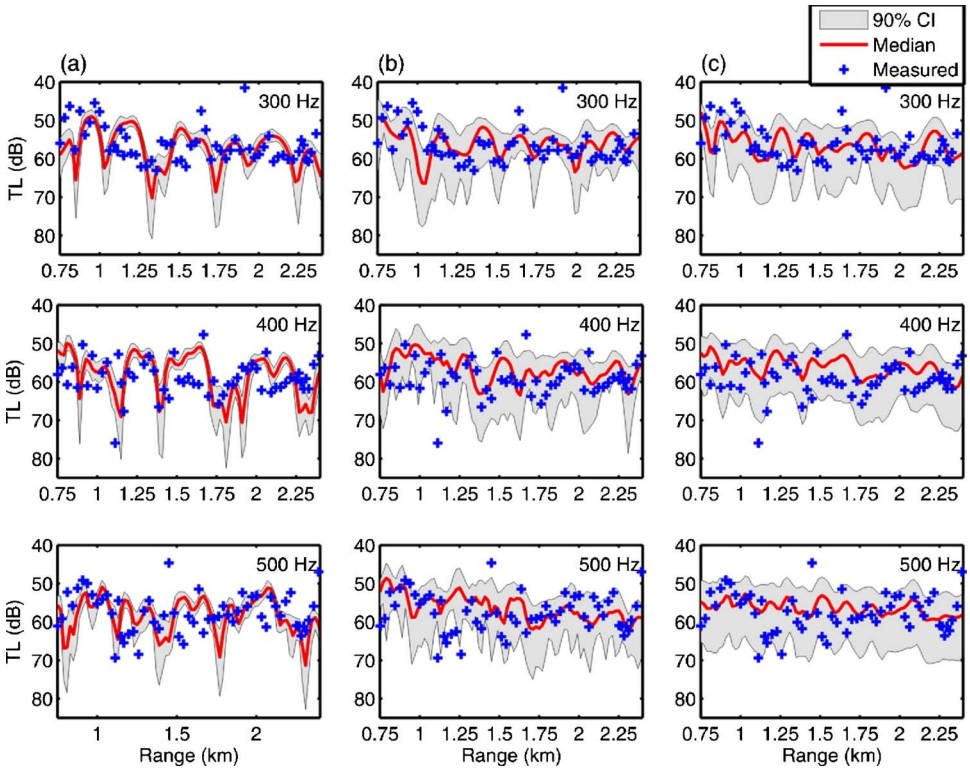


FIG. 7. (Color online) Predicted and measured TL (crosses) for frequencies 300, 400, and 500 Hz. The median of the predicted TL (heavy line) is shown together with the 90% CI (gray area). (a) Initial prediction. (b) Prediction with uncertainty introduced in source depth. (c) Prediction with uncertainty introduced in sound speed profile.

The ocean sound speed profile has been treated as known, but it also could be treated as an uncertain parameter [see Eq. (16)]. To observe the effects of sound speed profile variability on TL prediction, the variability was modeled with a Gaussian model with the measured values at each depth as the mean, and a uniform standard deviation along the whole water column (from 0 to 110 m with a 10 m spacing). Figure 7(c) shows the predicted results using a standard deviation of 1.7 m/s for the whole water column. This value gives >80% of the observed TL values falling within the predicted 90% CI.

Figure 8 plots on the right the variation of the percentage of the observed TL values falling within the predicted 90% CI versus the standard deviation of the modeled sound speed profile at 300 Hz. To the left, the median over range of the TL spread at a single range is plotted. As expected, the median of the spread of the predicted TL and the percentage of the observed TL values falling within the predicted 90% CI increases as the standard deviation increases. Table I summarizes the comparison of the measured and predicted TL from Fig. 7.

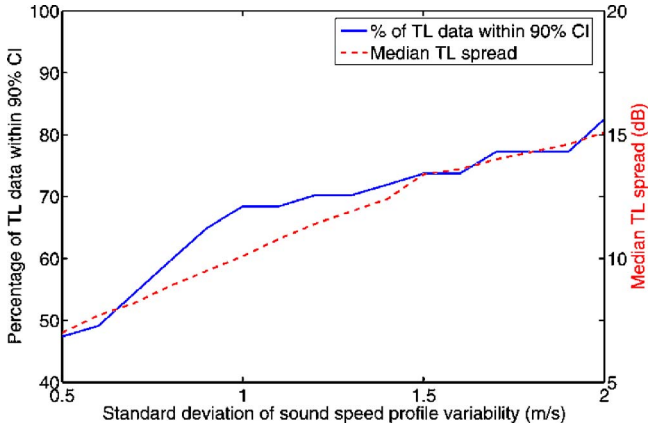


FIG. 8. (Color online) TL variation at 300 Hz versus standard deviation of the modeled sound speed profile. Right axis (solid): Variation of the percentage of the observed TL falling within the predicted 90% CI. Left axis (dashed): The median over range of the predicted TL spread at each range.

Complicated environments, such as spatial and temporal fluctuations in the water column, sediment, sea surface and water-sediment interface, are not modeled and this will increase the error. Further, not all noise sources have been taken into account. Therefore, the percentage of observed data points within the computed 90% CI is less than predicted.

The TL measurement was carried out with a different source depth at each range, due to the motion of the ship (the matched-field inversion indicated a standard deviation of 3 m). It is the uncertainty in source depth that is reflected in the large band of TL uncertainty in Fig. 7(b). Thus, it is expected that if the true source depth at each range was included in the statistical TL prediction, then the predicted TL would lie in a much narrower credibility band (about 8 dB as in Fig. 7(a)) and contain the same fraction of the measured TL. The statistical prediction of TL also could be used in a sensitivity study to explore the importance of each parameter. In such an approach, the reduction in TL uncertainty if one parameter was fixed could be explored systematically.

VII. CONCLUSION

In this paper, the use of geoacoustic inversion results to estimate statistically transmission loss are demonstrated for data received on a towed horizontal array.

A frequency-coherent likelihood function based on a noninformative prior probability distribution was derived for

TABLE I. Summary of transmission loss prediction performance at depth of 55 m. Numbers in dB indicated the median value of the predicted TL spread over range, while numbers in % represent the percentage of measured TL points that lie within the 99% CI.

Frequency (Hz)	Initial prediction	With uncertainty introduced to source depth	With uncertainty introduced to sound speed profile
300	49.1% / 5.2 dB	87.7% / 13.2 dB	84.2% / 15.3 dB
400	50.9% / 5.2 dB	87.7% / 15.5 dB	91.2% / 15.5 dB
500	59.7% / 6.0 dB	84.2% / 14.2 dB	89.5% / 16.4 dB

the horizontal array. A Markov chain Monte Carlo sampling was used to sample the posteriori distribution. Then, the posterior probability density of environmental parameters is utilized for the statistical estimation of TL. This is done by mapping of the probabilities from the geoacoustic domain to the transmission loss domain. It also is shown how additional uncertainty can be incorporated into the TL uncertainty.

The approach is demonstrated on towed array data from a sea trial in the North Elba area. Parameter uncertainties obtained from geoacoustic inversion are mapped into the transmission loss domain, where the probability distribution of transmission loss over different ranges and frequencies is obtained. The characteristic features such as the median and lower/upper percentiles from the distribution also are extracted. The predicted TL statistics are compared with TL measurements from the MAPEX2000 experiment. It is observed that more than 80% of the measured TL data fall within 90% of the range-varying predicted TL probability distribution, demonstrating that the statistical estimation approach presented is reasonable.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-05-1-0264. The data were collected by NATO Undersea Research Center NURC during the MAPEX2000 experiment. The assistance rendered by the Scientist-in-Charge Martin Siderius and the Engineering support staff onboard the NRV Alliance are greatly appreciated.

[1]D. F. Gingras and P. Gerstoft, "Inversion for geometric and geoacoustic parameters in shallow water: Experimental results," J. Acoust. Soc. Am. **97**, 3589–3598 (1995).
[2]M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).
[3]D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).
[4]L. Jaschke and N. R. Chapman, "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**, 1838–1851 (1999).
[5]D. M. F. Chapman, "What are we inverting for," in *Inverse Problems in Underwater Acoustics*, edited by M. I. Taroudakis and G. N. Makrakis, 1–14 (Springer-Verlag, New York, 2001), pp. 1–14.
[6]A. Caiti, S. M. Jesus, and A. Kristensen, "Geoacoustic seafloor exploration with a towed array in a shallow water area of the Strait of Sicily," IEEE J. Ocean. Eng. **21**, 355–366 (1996).
[7]M. R. Fallat, P. L. Nielsen, S. E. Dosso, and M. Siderius, "Geoacoustic characterization of a range-dependent ocean environment using towed array data," IEEE J. Ocean. Eng. **30**, 198–206 (2005).
[8]D. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, "Geoacoustic inversion of tow-ship noise via near-field-matched-field processing," IEEE J. Ocean. Eng. **28**, 454–467 (2003).
[9]D. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[10]D. Tollefsen, S. E. Dosso, and M. J. Wilmut, "Matched-field geoacoustic inversion with a horizontal array and low-level source," J. Acoust. Soc. Am. **120**, 221–230 (2006).
[11]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Validation of statistical estimation of transmission loss in the presence of geoacoustic inversion uncertainty," J. Acoust. Soc. Am. **120**, 1932–1941 (2006).
[12]P. Gerstoft, C.-F. Huang, and W. S. Hodgkiss, "Estimation of transmission loss in the presence of geoacoustic inversion uncertainty," IEEE J. Ocean. Eng. **31**, 299–307 (2006).
[13]P. Abbot and I. Dyer, "Sonar performance predictions based on environmental variability," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Dordrecht, 2002), pp. 611–618.
[14]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).
[15]S. Finette, "Embedding uncertainty into ocean acoustic propagation models," J. Acoust. Soc. Am. **117**, 997–1000 (2005).
[16]P. Gerstoft and D. F. Gingras, "Parameter estimation using multifrequency range-dependent acoustic data in shallow water," J. Acoust. Soc. Am. **99**, 2839–2850 (1996).
[17]F. B. Jensen and M. C. Ferla, *SNAP: the SACLANTCEN normal-mode acoustic propagation model*, SACLANT Undersea Research Center, SM–121, La Spezia, Italy (1979).
[18]L. T. Fialkowski, T. C. Yang, K. Yoo, E. Kim, and D. K. Dacol, "Consistency and reliability of geoacoustic inversions with a horizontal line array," J. Acoust. Soc. Am. **120**, 231–246 (2006).
[19]P. Gerstoft, *SAGA Users guide 5.0, an inversion software package*, An updated version of "SAGA 2.0," SACLANT Undersea Research Centre, SM–333, La Spezia, Italy (1997).
[20]P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[21]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **8**, 259–270 (2000).
[22]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[23]C. Soares and S. M. Jesus, "Broadband matched-field processing: Coherent and incoherent approaches," J. Acoust. Soc. Am. **113**, 2587–2598 (2003).
[24]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversions," J. Acoust. Soc. Am. **119**, 197–207 (2006).
[25]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "On the effect of error correlation on matched-field geoacoustic inversion," J. Acoust. Soc. Am. **121**, EL64–69 (2007).
[26]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge University Press, London, 1992).
[27]W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).
[28]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion I: A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[29]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E.

Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[30]W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," Biometrika **57**, 97–109 (1970).
[31]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).
[32]C. J. Geyer, "Practical Markov chain Monte Carlo (with discussion)," Stat. Sci. **7**, 473–483 (1992).
[33]C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer-Verlag, New York, 1999).
[34]C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Estimation of radio refractivity from radar clutter using Bayesian Monte Carlo analysis," IEEE Trans. Antennas Propag. **54**, 1318–1327 (2006).

A high-frequency warm shallow water acoustic communications channel model and measurements

Mandar Chitre[a)]
Acoustic Research Laboratory, National University of Singapore, 12A Kent Ridge Road, Singapore 119223

(Received 30 April 2007; revised 8 August 2007; accepted 9 August 2007)

Underwater acoustic communication is a core enabling technology with applications in ocean monitoring using remote sensors and autonomous underwater vehicles. One of the more challenging underwater acoustic communication channels is the medium-range very shallow warm-water channel, common in tropical coastal regions. This channel exhibits two key features—extensive time-varying multipath and high levels of non-Gaussian ambient noise due to snapping shrimp—both of which limit the performance of traditional communication techniques. A good understanding of the communications channel is key to the design of communication systems. It aids in the development of signal processing techniques as well as in the testing of the techniques via simulation. In this article, a physics-based channel model for the very shallow warm-water acoustic channel at high frequencies is developed, which are of interest to medium-range communication system developers. The model is based on ray acoustics and includes time-varying statistical effects as well as non-Gaussian ambient noise statistics observed during channel studies. The model is calibrated and its accuracy validated using measurements made at sea.
 [DOI: 10.1121/1.2782884]

PACS number(s): 43.30.Zk, 43.30.Cq [EJS] Pages: 2580–2586

I. INTRODUCTION

The ability to communicate effectively underwater has numerous applications for marine researchers, oceanographers, marine commercial operators, off-shore oil industry, and defense organizations. As electromagnetic waves cannot propagate over long distances in seawater, underwater acoustic communication is the core enabling technology for such applications. Underwater acoustic communications has been a difficult problem due to unique channel characteristics such as fading, extended multipath, and refractive properties of the sound channel (Stojanovic, 1996). Attempts at adapting communication techniques developed for other channels have yielded successful implementations in vertical deep water channels, but have had limited success in shallow water channels (Stojanovic, 1996). Although considerable progress has been made in shallow water communication over the past decade, the medium range channel of very shallow warm water, common in tropical coastal regions such as Singapore waters, still poses a challenge to most communication modems (Chitre, 2006). This warm shallow water acoustic (WSWA) channel presents two key features—extensive time-varying multipath (Chitre *et al.*, 2004) and high levels of non-Gaussian ambient noise due to snapping shrimp (Potter *et al.*, 1997a, b; Chitre *et al.*, 2004)—both of which limit the performance of traditional communication techniques. A good understanding of the communications channel is important in the design of a communication system. It aids in the development of signal processing techniques as well as in the testing of the techniques via simulation. In this article, we present a channel model for the WSWA channel and validate it with field experiments.

Ray theory and the theory of normal modes have provided the basis for channel modeling (Coates, 1989). At high frequencies, ray tracing is an appropriate model and is commonly used to determine the coarse multipath structure of the channel. The surface movement and environmental changes can contribute significantly to the variability of a channel. This variability has been modeled and measured (Owen *et al.*, 1994; Badiey *et al.*, 2000). A simpler stochastic channel model has also been developed (Galvin and Coates, 1994). The model has been validated against experimental results from a very shallow water channel in Southampton, UK. In wireless communications, it is common to model a multipath channel using a tapped delay line with tap gains as stochastic processes (Proakis, 1995). When the tap gains are modeled using complex Gaussian processes, the resulting channel is the well-known Rayleigh fading channel. Some researchers consider the shallow water medium range channel to exhibit Rayleigh fading (Catipovic, 1990). Experimental support for such claims is found (Galvin and Coates, 1994). However, the model has been challenged by others (Essebbar *et al.*, 1994; Geng and Zielinski, 1995). A hybrid model where the multipath tap delays are computed using a ray model and the tap variation is modeled using Ricean statistics is proposed (Geng and Zielinski, 1995). The motivation for such a channel model is discussed in depth but no results or validation is presented. Although the model is physics based, the model does not include known acoustic propagation physics such as spreading, absorption, etc. Based on a review of literature, currently there seems to be no consensus among researchers as to which channel model is best suited for very shallow water medium range channels.

In warm shallow waters, the ambient noise beyond about 2 kHz is commonly dominated by snapping shrimp (Potter *et al.*, 1997a, b). Snapping shrimp are impulsive sources and

a)Electronic mail: mandar@arl.nus.edu.sg

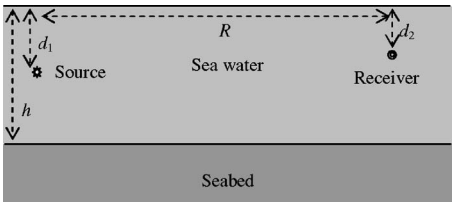


FIG. 1. Schematic showing a Pekeris waveguide model of the WSWA channel.

the pressure amplitude distribution is known to be described well by the symmetric α-stable (SαS) distribution (Chitre *et al.*, 2004). The noise amplitude distribution (pressure amplitude in case of acoustics) plays an important role in the analysis of communication schemes. A linear communication receiver designed with a Gaussian noise assumption is suboptimal in the presence of non-Gaussian noise. Armed with the knowledge of the noise distribution, optimal or near-optimal nonlinear receivers (Tsihrintzis and Nikias, 1995; Chitre *et al.*, 2006) and decoders (Chitre *et al.*, 2007) can be designed for improved communication performance.

In this article, we develop a physics-based channel model for the WSWA channel. The model is based on ray propagation in a well-mixed channel of constant depth. We model each ray to independently fade as described by the Rayleigh distribution. The arrival time of each ray is modeled to statistically vary over time. An additive SαS noise model is adopted to describe the ambient noise. The resulting channel model is calibrated and its accuracy validated using measurements made in Singapore waters.

II. CHANNEL MODEL

In this section, we develop a mathematical model for signal propagation through the WSWA channel. The model is primarily based on the physics of ocean acoustics. In addition, the model includes some statistical variations that are not included in the physics models that we have incorporated.

A. Physics of high frequency underwater acoustic propagation

Acoustic propagation in the ocean is governed by the wave equation. As solutions to the wave equation are difficult to find in general cases, approximations are often used to model propagation. The ray theory provides one such approximation, commonly used for high frequency propagation modeling (Brekhovskikh and Lysanov, 1991). As medium range communication signals are usually high frequency, we use the ray model as a basis for our model of channel propagation. In the ray model, sound energy is conceptualized as propagating along *rays*, which are straight lines in the case of a fluid medium with a constant sound speed (isovelocity). The rays are partially reflected and partially refracted when they encounter a discontinuity in sound speed. We model the WSWA channel as a Pekeris waveguide, consisting of an isovelocity layer over an isovelocity half-space (Fig. 1). The isovelocity layer models the seawater while the isovelocity half-space models the seabed. The isovelocity assumption for seawater is justified as WSWA channels are usually well mixed and have relatively small increase in pressure over the depth of the water column. The assumption is further supported by numerous sound speed measurements in Singapore waters (Chitre, 2006). The sea surface is modeled as a pressure-release boundary.

Let d_1 be the depth of the source, d_2 be the depth of the receiver, h be the height of the water column, and R be the transmission range. The distance traveled by the sound along various eigenrays can be computed using the method of images (Brekhovskikh and Lysanov, 1991). The distance along the direct eigenray is denoted by D_{00} given by

$$D_{00} = \sqrt{R^2 + (d_1 - d_2)^2}. \tag{1}$$

Let D_{sb} be the distance along an upward originating eigenray with s surface reflections and b bottom reflections. For such eigenrays, $0 \leqslant s-b \leqslant 1$ and

$$D_{sb} = \sqrt{R^2 + [2bh + d_1 - (-1)^{s-b} d_2]^2}. \tag{2}$$

Let $\underline{D}_{sb}$ be the distance along a downward originating eigenray with s surface reflections and b bottom reflections. For such eigenrays, $0 \leqslant b-s \leqslant 1$ and

$$\underline{D}_{sb} = \sqrt{R^2 + [2bh - d_1 + (-1)^{b-s} d_2]^2}. \tag{3}$$

We assume that our source is omnidirectional and therefore produces a spherical wavefront in an isovelocity medium. The energy intensity at any point along the wavefront will then reduce as the square of the distance traveled by the wave (Gauss' theorem). A factor representing the apparent loss in pressure amplitude due to spherical spreading along an eigenray of length D can be written as

$$L_{SS}(D) = \sqrt{\frac{1}{D^2}} = \frac{1}{D}. \tag{4}$$

When sound propagates in the ocean, part of the acoustic energy is continuously transformed into heat. This absorption is primarily due to the volume viscosity as a result of relaxation processes in seawater. An empirical expression for the attenuation coefficient β (dB/km) at a frequency f (kHz, between 3 and 500 kHz), salinity S (in ‰) and hydrostatic pressure P (kg/cm^2) is given by Brekhovskikh and Lysanov, 1991. At a nominal depth of 10 m, the hydrostatic pressure P is approximately 2×10^5 Pa (i.e., 2 kg/cm^2). Based on the attenuation coefficient, a loss factor (in pressure amplitude) can be computed to account for the absorption at distance D along an eigenray. We have

$$\begin{aligned} L_A(D) &= 10^{-[(D/1000)\beta]/20} = \exp\left[-\frac{D}{20\,000} \beta \log_e 10 \right] \\ &= \exp\left[-\frac{\log_e 10}{20\,000} D 8669 \left(\frac{S A f_T f^2}{f_T^2 + f^2} + \frac{B f^2}{f_T} \right) \right] \\ &= \exp\left[-0.998 D \left(\frac{S A f_T f^2}{f_T^2 + f^2} + \frac{B f^2}{f_T} \right) \right]. \end{aligned} \tag{5}$$

The attenuation coefficient does not change significantly with small changes in depth. As the depth is constrained in a very shallow water channel, we may use this expression in WSWA channels without a significant loss in accuracy.

The impedance mismatch between the seawater and air causes the sea surface to be a very good reflector. If the sea surface is calm, the reflection is close to perfect but includes a phase shift by π radians, i.e., the reflection coefficient is -1 (Brekhovskikh and Lysanov, 1991). If the sea surface is rough (due to waves), a small loss will be incurred for every surface interaction. We model this loss by allowing a constant loss factor of L_{SR} per surface interaction.

The impedance mismatch between the seawater and seabed causes the sea bottom to reflect some of the sound incident on it. Let ρ and c be the density and sound speed in seawater respectively. Let ρ_1 and c_1 be the density and sound speed in the seabed, respectively. For a smooth sea bottom, the reflection is angle dependent and is given by the Rayleigh reflection coefficient as (Brekhovskikh and Lysanov, 1991)

$$L_B(\theta) = \left| \frac{m \cos\theta - \sqrt{n^2 - \sin^2\theta}}{m \cos\theta + \sqrt{n^2 - \sin^2\theta}} \right|, \tag{6}$$

where

$$m = \frac{\rho_1}{\rho}, \quad n = \frac{c}{c_1}.$$

The angle of incidence θ can be computed based on the geometry of the Pekeris waveguide. Let angle θ_{sb} correspond to an eigenray D_{sb} and angle $\underline{\theta}_{sb}$ correspond to an eigenray $\underline{D}_{sb}$. Then, we have

$$\theta_{sb} = \tan^{-1}\left(\frac{R}{2bh + d_1 - (-1)^{s-b} d_2}\right),$$

$$\underline{\theta}_{sb} = \tan^{-1}\left(\frac{R}{2bh - d_1 + (-1)^{b-s} d_2}\right). \tag{7}$$

For a rough or absorbing sea bottom, additional reflection losses may be incurred. We model these losses by allowing an additional constant loss factor of L_{BR} per bottom interaction.

B. Time-varying statistical effects

The physics based model described above yields a time-invariant model of the channel. However, the WSWA channel is a time-varying channel. To model the time variation, we introduce some statistical variations in the propagation model.

In line with our initial experimental findings (Chitre *et al.*, 2004), we model the eigenray amplitude as a Rayleigh random process with a median determined by the physics based model and a time-correlation determined by the Doppler spread W_d. Rayleigh fading usually occurs in a sum of a large number of multipath. Rayleigh fading observed on each individual eigenray is thought to occur due to micro-multipath as the eigenray is scattered by small inhomogeneities in the medium and other suspended scatterers (Dashen *et al.*, 1979). This leads to a statistical tapped delay line model which is a special case of the model proposed in Geng and Zielinski (1995), where the tap variation is modeled using Ricean statistics.

The arrival time lag of each eigenray with respect to the direct arrival is related to the difference in path length along the eigenray and the direct path. However, the arrival time lag exhibits variation over time, probably due to movement of the source, receiver and the surface. The stability of the arrival time lag with respect to small changes in source/receiver position can be analyzed using the ray model. Let τ_{sb} be the arrival time lag of eigenray D_{sb} and $\underline{\tau}_{sb}$ be the arrival time lag of eigenray $\underline{D}_{sb}$. We have

$$\tau_{sb} = \frac{D_{sb} - D_{00}}{c},$$

$$\underline{\tau}_{sb} = \frac{\underline{D}_{sb} - D_{00}}{c}. \tag{8}$$

From Eqs. (1)–(3) and (9), we have

$$\tau_{sb} = \frac{1}{c}\left[\sqrt{R^2 + [2bh + d_1 - (-1)^{s-b} d_2]^2} - \sqrt{R^2 + (d_1 - d_2)^2}\right],$$

$$\underline{\tau}_{sb} = \frac{1}{c}\left[\sqrt{R^2 + [2bh - d_1 + (-1)^{b-s} d_2]^2} - \sqrt{R^2 + (d_1 - d_2)^2}\right]. \tag{9}$$

As we are interested in medium range communications in shallow waters, we assume $R \gg d_1$, $R \gg d_2$ and $R \gg h$. Using the Taylor series expansion $\sqrt{1+x} \approx 1 + \frac{1}{2}x$ if $|x| \ll 1$, we have

$$\begin{aligned}
\tau_{sb} &= \frac{R}{c}\left[\sqrt{1 + \left[\frac{2bh + d_1 - (-1)^{s-b} d_2}{R}\right]^2} - \sqrt{1 + \left(\frac{d_1 - d_2}{R}\right)^2}\right] \approx \frac{R}{c}\left[1 + \frac{1}{2}\left[\frac{2bh + d_1 - (-1)^{s-b} d_2}{R}\right]^2 - 1 - \frac{1}{2}\left(\frac{d_1 - d_2}{R}\right)^2\right] \\
&= \frac{1}{2Rc}\left[(2bh + d_1 - (-1)^{s-b} d_2)^2 - (d_1 - d_2)^2\right] \\
&= \frac{2}{Rc}\left[b^2h^2 + bhd_1 - (-1)^{s-b} bhd_2 + (s-b)d_1 d_2\right].
\end{aligned} \tag{10}$$

Similarly,

$$\underline{\tau}_{sb} \approx \frac{2}{Rc}\left[b^2h^2 - bhd_1 + (-1)^{b-s} bhd_2 + (b-s)d_1 d_2\right].$$

To test the stability of arrival time lag, we differentiate the previous equation with respect to the source depth d_1, receiver depth d_2 and range R to obtain

$$\frac{\partial \tau_{sb}}{\partial R} = \frac{\partial \underline{\tau}_{sb}}{\partial R} \approx 0, \tag{11}$$

i.e., the arrival time jitter is, to first order, independent of the range, and

$$\frac{\partial \tau_{sb}}{\partial d_1} = \frac{2bh}{Rc} + (s-b)\frac{2d_2}{Rc},$$

$$\frac{\partial \tau_{sb}}{\partial d_2} = -(-1)^{s-b}\frac{2bh}{Rc} + (s-b)\frac{2d_1}{Rc},$$

$$\frac{\partial \underline{\tau}_{sb}}{\partial d_1} = -\frac{2bh}{Rc} + (b-s)\frac{2d_2}{Rc},$$

$$\frac{\partial \underline{\tau}_{sb}}{\partial d_2} = (-1)^{b-s}\frac{2bh}{Rc} + (b-s)\frac{2d_1}{Rc}. \tag{12}$$

Using Δ as a small change operator, we can write the change in arrival time lag as a function of the change of the source depth, receiver depth and range as

$$\Delta\tau_{sb} = \frac{\partial \tau_{sb}}{\partial R}\Delta R + \frac{\partial \tau_{sb}}{\partial d_1}\Delta d_1 + \frac{\partial \tau_{sb}}{\partial d_2}\Delta d_2,$$

$$\Delta\underline{\tau}_{sb} = \frac{\partial \underline{\tau}_{sb}}{\partial R}\Delta R + \frac{\partial \underline{\tau}_{sb}}{\partial d_1}\Delta d_1 + \frac{\partial \underline{\tau}_{sb}}{\partial d_2}\Delta d_2. \tag{13}$$

As the derivatives with respect to R are 0, the arrival time lags are not sensitive to small changes in range. If we model Δd_1 and Δd_2 as Gaussian random variables with variance σ_d^2, we can estimate the variance of the arrival time lags as

$$\sigma_{sb}^2 = \sigma^2(\tau_{sb}) = \left|\frac{\partial \tau_{sb}}{\partial d_1}\right|^2\sigma_d^2 + \left|\frac{\partial \tau_{sb}}{\partial d_2}\right|^2\sigma_d^2 = \frac{4\sigma_d^2}{R^2c^2}[2b^2h^2 + (s-b)(d_1^2 + d_2^2 + 2bhd_1 + 2bhd_2)]. \tag{14}$$

Similarly,

$$\underline{\sigma}_{sb}^2 = \sigma^2(\underline{\tau}_{sb}) = \frac{4\sigma_d^2}{R^2c^2}[2b^2h^2 + (b-s)(d_1^2 + d_2^2 - 2bhd_1 - 2bhd_2)].$$

From this expression, we expect that we have less arrival time lag jitter at further ranges than nearer ranges. We also expect that the arrival time lag jitter increases as the number of surface and bottom interactions of the eigenray increases.

C. The channel model

In the previous two sections, we have developed the components of a time-varying channel propagation model. We now put the components together. Let $x(t)$ be a signal transmitted through the channel and $y(t)$ be the corresponding received signal. Ignoring the absolute time delay between transmission and reception, we can express $y(t)$ in terms of $x(t)$ as

$$y(t) = A_{00}(t)L_{SS}(D_{00})L_A(D_{00})x(t) + \sum_{s=1}^{\infty}\sum_{b=s-1}^{s} A_{sb}(t)L_{SS}(D_{sb})L_A(D_{sb}) \times (-L_{SR})^s L_{BR}^b L_B(\theta_{sb})^b x(t-\tau_{sb}+J_{sb}(t)) + \sum_{b=1}^{\infty}\sum_{s=b-1}^{b} \underline{A}_{sb}(t)L_{SS}(\underline{D}_{sb})L_A(\underline{D}_{sb})(\times - L_{SR})^s L_{BR}^b L_B(\underline{\theta}_{sb})^b x(t-\underline{\tau}_{sb}+\underline{J}_{sb}(t)) + n(t),$$

$$\therefore y(t) = \frac{A_{00}(t)L_A(D_{00})}{D_{00}}x(t) + \sum_{s=1}^{\infty}\sum_{b=s-1}^{s}\frac{A_{sb}(t)L_A(D_{sb})(-L_{SR})^s L_{BR}^b L_B(\theta_{sb})^b}{D_{sb}}x(t - \tau_{sb} + J_{sb}(t)) + \sum_{b=1}^{\infty}\sum_{s=b-1}^{b}\frac{\underline{A}_{sb}(t)L_A(\underline{D}_{sb})(-L_{SR})^s L_{BR}^b L_B(\underline{\theta}_{sb})^b}{\underline{D}_{sb}}x(t - \underline{\tau}_{sb} + \underline{J}_{sb}(t)) + n(t), \tag{15}$$

where $A_{sb}(t)$ and $\underline{A}_{sb}(t)$ are modeled as independent Rayleigh processes with unit mean and an exponential autocorrelation specified by the Doppler spread W_d. These processes represent the fading of individual eigenpaths. The quantities $J_{sb}(t)$ and $\underline{J}_{sb}(t)$ are random variables denoting the time jitter, modeled as Gaussian processes with zero mean, variance σ_{sb}^2 and an exponential autocorrelation specified by a transducer position coherence time τ_d. The noise $n(t)$ is modeled as a SαS process with characteristic exponent α and dispersion γ (Samorodnitsky and Taqqu, 1994). We assume the noise to be white over the bandwidth of interest. This assumption is reasonable provided the signal bandwidth is not very large.

Although the summations in Eq. (16) have an infinite number of terms, the terms diminish in magnitude and the summations converge. Only terms significantly larger than the ambient noise need to be included in practice. For most channels, it is sufficient to include the first few terms in both summations.

This channel model should be interpreted as a passband channel model. The L_A terms in the model are frequency dependent. If the bandwidth of the signal $x(t)$ is much smaller than the carrier frequency F_c, the L_A terms can be computed at the carrier frequency. For broadband transmission at long distances, this approach yields lower accuracy. However, for simplicity in the analytical model, we use the L_A terms at frequency F_c irrespective of the bandwidth of the signal.

A baseband channel model can also be implemented with some small changes. The L_A terms have to be computed at the appropriate carrier frequency. In addition, the real SαS ambient noise process $n(t)$ needs to be replaced with an isotropic complex SαS noise process (Samorodnitsky and Taqqu, 1994) to accommodate the complex noise requirement at baseband.

This channel model can be interpreted as a tapped delay line with time-varying tap weights and tap delays. Equation (16) can be written in a compact form with N significant terms, appropriate lumped coefficients B_j, delays τ_j, Rayleigh random processes $A_j(t)$, Gaussian random processes $J_j(t)$ and a SαS random process $n(t)$:

$$y(t) = A_0(t)B_0x(t) + \sum_{j=1}^{N-1} A_j(t)B_jx(t - \tau_j + J_j(t)) + n(t). \tag{16}$$

III. EXPERIMENTAL CALIBRATION AND VALIDATION

A. Validation of model against channel measurements

Channel measurements were conducted in February 2004 in Singapore waters to validate the model (Chitre *et al.*, 2004). The chosen location allowed measurements up to 1 km range in a relatively flat area with an average depth of about 15 m. Transmissions were made from an omnidirectional transducer located at the bottom of a 4 m pole mounted on the bow of a research vessel. The signal was received using a hydrophone located at the bottom of a 4 m pole mounted on the side of an anchored barge. The signal was sampled at 250 kSa/s and stored for later analysis. The research vessel moved to various locations and made transmissions. Global positioning system coordinates of the vessel and the barge were noted for range computation. The signal used was a 30 ms direct sequence spread spectrum binary phase-shift keying (BPSK) signal with a bandwidth of 40 kHz centered around 40 kHz. The signal was repeated 100 times at a rate of 10 Hz at each location of the research vessel. The different transmission locations corresponded to ranges of 50, 100, 550, 780, and 1020 m. The measurements were made in relatively calm weather over a period of an hour. The recorded signals were postprocessed using the sign correlator (Chitre *et al.*, 2006) to obtain estimates of the time variability of the multipath structure of the channel. The sign correlator is a near-optimal detector of signals in presence of impulsive snapping shrimp noise. The fading behavior of the signal was determined using the maximum likelihood (ML) estimator (Chitre *et al.*, 2006). The ML estimator is an optimal estimator of signal strength in the presence of snapping shrimp noise with a known probability distribution. The SαS distribution assumed for the ML estimator was calibrated using ambient noise samples recorded at the experimental site.

1. Fading statistics

The fading behavior of the direct arrival at short ranges was determined by estimating the signal strength of 100 μs sections of the received signal using the ML estimator. Doppler spreads in the range of 5–10 Hz were observed for all ranges. Figure 2 shows the fading of a single path as compared to Rayleigh fading at 50 m range. The observed fading is similar, but slightly less severe than predicted by the Rayleigh distribution. At longer ranges, multiple arrivals are often too closely spaced to separate. At 1020 m, the first arrival is formed by the interference of the direct and surface reflected arrival at this range. Hence one would expect that the fading could be explained as a function of the fading of each of the arrivals. As the time difference between the arrivals is small as compared to the reciprocal of the center frequency of the signal, we may assume that the two arrivals interfere

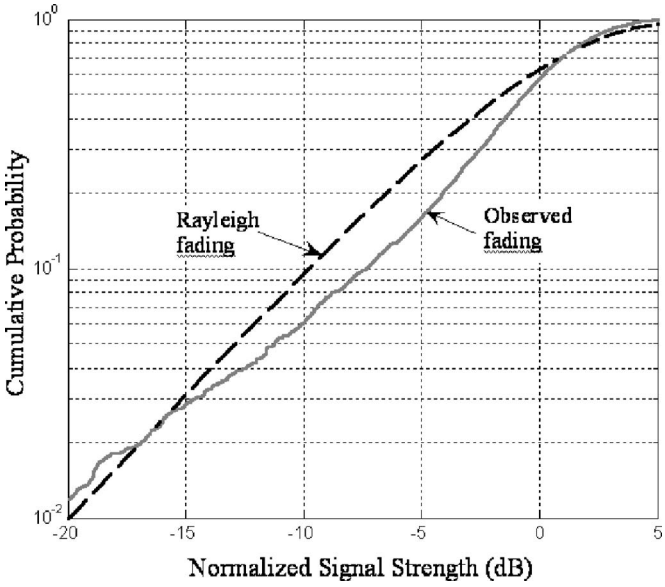


FIG. 2. Cumulative distribution function of signal strength showing fading behavior of direct arrival at 50 m, compared against Rayleigh fading.

destructively as the surface reflection coefficient is −1. The resulting first arrival would then be distributed as the difference of two independent Rayleigh random variables. Figure 3 shows the fading of the first arrival at 1020 m against simulated fading based on the difference of two Rayleigh random variables. The remarkable similarity in the simulated fading and the observed fading, suggests that the simple model presented earlier is a good approximation.

2. Multipath structure

We compare the measured multipath structure in the channel over 100 transmissions against simulation results. Figure 4 shows a comparison between experiment and simulation in a 15 m deep channel at a range of 100 m. Although

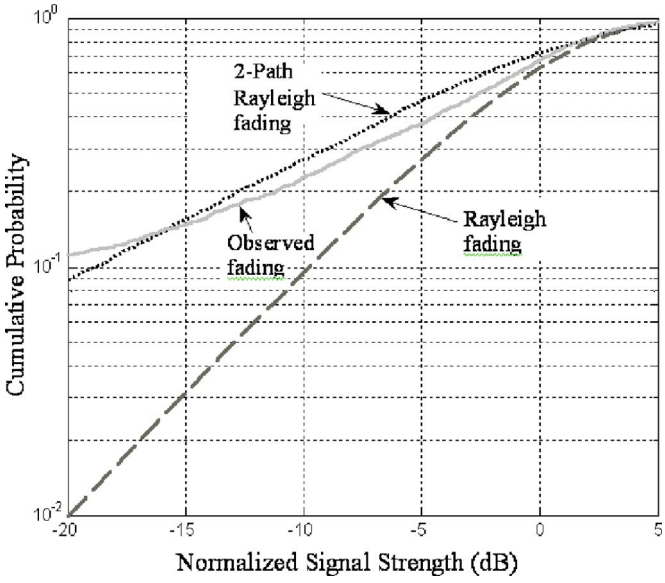


FIG. 3. (Color online) Cumulative distribution function of signal strength showing fading behavior of the combined arrival at 1020 m, compared against Rayleigh fading and simulated two-path Rayleigh fading.

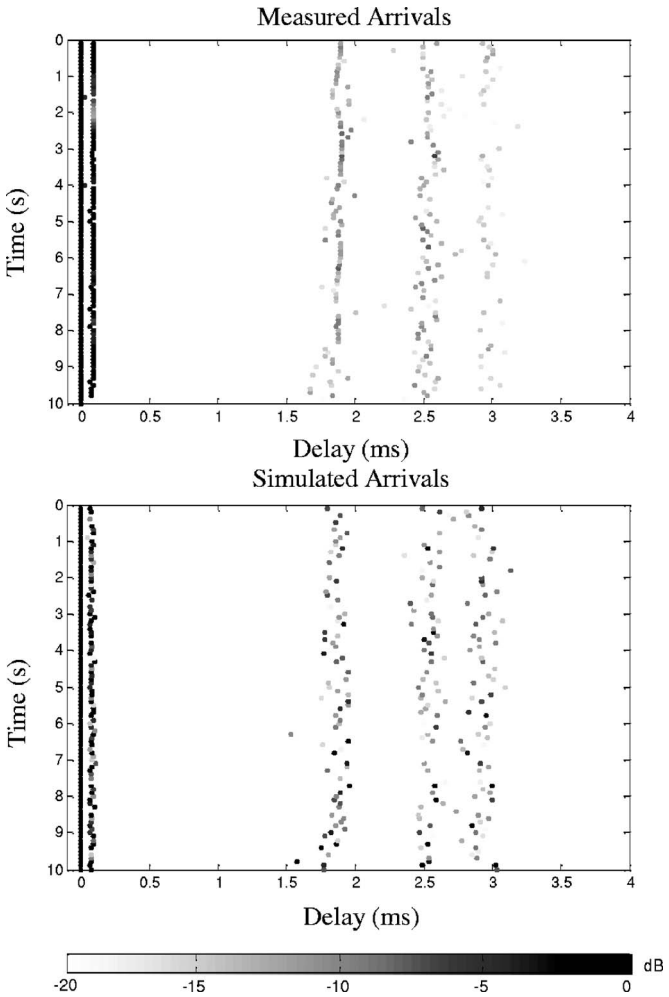


FIG. 4. Comparison of multipath arrival structure between experiment and simulation at 100 m.

similar validations were successfully performed at the longer ranges, we present the results at 100 m due to the high signal-to-noise ratio at this range. This enabled us to study the similarities and differences between the experimental measurements and simulation in more detail than at greater ranges. The simulation parameters were chosen based on our knowledge of the experimental environment. Some channel parameters were tuned slightly to obtain a close match between measurement and simulation. The parameter values used in the simulation are listed in Table I.

Both experimental and simulated arrival structures show five arrivals. The first arrival is the direct arrival, closely followed by the surface reflected arrival and then the bottom reflected arrival. The next three arrivals are a result of multiple surface and bottom interactions. We note that the mean arrival timings (relative to the direct arrival) agree closely between the simulation and measurement, as shown in Table II. The simulation also predicts the spread (standard deviation) of arrival time accurately. The measured data show some correlation between the arrival times of closely spaced iterations for some of the arrivals; this correlation is not captured in our simulation. The correlation is due to the fact that the channel varies slowly over time and hence is correlated over short time intervals. However, this correlation is not important if the transmission time between packets is large.

TABLE I. Parameters used for simulation of channel at 100 m.

Parameter	Symbol	Value
Range[a]	R	100 m
Water depth[a]	h	14.5 m
Source depth[a]	d_1	3 m
Receiver depth[a]	d_2	2 m
Center frequency	F_c	40 kHz
Sampling frequency	F_s	500 kHz
Surface reflection loss[b]	L_{SR}	3 dB
Bottom reflection loss[b]	L_{BR}	10 dB
Water density[a]	ρ	1023 kg/m^3
Water sound speed[a]	c	1539 m/s
Seabed density[b]	ρ_1	1500 kg/m^3
Seabed sound speed[b]	c_1	1650 m/s
Doppler spread[c]	W_d	10 Hz
Ambient noise α[a]	α	1.7
Ambient noise level (integrated over the frequency band of interest)[a]	$\gamma^{1/\alpha}$	120 000 μPa
Transducer RMS movement[c]	σ_d	0.25 m
Transducer position coherence time[c]	τ_d	0.25 s
Number of reflection terms	N	3

[a]The range, water depth, source, and receiver depths, water density, sound speed and ambient noise parameters were measured at the experiment site.
[b]The surface and bottom reflection losses and the seabed density and sound speed were fitted to the measured data. The values are consistent with known seabed type observed in the area.
[c]The Doppler spread and the transducer rms movement and coherence time were estimated based on the recorded data.

The average measured and predicted arrival strengths (relative to the direct arrival) are quite close. The simulation predicts a little more fading than observed in the direct and surface reflected arrival. It also predicts a few stronger arrivals than those observed in the bottom reflected ray and higher order reflected rays. The slight mismatch in fading statistics between observation and simulation is a limitation of our model as we modeled all arrivals with similar fading statistics. Although a more comprehensive fading model could have been developed, we opted to retain our simple fading model, as the proposed channel model seems to capture most of observed phenomenon in the channel. As we shall see in the next section, communication performance predictions based on this model agree closely with experiment.

TABLE II. Comparison of simulated channel statistics against experimental channel statistics at 100 m.

Arrival		Simulation	Measurement
2	Mean arrival timing (ms)	0.08	0.09
	Arrival timing spread (ms)	0.01	0.01
	Mean relative strength (dB)	0	1
3	Mean arrival timing (ms)	1.9	1.9
	Arrival timing spread (ms)	0.08	0.07
	Mean relative strength (dB)	−7	−8
4	Mean arrival timing (ms)	2.5	2.5
	Arrival timing spread (ms)	0.07	0.07
	Mean relative strength (dB)	−9	−9
5	Mean arrival timing (ms)	2.9	3.0
	Arrival timing spread (ms)	0.07	0.09
	Mean relative strength (dB)	−9	−11

TABLE III. Predicted and actual performance results of OFDM communication schemes at various ranges.

Number of carriers	Cyclic prefix length	Code	Communication range	Simulated BER	Experimentally measured BER
128	32	None	800 m	8.1×10^{-2}	8.0×10^{-2}
128	32	1/3	800 m	0.8×10^{-3}	0.7×10^{-3}
128	32	1/6	800 m	$<10^{-4}$	$<10^{-4}$
512	32	None	800 m	1.2×10^{-1}	1.2×10^{-1}
512	32	1/6	800 m	$<10^{-4}$	$<10^{-4}$
128	32	None	1 km	0.8×10^{-1}	1.0×10^{-1}
128	32	1/3	1 km	1.0×10^{-3}	1.5×10^{-3}
128	32	1/6	1 km	$<10^{-4}$	$<10^{-4}$
512	32	None	1 km	1.1×10^{-1}	1.0×10^{-1}
512	32	1/3	1 km	6.1×10^{-3}	6.0×10^{-3}
512	32	1/6	1 km	$<10^{-4}$	$<10^{-4}$

B. Validation of communications simulation against experimental results

Comparing simulated communication results against experimental results further validates the channel model. The communication scheme chosen was an interleaved coded orthogonal frequency-division multiplexing (OFDM) scheme as described in Chitre *et al.* (2005). The OFDM carriers were modulated using differential quadrature phase-shift keying (QPSK) and occupied a 24 kHz band centered at 62 kHz. Various combinations of the number of carriers, cyclic prefix length, and code rates were tested. Randomly generated data packets were coded using a convolution code or a serial concatenated convolution code. These data were interleaved and modulated using the chosen OFDM scheme and transmitted through the channel. At the other end, it was demodulated, deinterleaved and decoded using a 1-norm Viterbi decoder (Chitre *et al.*, 2007). By comparing the generated data against the decoded data, the bit error rate (BER) was estimated. Various transmission ranges were tested in simulation and in experiment.

The results are shown in Table III. The experimentally measured BER values closely agree with the BER predictions based on simulation using the channel model developed in this paper. Similar measurements with other communication schemes and channel parameters such as number of carriers (64, 128, 256, 512), prefix length (32, 64), code rate $(1,1/2,1/3,1/5,1/6)$, and transmission range (500 m, 800 m, 1 km, 1.2 km) showed that the predictions were robust to such changes.

IV. CONCLUSIONS

The primary aim of the channel model developed in this article is to enable the study of communication systems in warm shallow waters. The close agreement obtained between the simulated and experimental communication in warm shallow water channels suggests that our model captures the essential physics of such channels. The model captures the multipath arrival structure and includes statistical effects in the channel such as fading and arrival-time jitter. The model predicts the multipath arrival structure, fading probability distribution and arrival-time probability distribution given the channel geometry. However, the time correlation of the fading is defined by the Doppler spread, which is an input to the model. The Doppler spread has to be measured to calibrate the model and cannot be predicted by the model. The time correlation of the arrival-time jitter is not modeled. This does not affect short transmissions where the arrival-time does not vary over the length of the data packet. However, we expect that this may affect simulations where the data packet length is much larger than the coherence time of the arrival-time jitter.

ACKNOWLEDGMENTS

The author would like to acknowledge the Defence Science & Technology Agency (DSTA), Singapore, for their funding support for this research. He would also like to thank Mr. Shiraz Shahabudeen, Mr. Alan Low, Mr. Koay Teong Beng and Dr. Paul Seekings who helped collect the experimental data used for the validation of the model.

Badiey, M., Mu, Y., Simmen, J. A., and Forsythe, S. E. (**2000**). "Signal variability in shallow-water sound channels," IEEE J. Ocean. Eng. **25**, 492–500.

Brekhovskikh, L. M., and Lysanov, Y. (**1991**). *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, Berlin).

Catipovic, J. A. (**1990**). "Performance limitations in underwater acoustic telemetry," IEEE J. Ocean. Eng. **15**, 205–216.

Chitre, M. (**2006**). "Underwater acoustic communications in warm shallow water channels," *Faculty of Engineering* (National University of Singapore, Singapore).

Chitre, M., Ong, S. H., and Potter, J. (**2005**). "Performance of coded OFDM in very shallow water channels and snapping shrimp noise," IEEE Oceans'05, Washington, DC.

Chitre, M., Potter, J., and Ong, S. H. (**2004**). "Underwater acoustic channel characterisation for medium-range shallow water communications," IEEE Oceans 2004, Kobe, Japan, pp. 40–45.

Chitre, M. A., Potter, J. R., and Ong, S. H. (**2006**). "Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise," IEEE J. Ocean. Eng. **31**(2), 497–503.

Chitre, M. A., Potter, J. R., and Ong, S. H. (**2007**). "Viterbi decoding of convolutional codes in symmetric α-stable noise," IEEE Trans. Commun. (in press).

Coates, R. F. W. (**1989**). *Underwater acoustic systems* (Wiley, New York).

Dashen, R., Munk, W. H., Watson, K. M., Zachariasen, F., and Flatte, S. M. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge, UK).

Essebbar, A., Loubet, G., and Vial, F. (**1994**). "Underwater acoustic channel simulations for communications," IEEE Oceans'94, Brest, France, pp. III/495–III/500.

Galvin, R., and Coates, R. F. W. (**1994**). "Analysis of the performance of an underwater acoustic communication system and comparison with stochastic model," IEEE Oceans'94, Brest, France, pp. III/478–III/482.

Geng, X., and Zielinski, A. (**1995**). "An eigenpath underwater acoustic communication channel model," IEEE Oceans'95, San Diego, pp. 1189–1196.

Owen, R. H., Smith, B. V., and Coates, R. F. W. (**1994**). "An experimental study of rough surface scattering and its effects on communication coherence," IEEE Oceans'94, Brest, France pp. 483–488.

Potter, J. R., Lim, T. W., and Chitre, M. (**1997a**). "Ambient noise environments in shallow tropical seas and the implications for acoustic sensing," Oceanology International 97 Pacific Rim, Singapore, pp. 191–199.

Potter, J. R., Lim, T. W., and Chitre, M. (**1997b**). "High-frequency ambient noise in warm shallow waters," Sea Surface Sound, Southhampton, UK, pp. 45–54.

Proakis, J. G. (**1995**). *Digital Communications* (McGraw-Hill, Singapore).

Samorodnitsky, G., and Taqqu, M. S. (**1994**). *Stable Non-Gaussian Random Processes* (Chapman and Hall, New York).

Stojanovic, M. (**1996**). "Recent advances in high-speed underwater acoustic communications," IEEE J. Ocean. Eng. **21**, 125–136.

Tsihrintzis, G. A., and Nikias, C. L. (**1995**). "Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process," IEEE Trans. Commun. **43**, 904–914.

Acoustic fields of nonplanar radiators

A. P. Medley,[a] D. R. Billson, and D. A. Hutchins
School of Engineering, University of Warwick, Coventry, CV4 7AL, United Kingdom

A. Neild[b]
Department of Mechanical Engineering, Monash University, Victoria 3800, Australia

(Received 3 January 2007; revised 23 July 2007; accepted 10 August 2007)

A theoretical approach is described which predicts the fields of acoustic radiators with a predefined surface topography. This is achieved by dividing the surface of the source into small elements, each of which is oriented parallel to the tangent to the surface at that point. The result is an improved modeling performance, in that it is more efficient and requires far fewer elements compared to other numerical approaches using parallel elements. Theoretical predictions are compared to experimental results from curved electrostatic radiators, to demonstrate that the approach has promise.
 [DOI: 10.1121/1.2783117]

PACS number(s): 43.38.Bs, 43.38.Ja [AJZ] Pages: 2587–2593

I. INTRODUCTION

The radiated field of an acoustic source can be modeled in several different ways. For planar radiators, the traditional approach is to use a solution to Rayleigh's integral[1] in which the radiator is assumed to be comprised of an assembly of point sources. Many authors have proposed solutions to the field radiated by a plane piston, where the surface of the source is in one plane with a constant amplitude,[2] for both continuous and transient excitation, and for either rectangular or circular geometries.[3–5] The result can then be used for more complicated sources. For instance, the result from a pair of rectangular elements can be used to predict the field from two strips on either side of a central axis [see Fig. 1(a)] by simply subtracting the contributions from the two larger sources at each field point. Any source that has symmetry across the axis in one plane can then be simulated from the addition of these new elements shifted spatially. This is demonstrated in Fig. 1(b) for a curved radiator. Authors such as Reibold and Kazys[6,7] have demonstrated good agreement between theory and experimental results using such approximations.

The above-mentioned approach must be used with some care, however, as the number of elements and their size must be chosen appropriately. For a concave focusing radiator with a given fixed geometry, it is known that the axial distance from the source to the focal region within the radiated field decreases, as seen in Fig. 2, for lower drive frequencies.[8,9] This has consequences for modeling, as smaller elements (and hence a greater number of them) must be used for an accurate theoretical prediction at these lower frequencies. Thus, while it is generally accepted that the width of the strip-shaped elements should be less than the radiated wavelength (λ) in air, at low drive frequencies (f), extra care must be taken to ensure that they are small enough for the particular case being investigated, to reduce spatial sampling errors in the predicted field.

Consider, as an example, the radiated field in air from a rectangular device curved in the x direction, as illustrated in Fig. 3. The device has an overall width W and height h, and is divided up as shown into the conventional parallel strip elements of width w. The radius of curvature is R, and the axial distance to this point on-axis is z_f. Using such an approach for a radiator with $W=h=200$ mm and $R=150$ mm, at 10 kHz in air, the on-axis radiated fields are predicted to be those in Fig. 4 for four values of element width (w). It can be seen that the accuracy of the predicted focal point increases as the number of elements increases, such that there is a 20% difference in the results as the value of w becomes a smaller fraction of λ (=34.3 mm) at this frequency. However to achieve this, the number of elements has dramatically increased, increasing the computation time required to resolve the on-axis response by 59%. This would be even more pronounced had the whole field, rather than the on-axis response, been calculated.

It was thus considered that a more efficient approach would be to angle the elementary strips of a curved radiator such that they more accurately represent the overall shape of the transducer, as shown in Fig. 5. This simple solution is very effective, in that the process of rotating the strips means there is no longer the requirement to ensure that $w<\lambda$; thus the number of strips (and hence calculations) can be decreased, along with the computational time required to predict the resulting pressure field. This approach also enables the radiated field from transducers with varying geometries to be predicted with relative ease. Thus, while this step may be simple and obvious in hindsight, it does provide gains in computation efficiency. It is an approach that seemingly has not appeared in the acoustics literature for the calculation of radiated fields from acoustic sources. The attractiveness of the approach is that the normal methods of calculation via the Rayleigh integral can be applied, but modified using angled linear elements.

In the following, the theoretical steps required to predict the radiated field are outlined, along with the approach

[a]Electronic mail: a.p.medley@warwick.ac.uk
[b]Electronic mail: adrian.neild@eng.monash.edu.au

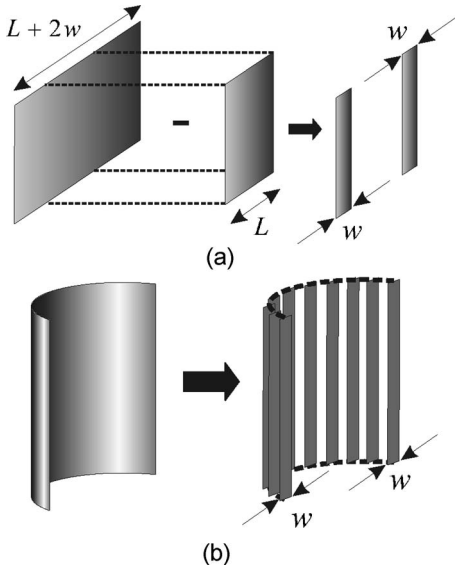


FIG. 1. (a) Calculation of the sound pressure field for a pair of strip sources. (b) Approximation used in the calculation of the theoretical sound pressure field from a cylindrical focusing transducer.

needed to angle the elements, together with some predictions of this approach. A comparison to experiment is also presented.

II. THEORETICAL PREDICTION OF SOUND PRESSURE FIELDS

The nature of the radiating field emanating from a sound source has received a lot of attention due to its importance in analyzing acoustic systems, with several methods being proposed.[2] In this work, a convolution method, based upon the Rayleigh surface integral, will be used, an approach sometimes called the impulse response method.[4] 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, each producing a hemispherical pressure wave. In this way, the pressure field can be modeled using a surface integral over the area of the source. Using the Rayleigh integral the time-dependent velocity potential $\phi(\bar{r},t)$ at a point, $P(l,z)$ can be stated as

$$\phi(\bar{r},t) = \frac{1}{2\pi}\int_s \frac{v(t-R/c)}{R} ds, \tag{1}$$

where c is the speed of sound in the medium and R is the distance between the observation point P and the elementary point ds on the surface of the piston. The expression $v(t-R/c)$ is the normal particle velocity at the front of the source, and $\bar{r}$ is the vector which defines the point ϕ at time t, as shown in Fig. 6.

The velocity potential $\phi(\bar{r},t)$ can be expressed as the convolution of the impulse response function $h(\bar{r},t)$ and the piston velocity $v(t)$ as follows:

$$\phi(\bar{r},t) = v(t) * h(\bar{r},t), \tag{2}$$

where the asterisk (*) denotes the convolution operation, and $h(\bar{r},t)$ is the velocity potential, or as it is alternately known the spatial impulse response.[10] The value at some point P which results from the velocity excitation of the source is defined as

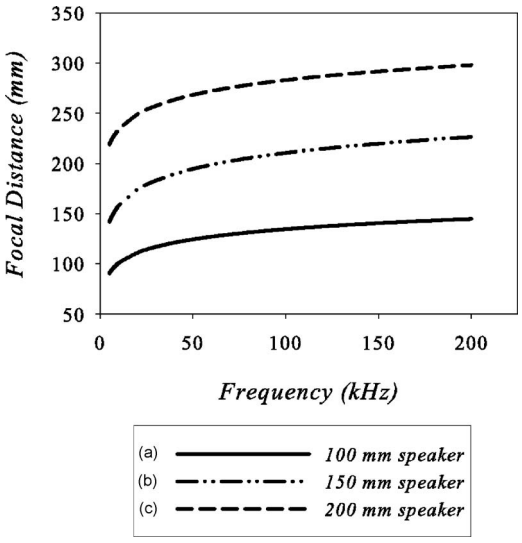


FIG. 2. Curves of focal distance as a function of frequency for a transducer that is (a) 100 mm, (b) 150 mm, and (c) 200 mm in diameter at its edge and whose focal distance are (a) 150 mm, (b) 225 mm, and (c) 300 mm, respectively.

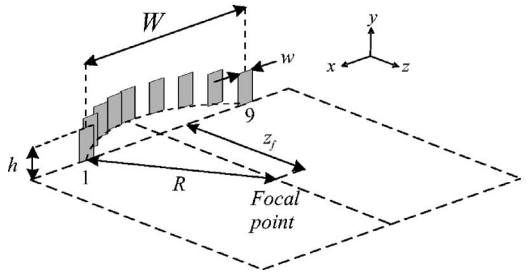


FIG. 3. Illustration showing an array of parallel elements, used to approximate the shape of a cylindrical focusing transducer.

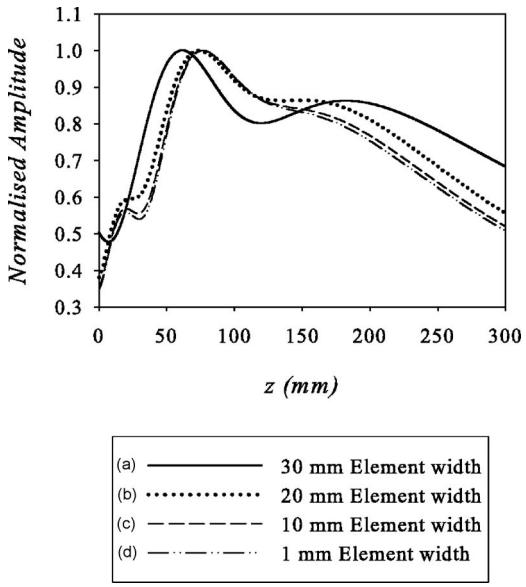


FIG. 4. Predicted field of a curved radiator in air at $f=10$ kHz, where $W=h=200$ mm and $z_f=150$ mm. The width of the elementary strips used in the simulation was (a) 30 mm, (b) 20 mm, (c) 10 mm, and (d) 1 mm.

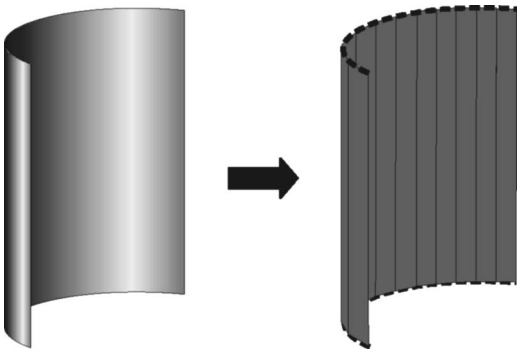


FIG. 5. Illustration of angled elementary strips to more accurately represent a curved radiating surface.

$$h(\bar{r},t)=\frac{1}{2\pi}\int_s \frac{\delta(t-R/c)}{R}dS. \tag{3}$$

The acoustic pressure $p(\bar{r},t)$ at the point P in front of the transducer is related to the velocity potential by

$$p(\bar{r},t)=\rho\frac{d\phi(\bar{r},t)}{dt}, \tag{4}$$

where ρ is the density of the propagating medium. The pressure can also be found in terms of the spatial impulse response function by combining Eqs. (2) and (4) to give

$$p(\bar{r},t)=\rho\frac{\partial[v(t)]}{\partial t}*h(\bar{r},t), \tag{5}$$

where $v(t)$ is the velocity of the transducer surface.

Figure 6 illustrates that only certain source locations will contribute to the scalar impulse response at a given time t, and these all lie on a circle centered at P', a projection of P onto the source, at a distance of $R=ct$ from the observation point. If the distance from P is R, then the radius of this circle is $l^2=R^2-z^2$. If the distance to point P is increased from R to $(R+\delta R)$, then the radius of this circle from the source to the observation point is $(l+\delta l)^2=(R+\delta R)^2-z^2$, and thus $l\delta l=R\delta R$.

The small area dS considered in Fig. 6 can also be expressed as $\Omega\cdot l\cdot\delta l$ or $\Omega\cdot R\cdot\delta R$, hence the surface integral of Eq. (3) can be expressed as an integral over the infinite range of R:

$$h(\bar{r},t)=\int_0^\infty \frac{\delta(t-R/c)\Omega R}{2\pi R}dR. \tag{6}$$

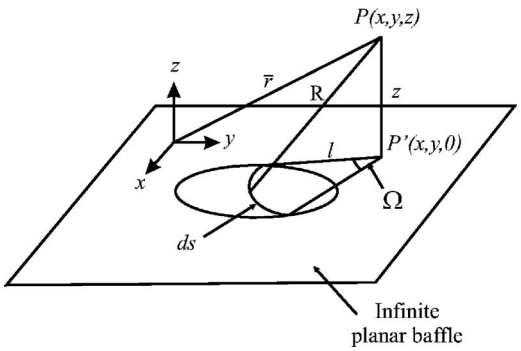


FIG. 6. Source geometry for determining the impulse response function.

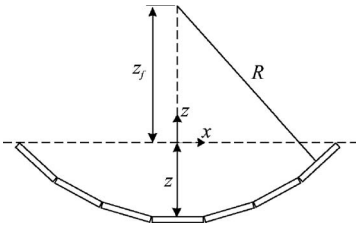


FIG. 7. An array of elements that have been rotated about their centers.

If $R=c\tau$ then,

$$h(\bar{r},t)=\int_0^\infty \frac{\delta(t-\tau)\Omega(\bar{r},t)c}{2\pi}d\tau. \tag{7}$$

By using the sifting function of the delta function,[3] Eq. (7) can be simplified to

$$h(\bar{r},t)=c\frac{\Omega(\bar{r},t)}{2\pi}, \tag{8}$$

where $\Omega(\bar{r},t)$ is the angle subtended by the piston in the plane of the source, from the point defined by $\bar{r}$ at a radius of $R=ct$. The scalar impulse response $h(\bar{r},t)$ therefore relates the radiated acoustic field to the geometry of the sound source.[2] The result is a time-limited function, where the limits are the longest and shortest path lengths from the observation point to the source.

III. IMPROVED GEOMETRICAL THEORY

To define the geometry of the transducer the array of parallel elements used to approximate the shape of a cylindrical focusing transducer, as shown earlier in Fig. 3, will be considered first. This will then be extended such that the elements are angled. The outermost elements of the array are arranged to lie in the $z=0$ plane. The focal point is at $(x,y,z)=(x_f,0,z_f)$. If the elements are arbitrarily numbered 1 to 9 from left to right, as shown in Fig. 3, then the location of the center of each element in the x direction is given by

$$x_{\text{center}}^n=\left(n-\frac{N+1}{2}\right)w, \tag{9}$$

where n is the element number, N is the total number of elements in the x direction, and w is the width of an individual element. If z_f in Fig. 3 is the distance from the focal point of the transducer to the x axis (i.e., at $z=0$), then the distance from the middle of any element to the axis can be calculated using

$$z=\sqrt{R^2-x_{\text{center}}^n}-z_f, \tag{10}$$

where R is the radial distance from the center of the element to the focal point of the array.

Now consider the case where the elements have been rotated about their centers, such that they more accurately define a cylindrically concave shape, see Fig. 7. In rotating the elements about their center point, it can be concluded that the distance from the center of the element to the x axis has not changed; thus the center of each element can still be found from Eq. (9). Each element is rotated by a different amount, enabling any size of transducer to be modeled. In

order to do so we must now define two more variables in the x direction, x_{low}, and x_{high}, which represent the nearest and farthest points on the element in the x axis (see Fig. 8), where it can be seen that

$$x_{\text{low}} = x^n_{\text{center}} - \left(\frac{w}{2}\right),$$
$$x_{\text{high}} = x^n_{\text{center}} + \left(\frac{w}{2}\right). \tag{11}$$

In a similar vein, two new variables in the z direction, namely z_{low} and z_{high}, which represent the similar positions of the element with respect to the z axis, must also be defined, as shown in Fig. 9. It is then possible to establish the distance in the z direction from the axis to both z_{low} and z_{high} as

$$z_{\text{low}} = \sqrt{R^2 - x^2_{\text{low}}} - z_f,$$
$$z_{\text{high}} = \sqrt{R^2 - x^2_{\text{high}}} - z_f. \tag{12}$$

Having established the distance from each end of the element to the axis, the distance from the observation point to the midpoint of the element can now be calculated:

$$z_{\text{new}} = \frac{z_{\text{high}} + z_{\text{low}}}{2} + z_f. \tag{13}$$

The angle, θ, by which the element has been rotated is equal to

$$\theta = \tan^{-1}\left[\frac{-z_{\text{low}} + z_{\text{high}}}{w}\right]. \tag{14}$$

Having rotated the element by θ, the element's length must change such that there are no gaps between adjacent elements. The new width of an element can be calculated as being equal to

$$w_{\text{new}} = \sqrt{(z_{\text{high}} - z_{\text{low}})^2 + w^2}. \tag{15}$$

As a further consequence of rotating the element about its midpoint the coordinate system must change, see Fig. 10, to take account of the fact that the element is no longer horizontal. Therefore from Fig. 10 it can be concluded that the new coordinate system is

$$x_{\text{angle}} = x^n_{\text{center}} \cos\theta + z_{\text{new}} \sin\theta,$$
$$z_{\text{angle}} = z_{\text{new}} \sin\theta - x^n_{\text{center}} \sin\theta. \tag{16}$$

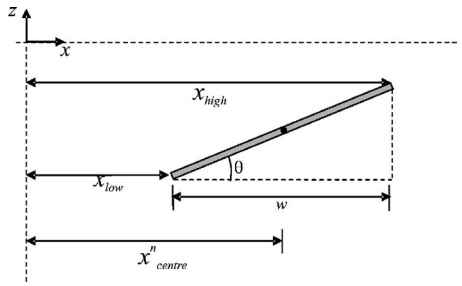


FIG. 8. Element variables used to define the position of each element in the x direction.

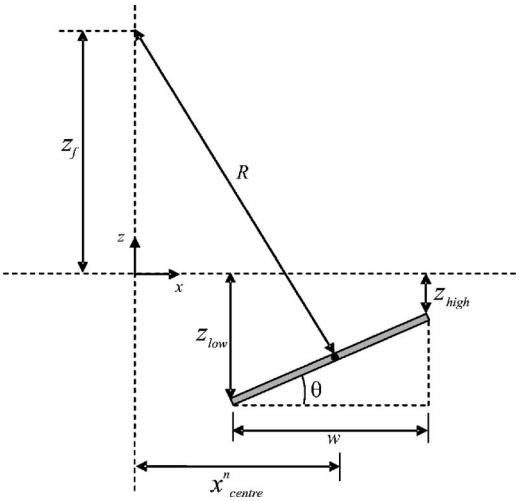


FIG. 9. Element variables used to define the position of each element in the z direction.

To establish the validity of this method, the theoretical sound pressure field for a 100 mm concave transducer has been calculated using the above-outlined theory outlined, and compared to the results obtained using the traditional parallel element approach[6,7] at $f=100$ kHz. To facilitate a direct comparison, the transducer was divided into 30 elements of width $w=3.33$ mm in each case. The results at $x=0$ mm (i.e., on axis) are shown in Fig. 11. In addition, Fig. 12 shows cross sections at an axial distance of $z=138$ mm, at the expected focal region of the transducer. As can be seen, the two approaches give similar results, with the size and location of both the focal point and near field response being analogous in both cases. However, the width w of the individual elements can be increased using the new approach, beyond the value of λ, without any detrimental effect to the results, decreasing the computational time as fewer elements are required. The results are shown in Fig. 13, with little difference in the on-axis response being evident. Using conventional parallel elements at low frequencies would require the use of a more restrictive element size, (i.e., $<\lambda$), in order to accurately resolve the radiated field, increasing the computational time required.

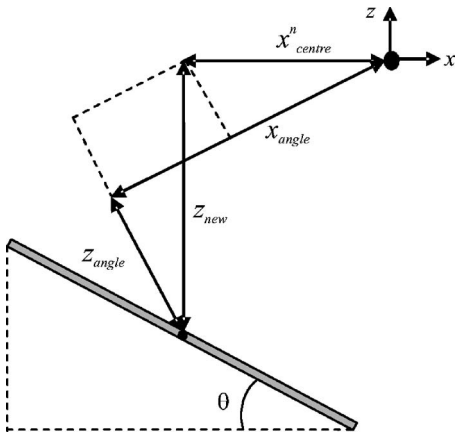


FIG. 10. Coordinate system employed by the new theoretical approach.

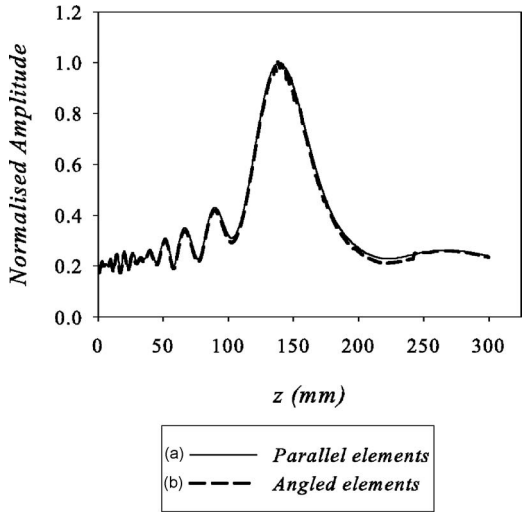


FIG. 11. Theoretical pressure field in air of concave rectangular source, assuming $W=h=100$ mm, $r_f=150$ mm, and $f=100$ kHz. Predictions are shown for both (a) parallel and (b) angled elements.

The number of elements chosen for field prediction will depend on the size of the radiator and the wavelength of operation. For a given source geometry the number of angled elements necessary for proper field prediction will have to be determined in each case. However, it has been determined that, for most conventional devices where the size of the radiator is not greater than approximately 30 wavelengths in diameter, the number of angled elements can be reduced to 10 with little degradation in prediction performance.

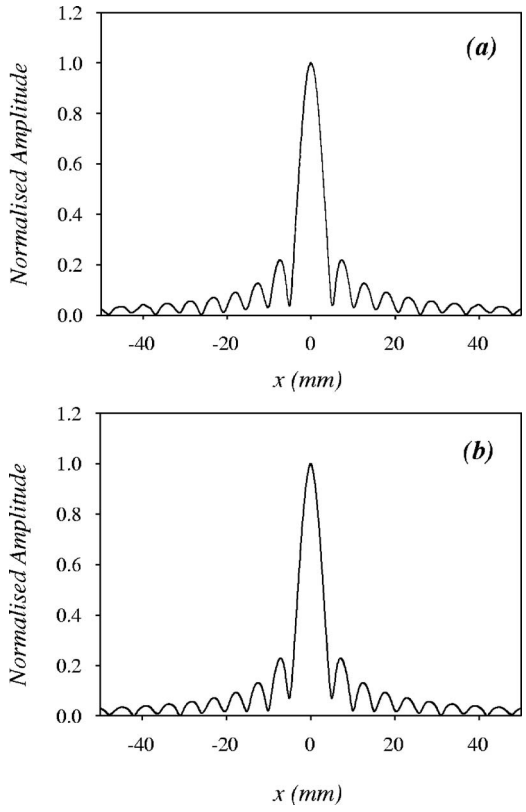


FIG. 12. Predicted cross sections at $z=138$ mm taken from the theoretical pressure fields shown in Fig. 9, using 30 elements of width $w=3.3$ mm. (a) Parallel and (b) angled elements.

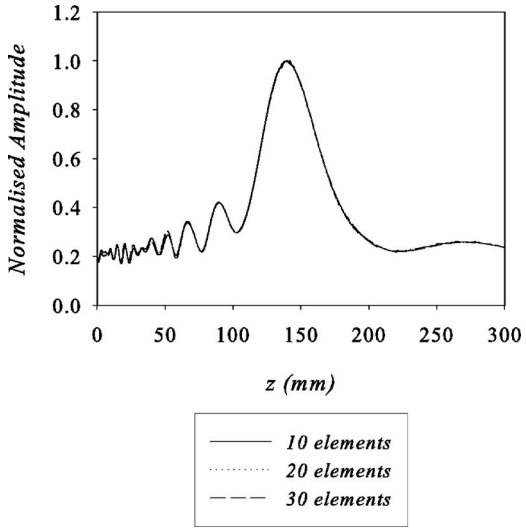


FIG. 13. Theoretical pressure field on-axis of a concave rectangular source in air using the new angled element approach, assuming $W=h=100$ mm, $r_f=150$ mm, and $f=100$ kHz. Predictions are shown as a function of the number of elements used across the face of the radiator.

To investigate the new method further, a source with $W=h=200$ mm was investigated for $R=150$ mm, at $f=10$ kHz in air. For the simulations, the transducer was split into a varying number of elements, the results of which are shown in Fig. 14. It can clearly be seen that there is very little difference in the predicted radiated field, with the size and position of the focal point remaining unchanged for the reduced number of elements. This is in direct contrast to the results obtained for the same loudspeaker using parallel elements, as shown earlier in Fig. 4.

It should be noted that only focusing transducers have been considered in this paper. However, the approach described here can also be easily extended to match any mathematically defined curvature (convex, sine, cosine, etc.). This can be achieved by modifications to Eq. (12), which defines the position along the z axis of the element to the axis. This process would then enable the radiated fields of alternative transducer shapes to be calculated.

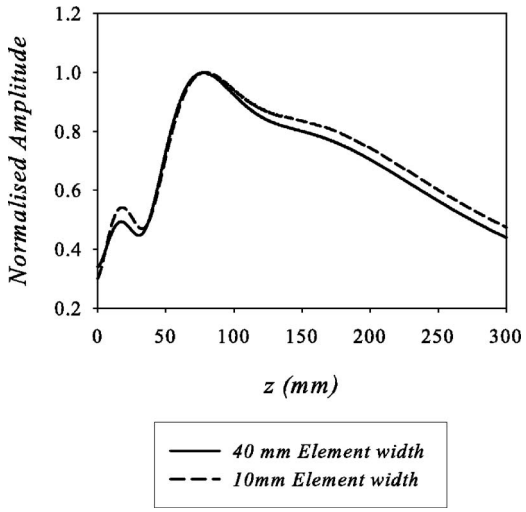


FIG. 14. Theoretical pressure field at 10 kHz in air for $W=h=200$ mm and $z_f=150$ mm, as a function of the width (w) of each element.

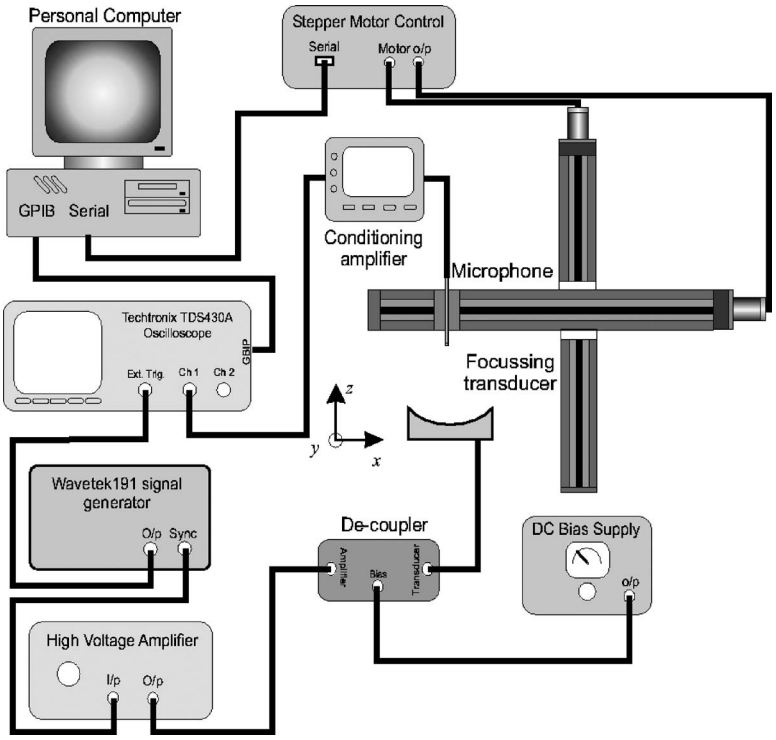


FIG. 15. Schematic diagram of the system used to measure the radiated sound pressure fields using a scanned microphone.

IV. COMPARISON OF THEORY AND EXPERIMENT

The above-mentioned predictions illustrate that the new geometrical formulation produces an accurate and reliable representation of the radiated fields of focusing transducers, which is more computationally efficient, greatly reducing the time needed to resolve the radiated pressure field. Therefore, having established the theoretical validity of this method, the pressure field from the transducer was now compared to experimental results. For the theoretical simulations, ten elements were used with $w=20$ mm. In each case, a simulated tone burst signal, as used by San Emeterio and Ullate,[4] was used to drive the transducer. Such a signal can be described as

$$v(t) = t^3 e^{-Kft} \cos(2\pi ft), \tag{17}$$

where f is the center frequency of oscillation and K is a constant that defines the bandwidth of the resultant wave form. A value of $K=0.4$ was used for both experimental and theoretical studies.

The experimental pressure field measured was that of a 200 mm square concave electrostatic loudspeaker, with $R=300$ mm. The construction and performance of these devices has been described elsewhere,[11] but they were chosen here as they allowed a variety of different shapes to be investigated. The radiated field was measured using a Bruel and Kjaer calibrated 1/8 in. microphone, scanned throughout the radiated field using a PC-controlled scanning system in the $x-z$ plane. The experimental arrangement is shown in Fig. 15, with a photograph of the electrostatic loudspeaker and the measurement microphone being shown in Fig. 16. For each scan, an area measuring 204 mm by 648 mm in the x and z direction, respectively, was examined using a spatial sampling of 4 mm in both the x and z directions, resulting in a scan 51 by 162 points in size. All scans were initiated at $z=15$ mm (± 0.5 mm) from the center of the transducer.

To enable a direct comparison between both theory and experiment, the resulting images have been produced as linear greyscale plots, as shown in Figs. 17 and 18 for driving frequencies of 10 and 50 kHz, respectively. Comparison of these results shows relatively good agreement, with similar interference fringes caused by the constructive and destructive interference of the tone-burst signal being evident on either side of the main peak. The plots also show that, as expected, the spatial extent of the focal region is a function of frequency, with the width of the focal point decreasing as the frequency is increased. It is also clear that the position of the focal point moves toward the geometric center of the transducer as the frequency of excitation increases.

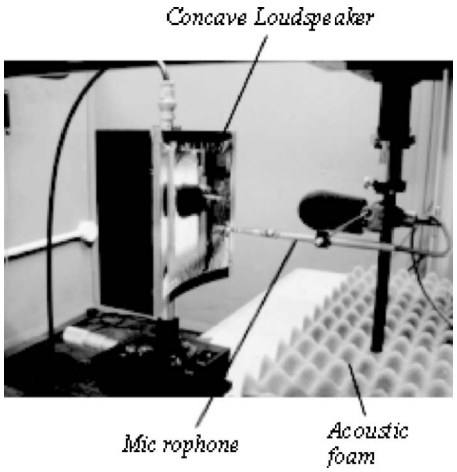


FIG. 16. Photograph of the concave loudspeaker and the 1/8 in. microphone.

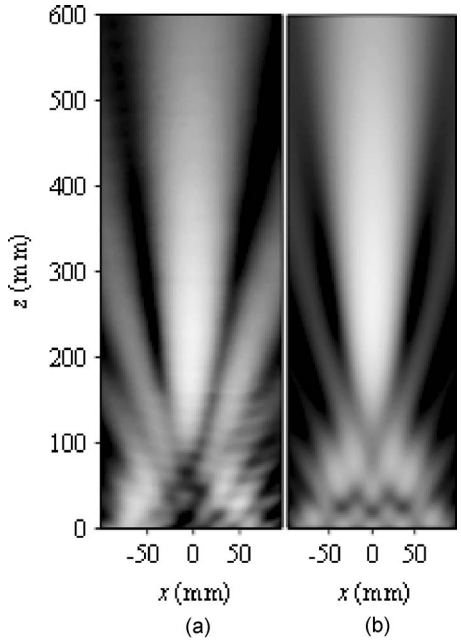


FIG. 17. Radiated pressure field in air at 10 kHz for $W=h=200$ mm and $z_f=300$ mm. (a) Experimental result and (b) theory.

V. DISCUSSION AND CONCLUSIONS

The prediction of radiated fields from acoustic sources is important in the characterization of acoustical systems. Recent developments in loudspeaker technology allow different profiles to be constructed, such that the radiated field can either be convergent (i.e., focused) or divergent. As a consequence of this, the radiated fields from a curved radiator have been studied over a range of frequencies, both theoretically and experimentally. As expected, the size and location of the focal region is a function of frequency, at an axial distance that is not equal to the radius of curvature of the device.

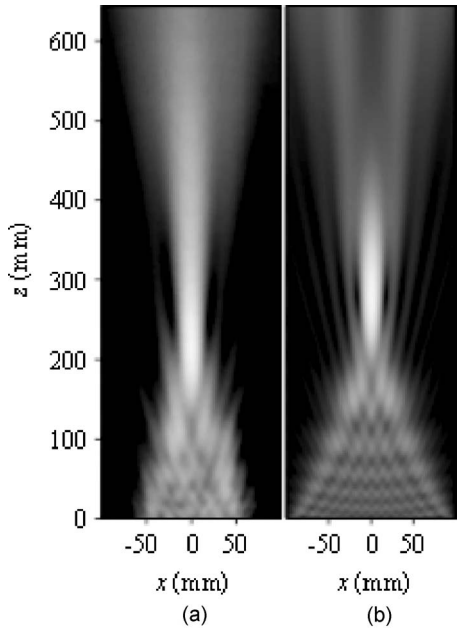


FIG. 18. Radiated pressure field in air at 30 kHz for $W=h=200$ mm and $z_f=300$ mm. (a) Experimental result and (b) theory.

It is possible to calculate the theoretical pressure field from a focusing transducer with a conventional approach, whereby the radiating surface is split into a series of horizontal elements each with an appropriate offset from the axis. It has been shown that this approach can accurately predict the radiated field of focusing transducers for certain geometries and frequency ranges. However, care must be taken to ensure that the elements are sufficiently small that errors are not introduced into the predicted sound pressure fields. Because of this, a new geometrical formulation was developed, in which the elements are angled such that they more accurately represent the overall shape of the actual transducer they are trying to replicate. Such an approach has a number of advantages, in that there is no longer a necessity to ensure that the width of the individual elements used in the calculation should be less than the wavelength of the driving wave form.

It has been determined that the number of angled elements in the calculation can be reduced to only 10, provided the ratio of source diameter to wavelength is less than 30. Typically, a calculation using the conventional approach using parallel elements would take approximately 2 h on a conventional PC. The resultant reduction in computation time occurs in all cases, and is a function of the reduction in number of elements possible in the new approach, which varies depending upon transducer size and wavelength. For a typical case, the computation time is reduced by 30%–50%. Note also that the method can also be easily modified to enable the radiated fields of different shaped transducers to be calculated.

This new approach has been validated by comparison to both the conventional simulation approach and to experiment. This has shown that the new approach is able to accurately predict the radiated field from concave loudspeakers at a range of frequencies. It is thus suggested that this method could have wide application in the area of radiated field prediction.

[1] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945).
[2] G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[3] P. R. Stepanishen, "Transient radiation from pistons in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1629–1638 (1971).
[4] J. L. San Emeterio and L. G. Ullate, "Diffraction impulse-response of rectangular transducers," J. Acoust. Soc. Am. **92**, 651–662 (1992).
[5] 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**, 735–741 (1973).
[6] R. Reibold and R. Kažys, "Radiation of a rectangular strip-like focussing transducer. 1. Harmonic excitation," Ultrasonics **30**, 49–55 (1992).
[7] R. Reibold and R. Kažys, "Radiation of a rectangular strip-like focussing transducer. 2. Transient excitation," Ultrasonics **30**, 56–59 (1992).
[8] D. A. Hutchins, H. D. Mair, and R. G. Taylor, "Transient pressure fields of PVDF transducers," J. Acoust. Soc. Am. **82**, 183–192 (1986).
[9] H. D. Mair and D. A. Hutchins, "Axial focusing by phased annuli," Proceedings of the International Conference of Underwater Acoustics, Halifax, Canada, 1987.
[10] A. J. Rutgers, "Spatial impulse response of an acoustic line radiator—A study of boundary diffraction-wave phenomena and their experimental detection," Ph.D dissertation, School of Engineering and Architecture, Catholic University of America, Washington, DC, 1977.
[11] A. P. Medley, D. A. Hutchins, and D. R. Billson, "Radiated fields of thin and flexible high frequency loudspeakers," Proceedings of the 149th Meeting of the Acoustical Society of America, Vancouver, Canada, 16–20, May 2005.

Experimental validation of band gaps and localization in a one-dimensional diatomic phononic crystal

Anne-Christine Hladky-Hennion
Institut d'Electronique, de Microélectronique et de Nanotechnologie, UMR 8520 CNRS, Department ISEN, 41 Boulevard Vauban, 59046 Lille Cedex, France

Michel de Billy[a)]
Institut Jean Le-Rond d'Alembert, UMR 7190 CNRS, Université Paris 6, 2, Place de la Gare de Ceinture, 78210 Saint Cyr l'Ecole, France

(Received 18 January 2007; revised 11 May 2007; accepted 6 August 2007)

The propagation of compressional ultrasonic pulses through a finite one-dimensional chain of various unit cells is investigated experimentally. The chain, initially compressed by an axially applied constant force, is excited by a periodic force, which acts in line with axis of bead chain. The experimental measurements giving the eigenfrequencies of the specimen are based on a Fourier analysis of the transmitted acoustic pulse. The results are compared with the numerical calculations and it is shown that the two approaches are well correlated. A phononic band structure is observed and under certain conditions, depending on the parity of the number and on the masses of the beads in the chain, it is shown that localized modes propagating in the forbidden band are exhibited. Much attention is devoted to the existence of these localized modes according to the mass ratio between two adjacent beads constituting the unit cell. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2779130]

PACS number(s): 43.40.At, 43.20.Ks, 43.20.Mv [RLW] Pages: 2594–2600

I. INTRODUCTION

In the recent past, many studies have been devoted to two- and three-dimensional periodic phononic crystals.[1–9] The main result of these investigations is proof of the existence of stop frequency bands in which the propagation of sound is forbidden and for which the mechanical properties of the composite system constituents determine their existence and width. The effects of boundaries, viscous damping, and imperfections on two-dimensional waveguides were investigated by Jensen[10] and predict the existence of so-called phononic band gaps (i.e., frequency intervals in which no propagating modes exist) for finite and infinite lattices; it is also shown that localized modes may be excited and associated with local defects from periodicity or with the ends of the lattice. During the last few years, some papers considered the propagation of sound through one-dimensional (1D) periodic waveguide systems such as regular and irregular string mass chains[10] and periodic arrays.[8–20]

In the present paper, experimental results on the vibrational response of finite and periodic structures are reported. They confirm the theoretical works published in Ref. 21 in which the case of a 1D diatomic chain of beads (the chain contains alternately two beads with different masses) with different acoustical properties is considered. Numerical calculations have shown that under certain conditions bearing on the mass ratio between adjacent beads in the diatomic unit cell, and on the parity of the number of the beads in the chain, it is possible to observe a band in which two localized modes propagate. This behavior is similar to the one observed in photonic crystals.[8,9]

Our main objective is to study the eigenstates of a 1D elastic chain submitted to a short compressional acoustic excitation. The periodicity is only over a finite and small number of unit cells. The unit cell is the elementary period of the lattice: It contains the two beads of different masses. First, the perfect system consisting of identical beads (same diameter and material) is examined. As is well known, no gap exists in the band structure for such an homogeneous chain. Then, we have investigated 1D waveguides described in terms of one "unit cell" consisting of two beads with different diameters and materials. We confirm that the periodicity and mechanical properties of the beads determine the position and the width of the band gaps as well as the existence of "localized modes" in the stop bands.[18,21]

II. BRIEF SUMMARY OF THEORETICAL RESULTS

In this section we recall the main theoretical results obtained in the case of chains containing identical or different beads.[21,22]

First, considering an infinite chain containing identical beads, the dispersion curves can be drawn and exhibit several branches, the first one being the acoustical branch. Considering now a finite chain containing N identical beads, N modes are obtained for each branch. In the case of a chain containing eight identical steel beads, excited by a longitudinal force applied at one extremity, Fig. 1 presents the variations of the displacement at the other extremity, in the direction of the chain, as a function of the frequency. The curve is generic, thus the scale is omitted. The frequency range of

[a)]Electronic mail: mdebilly@ccr.jussieu.fr

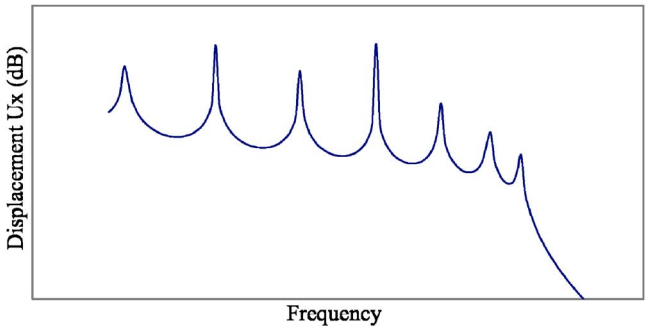


FIG. 1. (Color online) Theoretical curve showing the displacement U_x as a function of the frequency for a chain containing eight identical steel beads. The vertical scales are in decibels.

Fig. 1 corresponds to the first acoustical branch. These numerical results are obtained using the finite element method, with the help of the ATILA code.[23] It clearly shows seven peaks. The resonance frequency at zero always gives a peak but it is not reproduced in Fig. 1. At low frequency, the peaks are well separated. They are closely spaced when the frequency is increasing. It is possible to predict the position of the peaks with the help of a specific quantification.[22]

If we consider an *infinite* chain composed of N beads of types 1 and 2 alternately, there is a folding of the first branch, leading to the acoustical and optical branches, separated by a gap, the width of which depends on the mass ratio and on the coupling between the beads.[24,25] Considering a *finite* chain composed of N beads of types 1 and 2 alternately, two cases are distinguished, according to the parity of the number of the beads in the chain.

A. Even number of beads in the chain

In this case the chain is made up of beads the masses of which are m and M, designating the lighter and the heavier beads, respectively. A simple model of diatomic chain gives the vibration modes of the chain. The resonance frequency equal to zero ($f_0=0$) is always a solution to the system and corresponds to a global translation of the N beads. The resonance frequency,

$$f_{\mathrm{loc}} = [C(1/M + 1/m)]^{1/2}, \tag{1}$$

is always a solution too; C designates the coupling constant. This solution is located between the acoustical and optical branches and the calculation, using a simple atomic model or the finite element method,[21] shows that the amplitude of this vibration mode decreases from one cell to the next. This mode does not occur in the middle of the gap but is always located at the same position whatever the value of N and helps us to estimate the value of the coupling constant from Eq. (1).

B. Odd number of beads in the chain

The numerical study (Ref. 21) predicts the number of modes in the stop band as a function of α, which represents the ratio between the heavier (M) and the lighter (m) mass; α is always greater than 1.

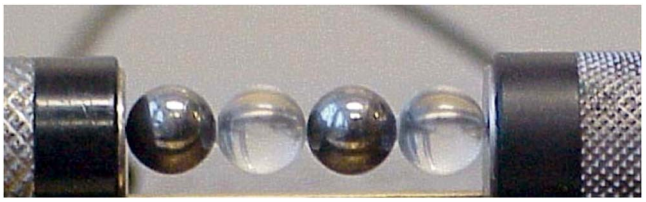


FIG. 2. (Color online) Picture of a sample consisting of two kinds of glued beads alternately distributed (aluminum and glass).

(1) If heavier beads are at both extremities, then no vibration mode is observed in the forbidden band.
(2) If lighter beads are at both extremities, two localized modes can be excited in the stop band as soon as the double inequality is verified by p (integer number) with $p=(N-1)/2$:

$$(p+1)/p < \alpha < p/(p-1). \tag{2}$$

Localized modes appear in the stop band only if the beads which are located at the extremities of the chain are the lighter. Equation (2) shows that, if the two types of beads are close ($\alpha \# 1$), then the localized modes will appear only if the chain consists of a large number of cells. Conversely, if the values of the mass are very different ($\alpha \gg 1$), the two localized modes exist even for very small values of N. Finally, when N is increasing, the difference between the two resonance frequencies of the localized modes is decreasing.

Another way for showing the results is to plot—for a given value of the integer N—the resonance frequencies. It allows one to point out two parts: the lower frequencies (acoustical branch) and the higher frequencies (optical branch).

The objective of this work consists in determining the positions of the peaks in the recorded power spectra obtained with various finite chains of different integer values of N and in showing if modes appear in the band gap. Four examples have been investigated for different values of α.

III. EXPERIMENTAL CONDITIONS

A. Experimental device

The experimental setup is a classical one used for acoustic transmission measurements: The specimen is clamped between two transmitters (type Panametrics V103) and a constant strength (35N) is applied along their common axis. The chain only supports the propagation of compressional acoustic waves generated at one of the extremities of the chain by a broadband contact transducer driven with short pulses. The scope averages 200 transmitted signals to improve the signal to noise ratio. The short signal generated by the incident vibration is transmitted through the system and recorded on a digital oscilloscope.

B. Description of the specimen

The samples consist of a series of glued beads as a function of length (Fig. 2). It is assumed that all the beads are perfectly aligned. The beads are glued with bicomponent epoxy. The coupling between two consecutive balls is roughly

TABLE I. Acoustic and mechanical properties of the materials used for calculations.

Material	Density ρ (kg/m^3)	C_L (m/s)	C_T (m/s)	Diameter (mm)
Glass	2600	5960	3760	8.0±0.1
Aluminum	2700	6420	3040	8.0±0.1
Steel	7800	5960	3235	10.0±0.1
				8.0±0.1
Bronze	8600	4582	2110	11.1+0.1

the same, thus we measure an averaged value of the coupling constant over all the contacts in the chain. This averaged value of the constant coupling ($\langle C \rangle$) is computed from Eq. (1) in which we replace f_{loc} by the average value of the frequency of the localized mode ($f_{\text{loc,exp}}$) identified in the experimental spectra obtained with chains containing an even number of spherical beads.

Four examples of specimens are considered in this paper according to four different combinations of two types of beads of different materials and diameters. They correspond to different values of α:

(1) Aluminum and glass beads (both 8 mm in diameter), $\alpha = 1.04$,
(2) Bronze (11.1 mm in diameter) and steel beads (10 mm in diameter), $\alpha = 1.47$,
(3) Steel beads (8 and 10 mm in diameter), $\alpha = 1.95$,
(4) Glass and steel beads (both 8 mm in diameter), $\alpha = 3.00$.

The acoustical and mechanical properties of the materials are reported in Table I.

C. Experimental procedure

A fast Fourier transform of the truncated recorded temporal signal is used to obtain the power spectrum of the transmitted signal. The spectra presents several peaks, revealing information about the eigenfrequencies. In the experimental power the displayed data are in decibels and the scale is linear.

IV. EXPERIMENTAL RESULTS

A. Case of even values of *N*

Figure 3(a) presents the experimental power spectrum for a finite chain made up of 12 steel beads of 10 and 8 mm in diameter alternately. We can clearly distinguish three sections: The first one, below 42 kHz, describes the acoustical branch; the third one, above 58 kHz, describes the optical branch. Between these two sections a localized mode is generated at 51.1 kHz. This type of frequency response will be observed for any sample including an even integer of beads.

In Fig. 3(b) the calculated displacement U_x of the last bead along the direction of the propagation is plotted. This harmonic analysis is performed by applying a force on one extremity of the sample and calculating the displacement at the other extremity. It reveals the existence of resonance peaks showing the acoustic and optical branches and a localized mode. In the experimental spectrum the acoustic modes are well separated and their positions comply with the peaks observed in the plot of the numerical displacement. Conversely the optical modes are not very well identified for large values of *N*. We observed that the maxima amplitudes of the optical modes in the power spectra are much smaller than the maximum amplitude of the acoustical modes. Then, it is difficult to identify the peaks associated with the optical range. In addition, at higher frequencies (case of optical branch) the modes become denser and it becomes difficult in separating them, the frequency resolution of the apparatus being ±0.4 kHz. For these reasons, the optical modes were not well identified or at least a great incertitude may affect the position of these modes. However it is conclusive that the experiment is in reasonably good agreement with the theory in that simple case and, consequently, measurements have been achieved on finite chains showing different unit cells and various values of N (2, 4, 6, 8, and 10 or 12 beads).

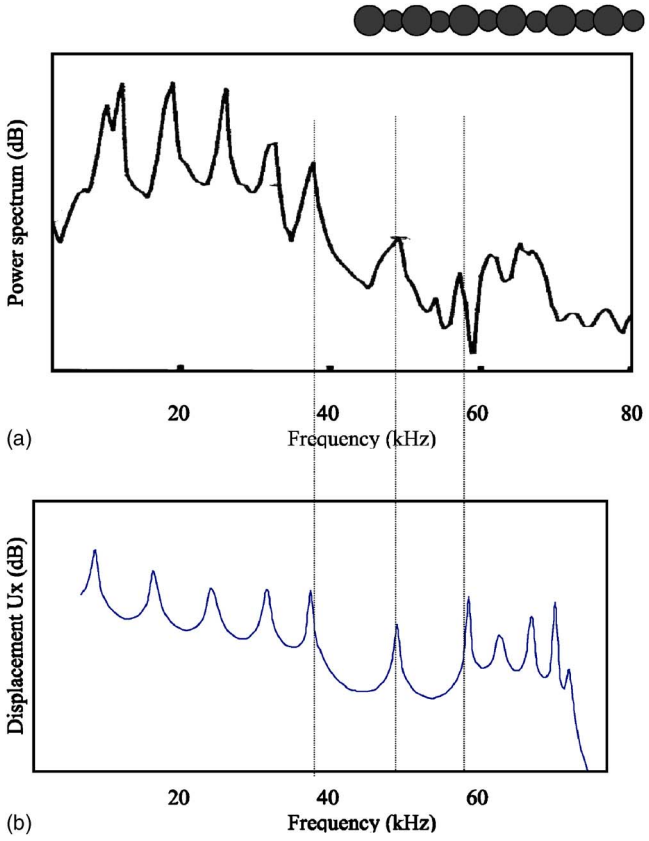


FIG. 3. (Color online) $N=12$. $\text{Steel}_{10}/\text{steel}_8$ ($\alpha=1.95$): (a) Experimental power spectrum. (b) Theoretical curve showing the displacement U_x as a function of the frequency. The vertical scales are in decibels. The vertical lines represent the correspondence between experimental and numerical peaks.

The experimental resonance frequencies are reproduced in Figs. 4(a)–4(d) for different values of α. Again we verify that the acoustical modes are well identified, whereas the distribution of the optical modes does not show the predicted distribution. Figures 4(a)–4(d) show the existence of the localized mode in the forbidden band. Its position does not change with the length of the chain. The frequency of the localized mode is calculated by averaging the frequency value $\langle f_{\text{loc,exp}} \rangle$ at which the localized mode is observed over

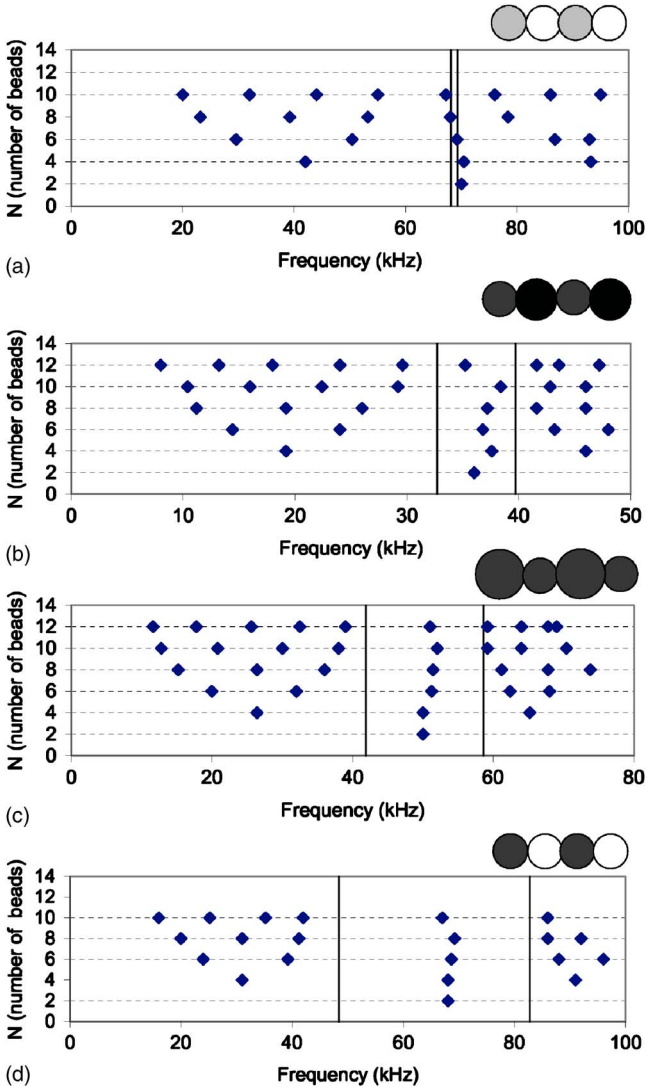


FIG. 4. (Color online) Experimental resonance frequencies of the samples, for chains containing an even number of beads. (a) $Alu_8/glass_8$ (α=1.04), (b) $bronze_{11}/steel_{10}$ (α=1.47), (c) $steel_{10}/steel_8$ (α=1.95), and (d) $steel_8/glass_8$ (α=3.00).

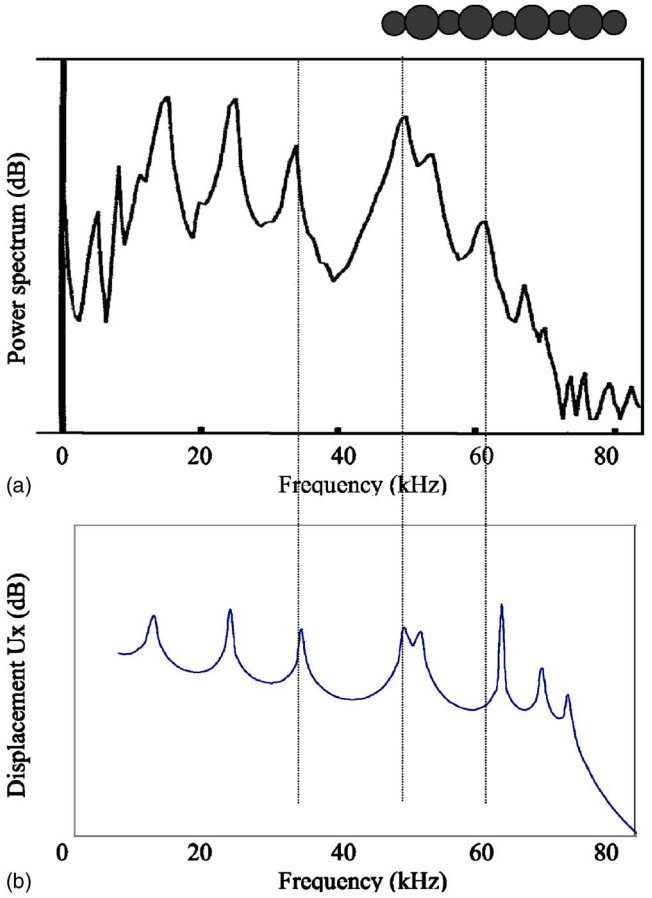


FIG. 5. (Color online) N=9, $steel_{10}/steel_8$ (α=1.95), lighter beads at both extremities. (a) Experimental power spectrum. (b) Theoretical curve showing the displacement U_x as a function of the frequency. The vertical scales are in decibels. The vertical lines represent the correspondence between experimental and numerical peaks.

the available measurements. Then, it is used to evaluate the average value of the coupling constant $\langle C \rangle$. These constant coupling values are then used for the evaluation of the limits of the passbands.[24] The results are presented in Table II for each case. One can notice that the identification of the localized mode for the chain made of steel and bronze beads [Fig. 4(b)] seems to be less consistent. In fact, the frequency scale is different from the other plots, which increases the dispersion of the modes. If we designate by Df the frequency dispersion affecting the localized mode, we get 3.2 kHz for the bronze/steel and aluminum/glass samples, and 2.2 kHz for the $steel_{10}/steel_8$ and glass/steel samples.

B. Case of odd values of *N*

For α=1.95, Fig. 5 presents the experimental and numerical spectra of a chain containing nine beads, when the two small beads are situated at both extremities. It clearly shows two modes in the band gap, as predicted by the theory. Figure 6 presents the experimental and numerical spectra of a chain containing nine beads, when the two large beads are situated at both extremities. It clearly shows no modes in the

TABLE II. For the different samples, mass ratio, frequency of the localized mode in the gap f_{loc}, coupling coefficient deduced from Eq. (1), and width of the band gap, deduced from Ref. 24.

Sample	$\alpha=M/m$	Frequency of the localized mode in the gap (kHz)	Coupling coefficient C (10^6 N/m)	Gap (kHz)
$Alu_8/glass_8$	1.04	68.9±1.6	1.67±0.04	68.1–69.3
$Bronze_{11}/steel_{10}$	1.47	36.8±1.6	3.32±0.08	32.7–39.7
$Steel_{10}/steel_8$	1.95	51.1±1.1	3.60±0.09	41.9–58.6
$Steel_8/glass_8$	3.00	68.1±1.1	2.40±0.07	48.4–82.8

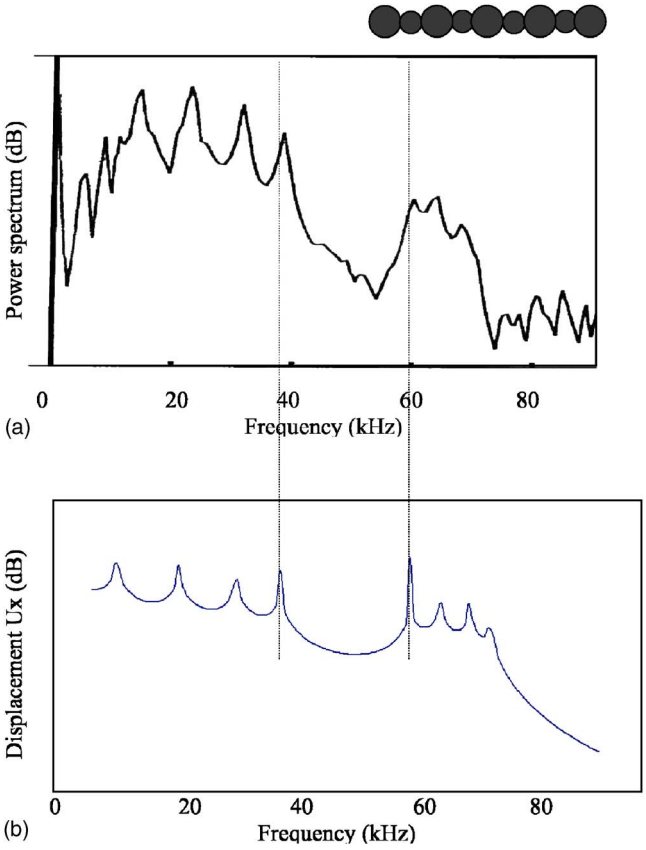


FIG. 6. (Color online) $N=9$, $\text{Steel}_{10}/\text{steel}_8$ ($\alpha=1.95$), larger beads at both extremities. (a) Experimental power spectrum. (b) Theoretical curve showing the displacement U_x as a function of the frequency. The vertical scales are in decibels. The vertical lines represent the correspondence between experimental and numerical peaks.

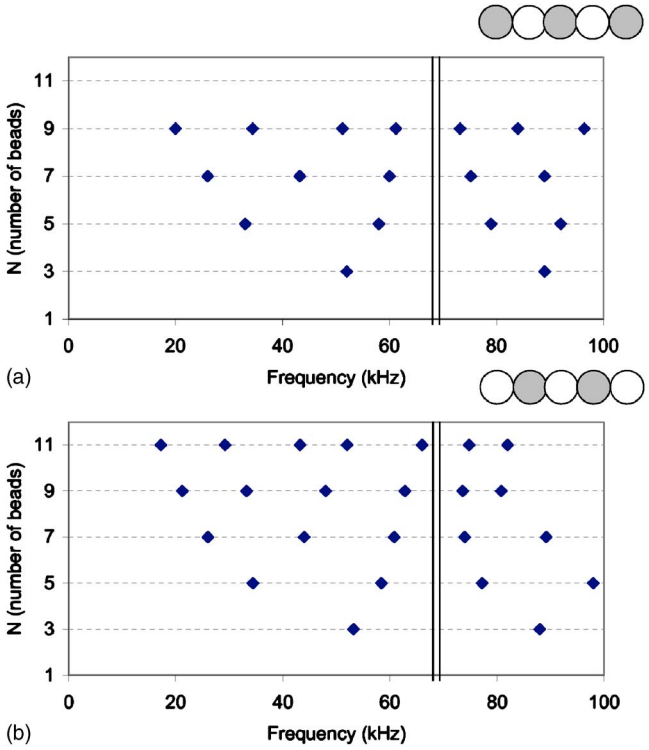


FIG. 7. (Color online) Experimental resonance frequencies of the samples, for chains containing an odd number of beads. $\text{Alu}_8/\text{glass}_8$ ($\alpha=1.04$): (a) Aluminum beads at both extremities (b) Glass beads at both extremities. The vertical lines represent the band gap limits.

band gap, as predicted by the theory. In these two cases, the experimental power spectra of the transmitted signals [Figs. 5(a) and 6(a)] are very similar to the variations of the numerical displacement obtained by the harmonic analysis [Figs. 5(b) and 6(b)].

Then the four cases, previously described in Sec. II, are experimentally tested and allows us to verify the law bearing on the N-value dependence of number and positions of the localized modes according to the relationships (2).

1. Case of $\text{Alu}_8/\text{glass}_8$ ($\alpha=1.04$)

In this case, as the masses are very close, the frequencies delimiting the stop band are very close, too (68.1–69.3 kHz). In Fig. 7(a) the experimental positions of the resonance obtained for a chain made up of nine beads are plotted, when the heavier bead is at both extremities. The vertical lines represent the calculated limits of the stop band, deduced from experiments (Table II). The stop band is very narrow, no modes are observed in it, and a very small gap separates the optical and the acoustical modes. Notice that these last modes are well identified.

When the lighter bead is at both extremities [Fig. 7(b)], no localized modes are observed in the forbidden band for the considered chains ($N=3-11$), because of the very small difference between the mass of the aluminum bead and of the glass bead (α close to 1). Localized modes could appear only if the chain contains a large number of cells, as predicted by Eq. (2).

2. Case of $\text{bronze}_{11}/\text{steel}_{10}$ ($\alpha=1.47$)

Figure 8(a) gives the experimental resonance frequencies for different values of N up to 11 when bronze beads are situated at both extremities of the chain. In that case no mode is observed in the forbidden band.

When the lighter bead is at both extremities [Fig. 8(b)] the calculations predict the presence of two localized modes if the number of beads in the chain is higher or equal to 7, according to Eq. (2). The experimental data reported in Fig. 8(b) confirm this value and the existence of two localized modes if N is greater than or equal to 7.

3. Case of $\text{steel}_{10}/\text{steel}_8$ ($\alpha=1.95$)

Similar curves are drawn in Fig. 9(a) for a chain containing steel beads of 8 mm in diameter and 10 mm in diameter, alternately. When the bigger beads are located at both extremities, no mode is observed in the forbidden band. The acoustical branches are well identified whereas the optical branches are not as easily detected in such a finite elastic specimen.

When the lighter bead is at both extremities [Fig. 9(b)], the calculations predict the presence of two localized modes if the number of beads in the chain is higher or equal to 3, according to Eq. (2). This prediction is well confirmed in

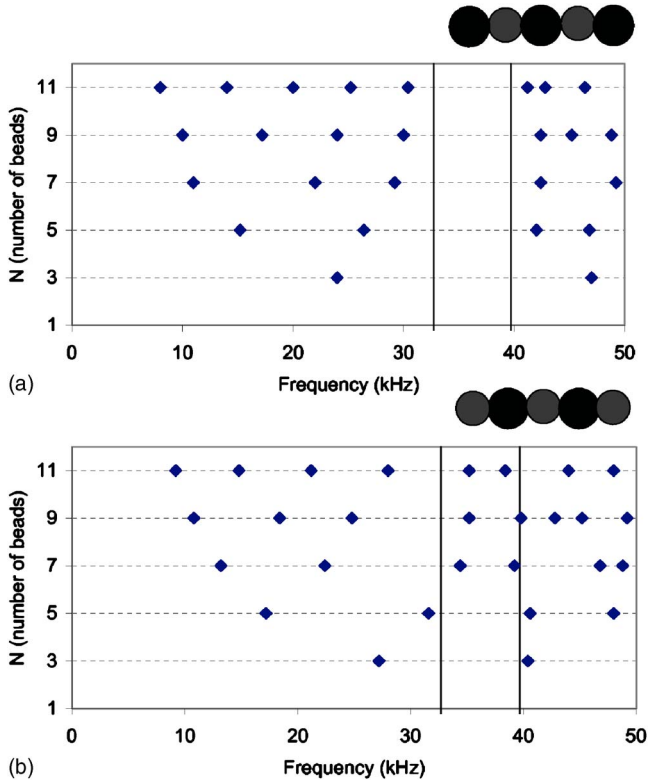


FIG. 8. (Color online) Experimental resonance frequencies of the samples, for chains containing an odd number of beads. $\text{Bronze}_{11}/\text{steel}_{10}$ (α=1.47): (a) Larger beads (bronze) at both extremities and (b) smaller beads (steel) at both extremities. The vertical lines represent the band gap limits.

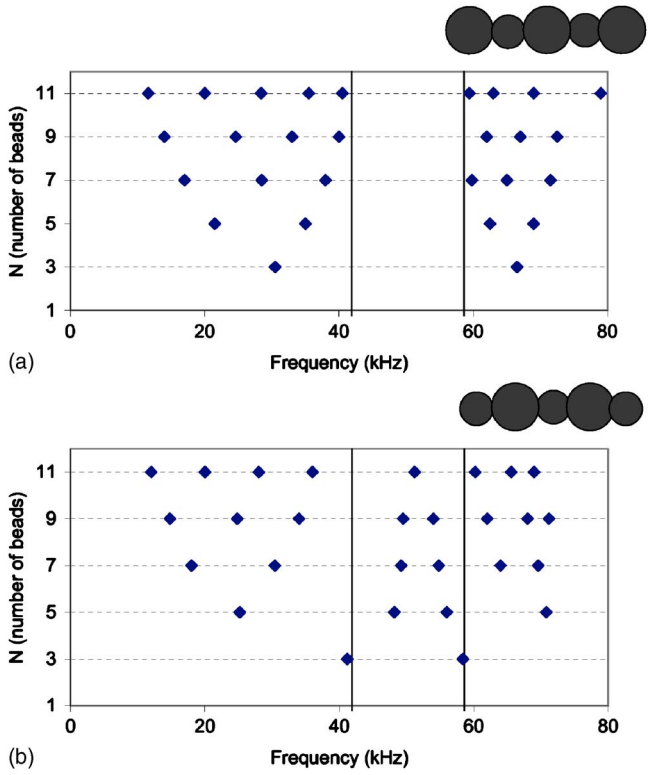


FIG. 9. (Color online) Experimental resonance frequencies of the samples, for chains containing an odd number of beads. $\text{Steel}_{10}/\text{steel}_{8}$ (α=1.95): (a) Larger beads at both extremities and (b) smaller beads at both extremities. The vertical lines represent the band gap limits.

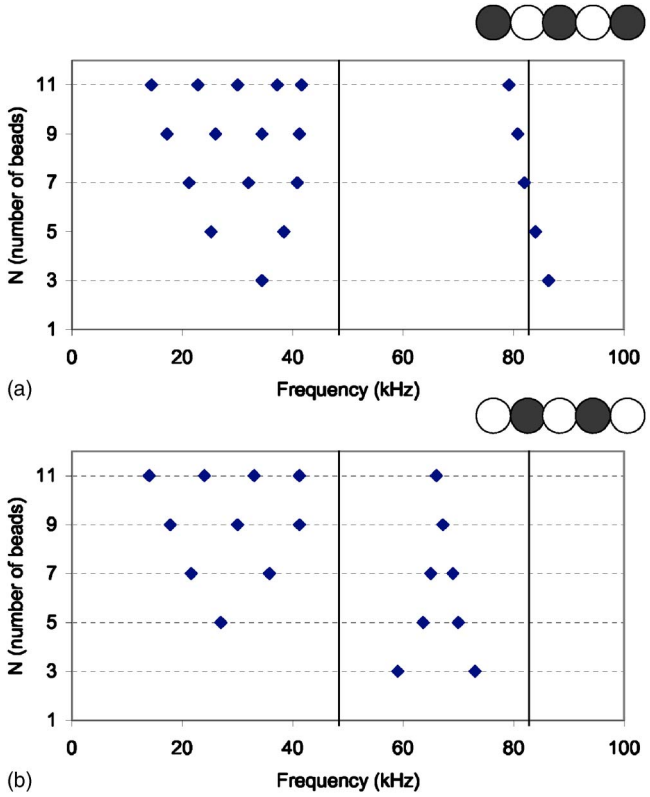


FIG. 10. (Color online) Experimental resonance frequencies of the samples, for chains containing an odd number of beads. $\text{Steel}_{8}/\text{glass}_{8}$ (α=3.00): (a) Steel beads at both extremities and (b) glass beads at both extremities. The vertical lines represent the band gap limits.

Fig. 9(b). For $N=11$, the two modes in the band gap are not separated because of the frequency resolution of the apparatus.

4. Case of $\text{steel}_8/\text{glass}_8$ (α=3.00)

Figure 10(a) gives the experimental resonance frequencies when steel beads are situated at both extremities of the chain. For higher values of N, the modes are unexpectedly within the forbidden region. In that case, the limits of the band gap are estimated using the localized mode of Fig. 4(d), which does not present a rectilinear line. Mistakes on its value imply mistakes in the calculations of the band gap limits. Moreover, only one mode is observed in the optical regime because the amplitude of the modes were so small that they were not significantly extracted from the noise in the spectra. When glass beads are at both extremities [Fig. 10(b)], we observe very clearly the existence of two localized modes as soon as N equals the lowest possible value ($N=3$), as predicted by the theory. For $N=9$ and $N=11$, the two modes in the band gap are not separated because of apparatus resolution.

V. CONCLUSION

In this paper we have examined the propagation of ultrasonic pulses through a 1D periodic waveguide system. The finite specimen consisted of a chain made up of glued beads. The physics is similar to that of electrons moving in a periodic potential. It is observed that allowed frequency

bands are separated by forbidden gaps. Under certain conditions bearing on the parity of the number of the beads and on the ratio of the two masses which constitute the unit cell, it was experimentally verified that two localized modes may be excited in the band gap. The experimental results comply with the calculations and demonstrate that this kind of limited specimen let us explore fundamental effects on the phononic band structure and can be used efficiently in filtering applications.

[1]Y. Pennec, B. Djafari-Rouhani, J. O. Vasseur, A. Khelif, and P. A. Deymier, "Tunable filtering and demultiplexing in phononic crystals with hollow cylinders," Phys. Rev. E **69**, 046608 (2004).

[2]S. Yang, J. H. Page, Z. Liu, M. L. Cowan, C. T. Chan, and P. Sheng, Ultrasound tunneling through 3D phononic crystals, Phys. Rev. Lett. **88**, 104301 (2002).

[3]J. O. Vasseur, P. A. Deymier, G. Frantziskonis, G. Hong, B. Djafari-Rouhani, and L. Dobrzynski, "Experimental evidence for the existence of absolute acoustic band gaps in two-dimensional periodic composite media," J. Phys.: Condens. Matter **10**, 6051–6064 (1998).

[4]R. Sainidou, N. Stefanou, and A. Modinos, "Linear chain of weakly coupled defects in a three-dimensional phononic crystal: A model acoustic waveguide," Phys. Rev. B **74**, 172302 (2006).

[5]Z. Hou, X. Fu, and Y. Liu, "Calculation method to study the transmission properties of phononic crystals," Phys. Rev. B **70**, 014304 (2004).

[6]Y. Lai and Z. O. Zhang, "Large band gaps in elastic phononic crystals with air inclusions," Appl. Phys. Lett. **83**, 3900–3902 (2003).

[7]A. Khelif, B. Djafari-Rouhani, J. O. Vasseur, P. A. Deymier, Ph. Lambin, and L. Dobrzynski, "Transmittivity through straight and stublike waveguides in a two dimensional phononic crystals," Phys. Rev. B **65**, 174308 (2002).

[8]J. O. Vasseur, "Matériaux à bandes interdites phononiques et photoniques," (Photonic and phononic materials with stop bands), Ph.D. dissertation (2002).

[9]J. O. Vasseur, A. Akjouj, L. Dobrzynski, B. Djafari Rouhani, and E. H. El Bondouti, Surf. Sci. Rep. **54**, 1–156 (2004).

[10]J. S. Jensen, "Phononic band gaps and vibration in one and two dimensional mass-string structures," J. Sound Vib. **266**, 1053–1078 (2003).

[11]W. M. Robertson and J. F. Rudy III, "Measurement of acoustic stop bands in two-dimensional periodic scattering arrays," J. Acoust. Soc. Am. **104**, 694–698 (1998).

[12]S. Parmley, T. Zobrist, T. Clougj, A. Perez-Miller, M. Makela, and R. Yu, "Phononic band structure in a mass chain," Appl. Phys. Lett. **67**, 777–779 (1995).

[13]R. James, S. W. Woodley, C. M. Dyer, and V. F. Humphrey, "Sonic bands, band gaps and defect states in layered structures—Theory and experiment," J. Acoust. Soc. Am. **97**, 2041–2047 (1995).

[14]J. N. Munday, C. B. Bennett, and W. M. Robertson, "Band gaps and defect modes in periodically structured waveguides," J. Acoust. Soc. Am. **112**, 1353–1358 (2002).

[15]D. Bria and B. Djafari-Rouhani, "Omnidirectional elastic band gap in finite lamellar structures," Phys. Rev. E **66**, 056609 (2002).

[16]J. S. Jensen and O. Sigmund, "Phononic band gap structures as optimal designs," *Proceedings of the IUTAM Symposium on Asymptotics, Singularities, and Homogenisation in Problems of Mechanics* (Klüwer Academic, Dordrecht, 2003), pp. 71–81.

[17]W. Chen, Y. Lu, H. J. Harris, and G. Xiao, "Picosecond ultrasonic study of localized phonon surface modes in Al/Ag superlattices," Phys. Rev. B **50**, 14506–14515 (1994).

[18]M. Shen and W. Cao, "Acoustic band-gap engineering using finite-size layered structures of multiple periodicity," Appl. Phys. Lett. **75**, 3713–3715 (1999).

[19]P. Gatignol, C. Potel, and J.-F. de Belleval, "Modal waves in periodically multilayered fluid structures," J. Acoust. Soc. Am. **105**, 1147–1148 (1999).

[20]N. F. Declercq and J. Degrieck, "A theoretical study of special acoustic effects caused by the staircase of the El Castillo pyramid at the Maya ruins of Chichen-Itza in Mexico," J. Acoust. Soc. Am. **116**, 3328–3335 (2004).

[21]A. C. Hladky-Hennion, G. Allan, and M. de Billy, "Localized modes in a one-dimensional diatomic chain of coupled spheres," J. Appl. Phys. **98**, 054909 (2005).

[22]A. C. Hladky-Hennion, A. Devos, and M. de Billy, "Quantitative analysis of the vibration modes in a finite set of coupled spheres," J. Acoust. Soc. Am. **116**, 117–124 (2004).

[23]ATILA Finite Element Code for Piezoelectric and Magnetostrictive Transducers Modeling, Version 5.2.1, User's Manual, ISEN, Acoustics Laboratory, Lille, France, 2002.

[24]C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996).

[25]L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1953).

Acoustical, sensory, and psychological research data and procedures for their use in predicting effects of environmental noises

Karl D. Kryter[a)]
School of Speech, Language, and Hearing Sciences, College of Health and Human Services, San Diego State University, San Diego, California 92182, USA

(Received 18 June 2007; accepted 30 June 2007)

A demonstration field-research study reveals that aircraft noise measured at two one-story houses is ~9 dB less attenuated from measured outdoor levels than is street traffic noise, and, found in other studies, ~14 dB less than railway noise. Comparable differences are found between these noises from the application of basic acoustical formulas for quantifying attenuations that occur on site of one- and two-story houses. Reasonably consistent with those findings are results from attitude surveys showing that daily exposure levels of aircraft must be ~8 dB less than levels of street traffic noise, and ~13 dB less than levels of railway noise to be perceived as an equal cause of annoyance and related adverse effects. However, USA government guidelines recommend that equal exposure levels of noise measured outdoors from vehicles of transportation should be considered as being equally annoying. Changes in present USA noise-measurement procedures and noise-control guidelines are proposed that provide more accurate predictions of annoyance, related adverse effects, and criteria for setting "tolerable" limits of noise exposure in residential areas. Key acoustical and psycho-acoustical principles and data pertaining to predicting correlations between dosages of environmental noises and its effects on people and land noise zoning in residential communities are examined. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2782748]

PACS number(s): 43.50.Ba, 43.50.Qp, 43.50.Lj [ADP] Pages: 2601–2614

I. INTRODUCTION

A. Quantification of exposures to aircraft, street traffic, and railway noise

Individual and multiple environmental noise events are generally quantified for government mandated "Environmental Impact Statements," Noise Control Act of 1972, in terms of one or more of the following metrics, or their equivalents. These metrics are either measured or calculated to be those present at a point outdoors usually near the closest source-facing façade of existing houses, or at some other common point on sites zoned for residential housing.

1. MaxdB: the maximum decibel (dB) level of instantaneous A-weighted sound pressures integrated over successive 1 s (one-second) intervals during the duration of a noise event. Throughout this text "decibel," unless otherwise specified, refers to units of sound measurements based on 1 s samples of A-weighted sound spectra.
2. SENEL (Single Event Noise Exposure Level, dB): the 1 s average of sound energy integrated over the time a given noise event exceeds the level of background noise, or as a practical matter is less than 10 dB below its MaxdB level. SENEL of point-source moving street, railway and aircraft noise decreases 3 dB per doubling of distance between the source and a more or less stationary receiver.
3. Annualized DNL, or Ldn, (day/night level), dB: the 1 s energy average of the SENEL's of the noise events from a given source at a given location during an annualized average 24 h day. A 10 dB "penalty" is added to events occurring during the time period of 10 PM–7 AM. DNL is the standard environmental noise metric used in the USA. An alternative metric, DENL (day/evening/night level), used in the European Union applies a 5 dB penalty to noise events occurring in the 7 PM–11 PM time period, and a 10 dB penalty to the events occurring between 11 PM and 7 AM. An average difference of ~1 dB is found between DNL and DENL, with DENL being the larger (Miedema and Oudshoorn, 2001).
4. The primary measure for quantifying the annoyance/disturbance felt by people in residential communities to different exposure levels of different environmental noises is the percentages of thusly exposed people who rate their subjective feelings of noise annoyance on numerical scales ranging from 0% to 100%. This numerical scale has been verbally categorized in surveys of environmental noise annoyance as, for example, "not or a little annoyed" (0%–20%); "moderately or more annoyed" (21%–40%); "middle of annoyance scale" (41%–60%); "very or more annoyed" (61%–70%); "highly or more annoyed" (71%–80%); "extremely or more annoyed" (81%–100%). Somewhat different partitioning and verbal categorizations have been applied in different research surveys of noise annoyance, or synonymously, noise disturbance.

[a)]Electronic mail: kdkryter@cox.net

B. The Schultz curve of annoyance from exposure to noise from transportation vehicles

Under contract with the United States Department of Housing and Urban Development, Schultz (1978, 1982) concluded that a single "trend curve" reasonably well describes the relation between: (a) the DNL of either aircraft, street traffic, or railway noise present at some point outdoors near the street-facing façades of houses, and (b) the percentages of thusly noise-exposed residents who report being "highly or more annoyed or disturbed." The surveyed residents reported that the annoyance/disturbance was primarily due to the fact that the noise masked the hearing of speech and other wanted sounds, interfered with sleep and rest, distracted the attention of the listener away from normal behavior and caused startled reactions for impulsive or unexpected sounds.

On the basis of Schultz's findings, USA federal government guidelines hold that equal DNL levels of either aircraft, street traffic or railway noise found outdoors of houses will cause equal annoyance to the residents in the houses. Accordingly as a general criterion, people who rated themselves to be "highly or more annoyed" (in the top ~30% of an annoyance scale ranging from "not or a little annoyed" to "extremely or more annoyed") could, from a public health and welfare point of view, be considered as living in an "unacceptable" level of environmental noise.

Relevant guides include those of the National Academy of Sciences (1977); Environmental Protection Agency, EPA (1978); Federal Interagency Committee on Urban Noise (FICUN), 1980; and Federal Interagency Committee on Noise (FICON), 1992. However, basic data and conceptual models for the assessment and prediction of the effects of these noises have been controversial over that period of time and to the present day (Kryter, 1982, 1983, 1985, 1994; Job, 1988; Schomer, 2002; Fidell, 2003).

II. ACOUSTICAL DOSAGE OF STREET TRAFFIC AND AIRCRAFT NOISE

The goals of the following demonstration study and analyses of on-site noise attenuation were to:

1. Measure, for two one-story houses, the house barrier and transmission attenuation between the outdoor and indoor, rear-to-front and front-to-rear yard, levels of noise events from both aircraft and street traffic.
2. Calculate, using acoustically principled mathematical formulas, the amounts of house barrier and transmission attenuation of aircraft, street traffic and railway noise between the outdoor and indoor, and between the front and back yard levels at one- and two-story houses.

A. Measurement of on-site attenuations of environmental noises

Recordings were made of the noise at ground level from commercial jet aircraft following takeoff operations and of the noise from passing street traffic. The recordings were made of signals from microphones placed at points in the front yards, rooms and back yards of two typical single-story stucco, brick and wood-sided, wood-framed, gabled houses near the Burbank California International Airport, see Fig. 1. The front façades are 14 m from the center of the street, the sides of adjacent neighborhood houses are separated by ~3.1 m and the back yards are ~15 m deep. The windows and doors of the houses were closed when the recordings were made.

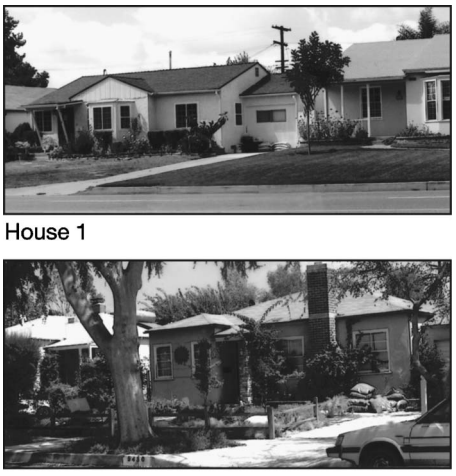


FIG. 1. Photographs of House 1, on Hollywood Way, and House 2, on Auckland Street, Burbank, California.

Microphone M1 was located in the front yard, microphones M2, M3 and M4 were placed in various rooms, and M5 in the rear yard of each house, see Fig. 2. The outdoor microphones, M1 and M5, were placed 1.82 m above ground level and 3 m from the front and rear house façades. The indoor microphones were placed 1.4 m above floor level in mid-room locations.

The 3 m placement of M1 and M5 from house façades was chosen to avoid including significant reflections of sound from the house façade in the noise measurements. It is

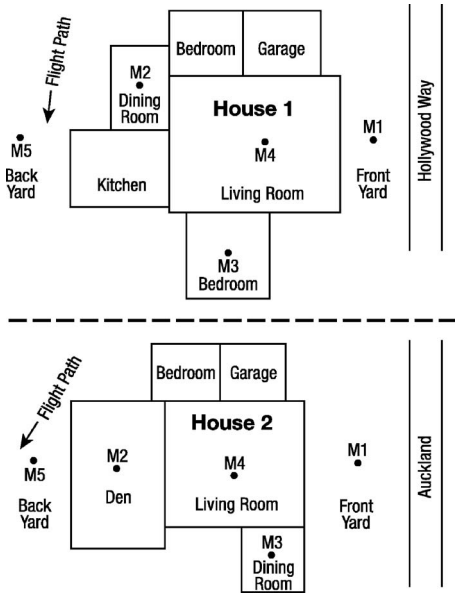


FIG. 2. Schematic drawing of Houses 1 and 2 and locations of microphones outside (M1, M5) and inside (M2, M3, M4) the houses.

presumed that the actual level of a noise at the closest façade surface from the source is fairly represented by the level found at the 3 m (10 ft) away point.

The noise events were recorded and analyzed by Veneklasen Associates of Santa Monica, California. The noises were analyzed for their A-weighted MaxdB, SENEL and octave band levels. A multi-channel digital recorder (Sony, PC16AX1), 1/2 in condenser microphones (Bruel & Kjaer 4155, and Larsen Davis 2559), with appropriate amplification and calibration equipment were used for making the recordings. Instrumentation for data analyses included: the Sony, PC216AX; Bruel & Kjaer 2230 SLM #2 w/Filter; Bruel & Kjaer 2260 Noise Investigator and Enhanced Sound Analysis Software.

B. Special factors

The following factors are germane to understanding the data and discussions to follow:

Factor 1. As a matter of safety for persons and property on the ground, a Federal Aviation Administration (FAA) regulation restricts commercial aircraft from flying over residential areas at altitudes below 1500 ft, ~458 m (FAA Regulation, Title 14, 1991). Furthermore, engine power settings and number of aircraft takeoff and landing operations must be such that the DNL in residential areas directly under the aircraft flight paths are 65 DNL or less [Federal Interagency Committee on Urban Noise (1980); Federal Interagency Committee on Noise (1992)].

With these restrictions and considerations in place, the slant range distance between the aircraft and houses to the right or left of the aircraft flight path will normally become longer. As a result, the noise level at the houses farther from either side of the flight path will be less than DNL 65 dB (or any other designated allowable limit), and are of only academic interest for purposes of aircraft noise control in residential areas.

Note in this regard, that for an aircraft flying at FAA prescribed minimum altitude of 458 m, one finds that with a slant angle of 75°, the slant range distance from the aircraft path to the house is increased to 515 m, and the Leq is decreased by ~0.4 dB. With a slant angle of 45° the slant range distance is increased to 793 m and the Leq is decreased by 2.4 dB. For purposes of regulating noise from commercial aircraft, the slant angle between the aircraft and residential houses should not be less than ~75°, otherwise the estimated impact of the aircraft relative to that of street traffic noise would be unjustifiably decreased.

Parenthetically, the average ~305 m altitude of the jet aircraft when passing H1, as will be reported below, is not anomalous with respect to the FAA regulation of 458 m concerning commercial aviation flights over residential housing. This occurs for this situation because the aircraft were flying over airport and not residential property.

Factor 2. As a general rule, except for the "corner" houses of city blocks, street traffic noise events will almost always move past the front yard and façade of houses on both sides of a street, and the house will significantly barrier-shield the street traffic noise from the rear façade and yard of houses. However, the noise from aircraft flying overhead at

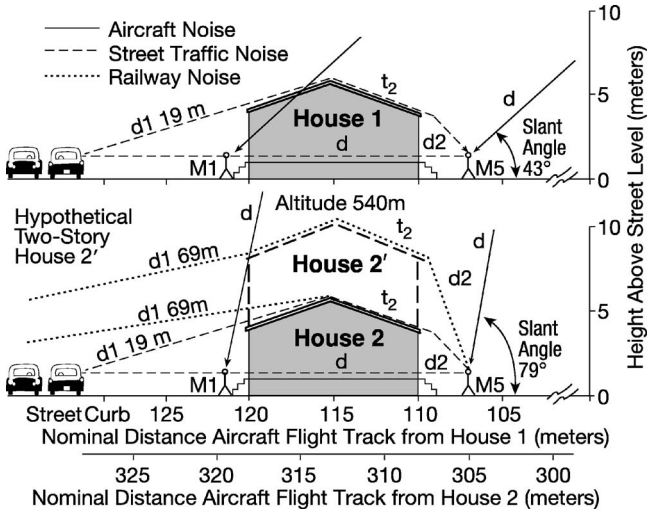


FIG. 3. Schematic showing House 1 and House 2, and hypothetical two-story House 2′, and orientation of street traffic and aircraft to the houses: House 1, aircraft: slant angle ~43°, slant range to M1, 431 m; House 2, aircraft: slant angle ~79°, slant range to M1, 622 m; orientation of street traffic to roof of gable of Houses 1 and 2, slant angle ~16°, range ~19 m; railway to Houses H2 and H2[1], range 69 m.

normal altitudes in residential areas may fall with minimal house-barrier attenuation on both the front and rear yards and façades of the houses.

For aircraft on flight paths that are to some degree parallel with and over or somewhat to a side of the street, for the houses on one side of the street, the closest house façade to the aircraft will: (a) be the rear yard and not street-facing façade, and (b) at the same time, for the houses on the other side of the street the closest house façade will be the reverse, the front yard street and the street-facing façade. The closest house façade to aircraft on flight paths that are to some degree perpendicular to the streets, sometime during the noise event the closest façade for houses on both sides of a street will be: (a) the rear yard and not street-facing façade, and also (b) the front street and yard facing façade of house.

Factor 3. As a practical matter for the present study it is assumed that any reflection or attenuation effects on aircraft or ground based vehicle noise when passing through spaces between the façade and roof surfaces of adjacent houses in typical airport-area residential neighborhoods will be of about the same magnitude for each of the subject noises.

Factor 4. The data in this part of the paper are concerned with differences in source and on-site physical sound and auditory attenuation principles that can cause differences in the levels of the subject noises as heard and perceived by listeners in the interior spaces and yards of houses. These include:

(a) Transmission attenuation due to geometric spherical divergence of sound waves from a moving source as a function of the distance between the source and receiver, herein labeled Adiv(ergence).
(b) Barrier attenuation due to a "thick" barrier, such as a house, of sound waves during transmission from the front yard to the rear yard (or vice-versa) of the house, herein labeled Abar(rier). The graphs in Fig. 3 schematically illustrate straight lines of sight (d) and interposed paths (d1, t,

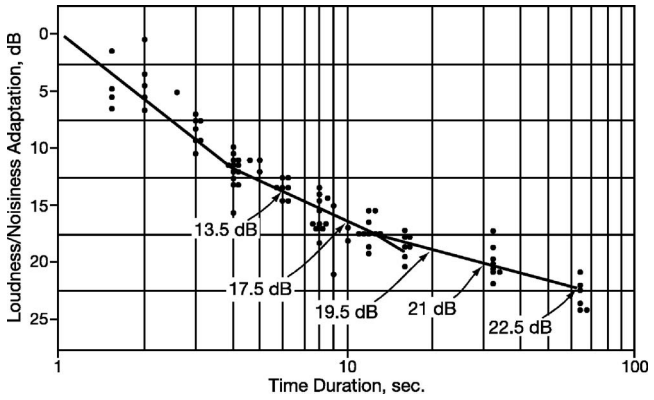


FIG. 4. Relative effect of duration on the perceived attenuation of loudness and noisiness of a sound/noise event. After Kryter and Pearsons (1963); Pearsons (1966).

and d2) of noise from the street traffic and aircraft sources to the front and rear yards, with H1, H2 and a hypothetical two-story house (H2′) serving as sound barriers. The Department of Engineering of the Burbank Airport furnished the distances and altitudes of the nominal flight paths and altitudes of jet aircraft from the locations of H1 and H2. That information was used to calculate the slant ranges and slant angles given in the caption to Fig. 3.

(c) Absorption attenuation and reflection of sound energy during transmission over ground, herein labeled Agro(und).

(d) Equivalent attenuation of the noises from the different types of transportation vehicles due to differences in the relative accessibility of their sound waves to the façades and roof surfaces of a house, and their transmissibility into a house interior, herein labeled Aacc(ess)/tran(mission):

(e) Equivalent attenuation of the noises from transportation vehicles due to a diminution in the loudness of sound or noise as a function of its duration, herein labeled Aadapt. This "equivalent-to-acoustical" attenuation is attributed to adaptive auditory neural equilibration and/or to temporary auditory receptor-fatigue processes (see Small). The effect appears to be based on physiological mechanisms, and results in decreased sensitivity to components of sounds or noises (see Fig. 4).

C. Results of research study of on-site noise attenuations

1. Measured MaxdB and SENEL

Table I shows the MaxdB and SENEL levels of street traffic and aircraft noise events found at each of five microphone locations at H1 and H2. The attenuations and differences in MaxdB and SENEL of the averages of street traffic and aircraft noise events at each house were derived from the data in Table I. The results are shown in Table II.

The last column of Table II shows that the average differences between SENEL and MaxdB measures of house attenuation is less than 1 dB. In these tables, and those to follow, the raw data are computer carried to multiple decimal places, but, with a few exceptions, i.e., in part of Table II, most of the table-printed data are in whole numbers rounded upward from decimals >0.49 and downwards from decimals <0.50. As a result the average sums of printed whole numbers may differ from their individually printed whole numbers by 1 integer.

TABLE I. MaxdB and SENEL for noise events at each house. Average MaxdB's and SENEL's based on event antilogs.

	Street Traffic			Aircraft	
House(H)1/	M1 Front		House(H)1/	M1 Front	
House(H)2	MaxdB	SENEL	House(H) 2	MaxdB	SENEL
H1, Truck	78	83	H1, Aircraft	80	87
H1, Motor Cy	74	81	H1, Aircraft	88	95
H1, Truck	72	78	Average, H1	86	93
Average, H1	76	81	H2, Aircraft	86	94
H2, Truck	80	85	H2, Aircraft	69	77
H2, Car	66	72	H2, Aircraft	77	86
Average, H2	77	82	Average, H2	82	90
	M5, Rear			M5, Rear	
	60	67		85	92
	56	64		91	99
	53	58	Average, H1	89	97
Average, H1	57	64		86	95
	60	67		72	81
	49	54		78	86
Average, H2	57	64	Average, H2	82	91
	M2, Rm			M2, Rm	
	58	64		63	71
	51	58		69	77
	51	54	Average, H1	67	75
Average, H1	55	60		Not Recorded	
	56	63		51	61
	41	48		61	69
Average, H2	53	60	Average, H2	58	66
	M3, Rm			M3, Rm	
	45	52		58	66
	48	54		64	72
	44	51	Average, H1	62	70
Average, H1	46	52		57	67
	52	58		50	58
	39	47		Not Recorded	
Average, H2	50	55	Average, H2	54	65
	M4, Rm			M4, Rm	
	49	56		66	73
	48	52		72	80
	50	52	Average, H1	70	78
Average, H1	49	54		41	69
	48	52		47	57
	34	43		53	61
Average, H2	45	50	Average, H2	49	65
	N=25	N=25		N=23	N=23
Aver. H1	56	62	Aver. H1	75	82
Aver. H2	56	62	Aver. H2	54	63

2. Calculated attenuation of overall and octave band levels of street traffic and aircraft noise

Two procedures are followed for quantifying differences in on-site attenuations of street traffic, railway and aircraft noise. One procedure is the analysis of the actual measured noise attenuation data from the present study, as shown in Table II. The second is the application of generalized acoustical formulas and principles available for the calculation of expected on-site attenuations of the subject noises.

TABLE II. Measured attenuations: Upper section: differences in levels of street traffic and aircraft noise between M1 and the other microphones; Lower section: same for street traffic but, for aircraft noise, between M5 and the other microphones. Based on Table I.

	Street M1-M2[a]	Aircraft M1-M2[a]	Str.-Air.	Street M1-M3[a]	Aircraft M1-M3[a]	Str.-Air.	Street M1-M4[a]	Aircraft M1-M4[a]	Str.-Air.	Street M1-M5[a]	Aircraft M1-M5[a]	Str.-Air.	M1(M2,3,4)[c] Street	M1(M2,3,4)[c] Aircraft	Str.-Air.	
H1, MaxdB	21	19	2	30	23	6	27	16	11	19	−4	22	26	19	6.6	
H1, SENEL	21	18	3	29	23	6	26	15	10	17	−4	21	27	19	8.2	Av.
Max-SENEL	0	1	−1	1	0	0	1	1	1	1	0	1	−1	1	−1.6	0.5
H2, MaxdB	24	23	1	27	26	1	32	26	6	20	0	21	28	25	2.7	
H2, SENEL	22	24	−2	27	25	2	31	25	6	18	−1	19	27	25	2.3	
Max-SENEL	3	0	3	1	2	−1	1	1	0	1	0	1	1	0	0.4	1.0
Av., MaxdB	23	21	2	29	25	4	29	21	9	19	−2	21	27	22	4.6	
Av., SENEL	21	21	1	28	24	4	28	20	8	18	−2	20	27	22	5.2	
Max-SENEL	1	0	1	1	1	0	1	1	0	2	0	2	0	1	−0.6	0
		M5-M2[b]			M5-M3[b]			M5-M4[b]			M5-M1[b]			M5(M2,3,4)[d]		
H1, MaxdB	21	22	−1	30	27	3	27	19	7	19	4	15	26	23	3.1	
H1, SENEL	21	22	−1	29	27	2	26	19	6	17	4	13	27	23	4.0	
Max-SENEL	0	1	−1	1	0	1	1	0	1	1	0	1	−1	0	−0.9	0.5
H2, MaxdB	24	24	1	27	27	1	32	26	6	20	0	20	28	26	2.0	
H2, SENEL	22	25	−3	27	27	0	31	26	5	18	0	18	27	26	0.7	
Max-SENEL	3	−1	4	1	0	1	1	0	1	1	0	1	1	−1	1.3	0.9
Av., MaxdB	23	23	0	29	27	2	29	23	7	19	2	17	27	24	2.5	
Av., SENEL	21	23	−2	28	27	1	28	23	6	18	2	16	27	24	2.4	
Max-SENEL	1	0	2	1	0	1	1	0	1	2	0	1	0	0	0.2	0.7
Av. Max+SEN	22	23	−1	28	27	1	29	23	6	18	2	16	27	24	2	0.6

[a]In dB, av. anti-logs M1-M2; M1-M3; M1-M4; M1-M5.
[b]In dB, av. anti.logs (M5-M2; M5-M3; M5-M4).
[c]In dB M1-av. sum anti-logs: M1-(M,2,3,4/3).
[d]In dB, av. anti-logs, aircraft noise: M5-(2,3,4/3).

Figure 3 illustrates straight lines-of-sight (d) and interposed paths (d1, t, and d2) of the noise to the front and rear yards of H1 and H2. When the line-of-sight path between a receiver and a source is below the top edge of a thick barrier, such as a house, the receiver is in a "shadow" zone of sound attenuation. When the line-of-sight path from the receiver to the source is at, or somewhat above, the top edge of a barrier, an excess amount of shadow attenuation of from 0 to 5 dB occurs due to sound refraction-interference effects (see Fig. 7.8 in Kurze and Beranek, 1988). Piercy and Daigle (1991) set this "excess" attenuation at a maximum of about 5 dB at a Fresnel N of −0.30.

Table II shows the H1+H2 average of the decibel amounts of house barrier attenuation calculated from measured octave band levels having center frequencies ranging from 125 to 4000 Hz. Similar calculations were applied to average source-to-receiver (human listeners) distances for the one-story houses, and a hypothetical similarly sized footprint two-story house, H2′ (see Fig. 3).

The measured MaxdB street traffic and aircraft noise events data differ from measured SENEL by less than 1 dB on average. For the purposes of measuring or calculating the decibel magnitude of attenuation of the subject noises, one focus of the present study, MaxdB and SENEL provide essentially the same results. This would indicate that for these noises some acoustical effects, such as Doppler frequency shifts, are practically insignificant (<1 dB).

3. Attenuations of street traffic, aircraft, and railway noise

Note that in Table III Abar for street traffic vs. aircraft noise is 17 dB measured and 21 dB calculated for outdoor listening. Ortega and Kryter (1982) found the measured difference in Abar of street traffic vs. aircraft noise by one-story houses to be 19 dB. However, the ~2 dB difference in the respective measured average Abar's is to be expected since the slant angle between the aircraft and one-story houses was somewhat less than the 90° directly overhead aircraft flyovers in the Ortega and Kryter (1982) study.

Fields and Walker (1980, 1982) surveyed annoyance reactions to railway noises of some 1453 respondents in Great Britain, 50% of whom lived in houses whose front, and closest façades were within ~68 m (225 ft) of the railway tracks. In a similar study Andersen *et al.* (1983), with comparable findings to Fields and Walker (1980, 1982), the median distance between the railroad tracks and noise-surveyed houses was shorter, ~38 m.

The average median distances from railroad tracks to median location of the median number of surveyed respondents for these two studies is 53 m. For present purposes, this distance is taken to represent a nominal distance between railway tracks and houses in urban residential areas. In the present study 14 m is the "closest-to-source" distance for street traffic noise.

TABLE III. Comparison measured versus calculated Abar of street and aircraft noise by one-story and hypothetical two-story houses.

	One-story house, average H1+H2										Hypothetical two-story house		
Oct. band c.f., Hz	Street meas., M1 SENEL[a]	Street meas., M5 SENEL[a]	Street M1-M5 Bmeas[b]	Str. calc. Bcalc[c]	Aircraft meas., M5 SENEL[a]	Aircraft meas., M1 SENEL[a]	Aircraft M5-M1 Bmeas	Airc. calc. Bcalc[d]	Str.-Airc. Bmeas	Str.-Airc. Bcalc	Street calc. M1-M-5. Bcalc	Aircraft calc. M5-M1. Bcalc[d]	Str.-Airc. Bcalc
125	63	48	15	13	70	68	−2	n	17	13	20	n	
250	70	51	19	16	72	75	3	n	16	16	23	n	
500	69	52	18	19	81	80	−1	n	19	19	25	n	
1000	71	52	19	22	79	84	5	n	15	22	28	n	
2000	68	49	20	25	74	74	1	n	19	25	31	n	
4000	57	38	19	28	57	60	2	n	17	28	34	n	
Arith. Av.	67	48	18	21	72	73	1	1	17	21	27	3	24
	Street traffic, Bcalc-Bmeas			2	Aircraft, Bcalc-Bmeas			0	Bcalc Str.-Bcalc Air; aver. one+two story				3

[a]Measured octave band levels of street traffic and aircraft noise when events reached maximum overall MaxdB.
[b]Bmeas: octave band levels measured for closest façade, M1, for street traffic, and M5 for aircraft noise, see Table II.
[c]Bcalc: Eq. (3.13): N, (Freznel)=(2/w) (d1+t+d2−d), Eq. (3.14): 10 log(3+30 NK)- Aground, -Aair, Eqs. from Piercy and Daigle.
[d]Aircraft noise, N's are negative (n), set to zero. Transition zone excess attenuation estimated, see Fig. 7.8, Kurze and Beranek.

Table IV shows that for one-and two-story houses, the average amount of calculated geometrical spherical divergence attenuation (A_{div}) of street traffic vs. railway noise between (a) the front façade of houses and the interior rear wall of the house is 2 dB if heard indoors (3−1 dB), and (b) is 3 dB if heard outdoors in the rear yard of the house (4 −1 dB). The distances between the source of the noise and the nominal interior rear wall of the house is 27 m for street traffic and 65 m for railway noise. The noise levels are estimated to be practically homogeneous within rooms due to sound reflections and reverberation. The distance between the source and rear yard is 35 m for street traffic and 75 m for railway noise. The differences in Adiv for the aircraft noise is 0 dB due to the hundreds of meters between the aircraft and noise-measurement at any points around a house.

In real life, the distances between listeners and the levels found at the house façade closest to the source of the noise are not acoustically directly relatable to Adiv at the receiver's (listener's) positions. Piercy and Daigle's (1991) equations (Eqs.) 3.17, 3.18, 3.19, the term r in "*ref*-20, 10, and $5\log_{10}(r/r_{ref})$", r represents the distance from source to ultimate receiver (listener), and the term r_{ref} is the measured or estimated level that is intermediate between the source and the listener. That is, the effective Adiv is the distance from the source to the listeners, not the distance from the point at of the noise level measurements and the listeners.

Table IV shows that the differences between closest-to source (M1) versus farthest-to-source façade (M5) for the calculated Abar attenuations of street and railway noise are: (a) the same for both one- and two-story houses (0 dB indoors, and 21 dB outdoors); (b) outdoor exposure levels at closest-to-source house façade versus their levels at façades farthest from source (M1-M5) are 6 dB greater for two, than for one-story houses (21 vs. 27 dB), an average of 24 dB for both types of houses.

Table IV shows the differences between closest-to-source façade (M5) versus farthest-to-source facade (M1) Abar attenuations of aircraft noise are: (a) 0 dB attenuated indoors by both one- and two-story houses; (b) and 1 dB attenuated outdoors by one-story and 3 dB attenuated outdoors by two-story houses, an average of 2 dB for both types of houses.

TABLE IV. On-house-site decibel attenuations, and equivalents, of street, railway, and aircraft noise for one- and two-story houses and their sums.

Source to	Street	Street	Street	Street	Street	Street	Rail	Rail	Rail	Rail	Rail	Rail	Airc.	Airc.	Airc.	Airc.	Airc.	Airc.
mics, meters	15/27/37 Adiv[a]	M1-M5 Abar[b]	15m Agro[c]	M1-M[f] Aaccd/tran[d]	6s Aadapt[e]	Sum A's	53/65/75 Adiv	M1-M5 Abar	53m Agro	M1-M[f] Aacc/tran	20s Aadapt	Sum A's	458m Adiv	M5-1 Abar	0m Agro	M5-M[g] Aacc/tran[d]	10s Aadapt	Sum A's
One-story houses																		
Indoors	3	0	0	27	14	43	1	0	4	27	20	51	0	0	0	24	18	42
Outdoors	4	21	0	0	14	38	1	21	4	0	20	46	0	1	0	0	18	19
Two-story houses																		
Indoors	3	0	0	27	14	43	1	0	4	27	20	51	0	0	0	24	18	42
Outdoors	4	27	0	0	14	44	1	27	4	0	20	52	0	3	0	0	18	21
Average one and two-story houses																		
Indoors	3	0	0	27	14	43	1	0	4	27	20	51	0	0	0	24	18	42
Outdoors	4	24	0	0	14	41	1	24	4	0	20	49	0	2	0	0	18	20

[a]Adiv 15m to front, 27m rear wall, 37m rear patio; Rail: 53m front, 65m rear wall, 75m rear patio. Eq. 3.18: $L_{eq\ ref}-10\log_{10}(r/r_{ref})$, Piercy & Daigle.
[b]Abar: Measured M1-M5 street, presumed same for railway noise, M5-M1 for aircraft noise, Table III, & Eqs. 3.14, 3.13 and 3.14, Pierey & Daigle.
[c]Agro: Front yrd. 15m street noise, 53m rail noise. Oct. band 500 Hz, source 0.3 m above grass. See Table 3.2, Piercy and Daigle.
[d]Aacc/trans: diff. in size of house facades and roof areas exposed to aircraft vs. street/railway noise (diff. of ~1/2 overall of street/railway vs. airc.).
[e]Aadapt:a loudness/noisiness dimunition due to duration of noise event. Equivalent, on a decibel basis, the exposure level of the noise event. See Text.
[f]M1 – (Average indoor microphones, (M2, M3, M4)/3, Street and Rail. Noise. See Table II.
[g]M5 – (Average indoor microphones, (M2, M3, M4)/3, Aircraft noise. See Table II.

Table IV, based on Table 3.2 of Piercy and Daigle, shows that front yard grass/vegetation/surface sound absorption-reflection (A_{gro}) is taken to be equivalent to an increase of 0 dB for street traffic, and a decrease of 4 dB for railway noise. It is assumed that A_{gro} of these noises in the rear yards are insignificant.

Figure 3 shows that the noise from aircraft falls nearly freely on street-facing and rear façades and roofs of houses, but that the noise from street and railway transportation vehicles will fall freely on only the street- and railway-facing house façades, and approximately one half of the roof surfaces. This is equivalent to the aircraft noise being ~3 dB less $A_{acc/trans}$ than street traffic and railway noise. The measured outdoor levels of street traffic and railway noise were indeed 3 dB less indoors (27 dB attenuation, M1- (M2, 3, 4)/3)), than were the measured outdoor levels of aircraft noise (24 dB attenuation, M5- (M2, 3, 4/3)), see Table II. Likewise, for calculated outdoor levels of aircraft noise, the equivalent measured $A_{acc/trans}$ is 3 dB, 27 dB vs. 24 dB (see Table IV).

The equivalent nominal A_{adapt} attenuations are calculated to be: 14 dB for street traffic (6 s duration), 27 dB for railway noise (20 s duration), and 24 dB for aircraft noise (10 s duration). These values are estimated from the function shown in Fig. 4. Table IV shows that for one- and two-story houses the sums of A's are: (a) for street traffic: 43 dB indoors and 41 dB outdoors, (b) for railway noise: 51 dB indoors and 49 dB outdoors, and (c) for aircraft noise: 42 dB indoors and 20 dB outdoors.

III. EDNL

A. Introduction

The DNL metric is denigrated for its originally intended purpose because it does not adequately take into account established physical principles involved in the transmission and attenuation of environmental noise to the ears of listeners indoors or in the yards of their homes, nor with physiological stimulation of the human auditory system. It is proposed that the findings shown in Table IV be used as a basis for the formulation of a new metric for DNL. This new metric is herein labeled E(ffective)DNL, or EDNL, of transportation vehicle, or other sources of environmental noise.

Four basic premises of the proposed EDNL metric are that:

1. Airborne acoustical energy in environmental noise events present at, or near, the external ear(s) of listeners is the normal adequate stimulus for auditory sensory sensations and perceptions.
2. Squared sound pressure and duration in seconds contribute equally to the perceived exposure level of a sound or noise event.
3. The normal human auditory is a *par excellence* receptor and near real-time analyzer of the spectral and durations aspects of sound and noise.
4. EDNL is applicable in the prediction of the adverse effects of given dosages of noise events on people who have experienced and come to expect the noise events to be part of their living environment.

The proposed EDNL metric is primarily concerned with the measurement and prediction of annoying/disturbing and associated adverse effects of noise due acoustical, sensory, perceptual auditory attributes of unwanted sound. These attributes will be discussed more fully later.

B. Calculation procedures

EDNL of a given, nonaircraft, environmental noise is its measured or calculated DNL value, plus the summed decibels of its on-site attenuations (including equivalent attenuations) minus the on-site attenuations of aircraft, or another environmental designated-reference noise. The EDNL of a given environmental noise compared to the impact of aircraft noise of a given DNL is calculated as follows:

$$\begin{aligned} \mathrm{EDNL}_x = \mathrm{DNL}_x & \\ & + (\text{arithmetically summed decibels A's}_x) \\ & - (\text{arithmetically summed A's}_{\text{aircraft}}), \end{aligned} \quad (1)$$

where x is a given environmental noise.

Aircraft noise is chosen as an "anchor" reference for effective EDNL of environmental noises in general because it has been the dominant and most researched environmental noise. Note that EDNL does not disturb the numerical relations between the Federal Interagency Committee on Urban Noise (1980) and the Federal Interagency Committee on Noise (1992) calculated DNL and annoyance effects but serves as a means of converting DNL into a more truly *effective* noise-dosage level. Attention is invited to the possibility that to make more accurate and general the DNL and EDNL metrics, it would be appropriate to substitute the so-called D weighting for the A weighting in the measurement or calculation of SENEL of the noises involved. This matter will be discussed later in the paper.

1. Indoors and outdoors listening

Because DNL and EDNL use the equal-energy equation, the effective EDNL for different ratios of the hearing indoors and outdoors of environmental noise events during a typical 24 h day can be calculated as follows:

Step 1. Multiply the arithmetically summed decibels of attenuations, A's, of street, railway, or aircraft noise shown in Table IV, by chosen fractions of annualized 24 h days, when listeners are indoors and when outdoors (in yards) of their homes.

Step 2. The indoor and outdoor decibel sums from Step 1 are added on an energy, antilog, basis with the resulting sums converted to decibels for each type of noise source.

Step 3. Subtracting the results of Step 2 for one noise source from another source gives the relative differences in daily effective, EDNL, noise dosages between the two sources.

Table V shows for various ratios of average percentages of annualized daily time spent by residents indoors vs. outdoors of homes, the decibel sums of on-house-site attenuations (including equivalent attenuations) applied to the dif-

TABLE V. Sums of A's, % indoor +% outdoor listening, and differences between sources.

Listeners	Average for one and two-story houses: Street Sum A's	Rail Sum A's	Aircraft Sum A's	Diff. dB atten., diff. sources: Rail-Str.	Str.-Airc.	Rail-Airc.
100% Indoors	43	51	42	8	1	9
100% Outdoors	41	49	20	8	21	29
90% Indoors	39	46	38	7	1	8
10% Outdoors	4	5	2	1	2	3
dB attenuation: % ind. + % out.=				8	3	11
80% Indoors	34	41	34	6	1	7
20% Outdoors	8	10	4	2	4	6
dB attenuation: % ind. + % out.=				8	5	13
75% Indoors	32	38	32	6	1	7
25% Outdoors	10	12	5	2	5	7
				8	6	14
70% Indoors	30	36	29	6	1	6
30% Outdoors	12	15	6	2	6	9
dB attenuation: % ind. + % out.=				8	7	15
60% Indoors	26	31	25	5	1	5
40% Outdoors	16	20	8	3	8	12
dB attenuation: % ind. + % out.=				8	9	17

ferent transportation noises, and the differences in summed daily attenuations between each of the respective types of noise.

The selection, as will be discussed later, of the 70% indoor with 30% outdoor "at-home-living" is based on the greater similarity (average differences of less than 2 dB) of the decibel averages of measured synthesis attitude annoyance survey data to calculated on-site attenuations of street traffic and railway noise in comparison to aircraft noise. Table V shows that with less than ~30% outdoors, or more than ~70% indoors, the similarities between annoyance scores and noise attenuations are somewhat less. Annualized 30% outdoor living is perhaps tantamount to about 50% or so outdoor daytime/summer time living with open windows and doors; and that annualized 70% indoor living is perhaps equal to 90% or so indoor/winter time living with closed windows and doors.

2. Monthly annualized effective levels of noise exposure

Figure 5 shows the monthly telephoned rate, compared to annual rate, of annoyance complaints received about aircraft noise as related to the average number per year over a four-year period. The complaints were from residential neighborhoods near four airports in the northeastern quarter of the United States. It is seen that for the about six months of April to October, the average monthly number of complaints increases to a peak in July that is the equivalent of 8 dB increase in SENEL (or DNL) levels.

The "tick marks" above the abscissa on Fig. 5 show the months in which the attitude surveys of noise annoyance in studies were conducted in the fall, winter, or spring in temporal climates. The tick marked studies represent those used by Schultz, and Miedema and Oudshoorn in the derivation of their synthesis curve.

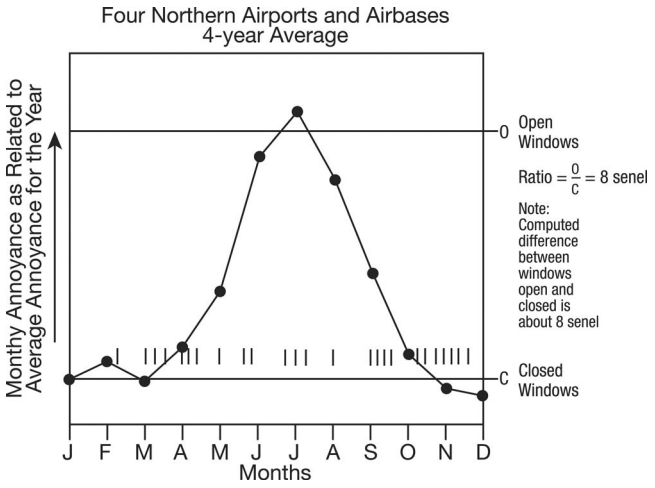


FIG. 5. Total monthly volunteered, via telephone, annoyance complaints over a four-year period by residents near four airports in the northeastern quarter of the United States. After Beranek, Kryter, and Miller (1959). Tick marks above abscissa indicate approximate months of the year attitude surveys of aircraft noise used in the Schultz and Miedema and Oudschoorn synthesis were conducted.

It is surmised that the magnitude of annoying/disturbing effects from transportation noise shown in Fig. 5 vary as function of seasonal temperature changes during typical years. The approximately six months during which the subject annoyance surveys were conducted are those when the residents would spend more time indoors with windows closed, and thus be less impacted by the transportation noises than in the warmer months when the residents would spend more time outdoors and indoors with windows open.

Note that Miedema, Fields and Vos (2005), found a similar pattern of higher annoyance reactions during 'summer' compared the 'colder' months of the year to transportation noises. The difference was statistically significant. However, the scoring procedures employed on these later studies does not permit a direct equivalent decibel comparison with the data in Fig. 5.

The 8 dB SENEL peak-month different level shown in Fig. 5 in effective exposure level translates into ~2 dB lesser SENEL/DNL value when referenced to a 12 month averaged effective level of noise from a moving source. It could be argued that annualized DNL or EDNL should be decreased by ~2 dB in order to best represent the average monthly effective exposure level over a calendar year.

The important empirical question is: how closely do different methods for measuring or calculating real-life dosages of environmental noise correlate with listener's self-rated subjective annoyance/disturbance effects and with objectively measured, by independent experts, effects on public health, well-being and welfare? The next section of the paper addresses that issue.

IV. PSYCHOLOGICAL ANNOYANCE AND ASSOCIATED EFFECTS

A. Introduction

Acoustical "noise" is defined as sound that is perceived as being unwanted by a listener. The auditory physiological-

TABLE VI. Percent of people highly or more annoyed by aircraft, street and railway traffic noise, and % point differences.

Section A	DNL55 Aircraft	DNL60 Aircraft	DNL65 Aircraft	DNL55 Street	DNL60 Street	DNL65 Street	DNL55 Railway	DNL60 Railway	DNL65 Railway
Synthesis, Schultz, 1978	3	8	15	3	8	15	3	8	15
Synthesis, Fidell, *et al.*, 1991	3	8	15	3	8	15	3	8	15
Synthsis. Finegold *et al.*, 1994	3	8	14	3	8	14	3	8	14
Average	3%	8%	15%	3%	8%	15%	3%	8%	15%
Section B									
Study, McKennell, 1963[a]	13	20	24	No data	No data	No data	No data	No data	No data
Study, Hall *et al.*, 1981	13	16	29	4	10	20	No data	No data	No data
Synthesis, Kryter, 1982	10	16	29	0	4	10	No data	No data	No data
Synth., Mied. & Oud.'01	14	20	29	6	10	17	2	4	8
Average	12%	17%	29%	3%	8%	16%	2%	4%	8%
% points diff. Section B-A:	+10% pts.	+10% pts.	+13% pts.	+1% pts.	+1% pts.	+1% pts.	−1% pts.	−4% pts.	−7% pts.

[a]McKennell, in Committee on Problem of Noise, Appendix X1, Final Report (1963). HM Stationery, Office London.

sensory and perceptual mechanisms involved in this unwanted ness are in brief that sound:

1. Stimulates sensory-neural cells in the cochlea of the peripheral ear by causing turbulence in the fluids in the cochlea that mask or distort turbulences being simultaneously, or nearly simultaneously, caused by other then occurring sounds that may convey wanted auditory information or auditory sensations. For example: (a) music can be perceived as being noise by a listener who is trying to understand wanted speech sounds that are being physically masked in the cochlea by unwanted music sounds, and (b) speech sounds can be perceived as noise by a listener who is trying to hear wanted music sounds that are being physically masked in the cochlea by unwanted speech sounds.
   Masking automatically occurs independent of any cognitive higher-brain neural activities. Masking is a basic attribute of hearing and a pervasive cause of subjective annoyance and related effects in everyday living activities independently of any psychological attitudes, biases, or abilities that may differ among people or peoples.
2. Such mechanical-neural activity in the cochlea in turn neurally stimulates higher brain centers, vegetative, homeostatic organs and autonomic systems of the body, particularly if the sound (noise) is unexpected. Such reactions arouse a person from quiescence or sleep and may cause emotional feelings of annoyance, fear, and anger. However, cognitively learned "meanings" are not as directly causally related to, controlled by, dependent upon, or validly regulateable, as are the basic unlearned auditory system sensory-perceptual and bodily interactions with unwanted familiar sounds or noises.
3. Cochlear masking and autonomic system stimulations from exposures to noise can also distract the attention of listeners away from, or require more than normal effort to adequately perform mental and motor behavioral tasks that depend upon correctly hearing speech, sounds from operating machinery, audible warning signals, etc.
4. If at a sufficiently intense level, above ~80 dB, and brief, less than ~0.5 s duration, impulsive sounds, e.g. gunshots, sonic booms, breaking objects, etc., cause autonomic-system startle reactions that create feelings of annoyance and fright.
5. The noises of most public concern and research study have been from vehicles of transportation, but the basic acoustical, sensory and psychological research findings from those studies are presumably applicable to assessment of the effects of other environmental noises, such as those from construction activities and industries.

B. Studies of annoyance from aircraft, street traffic and railway noise

The most recent analyses and synthesis of annoyance as a function of DNL and DENL exposure levels to transportation noise are those of Miedema (1993); Miedema and Vos (1998); Miedema and Oudshoorn (2001). Miedema and Oudshoorn (2001) mathematically modeled the data from 20 surveys of annoyance from aircraft, 18 from road/street traffic, and 9 from railway noise. Strict criteria were followed in scaling the annoyance data to insure comparability among the different surveys. The surveys were conducted in residential areas in Australia, Canada, Europe and the United States. Some of their findings, and those from other relevant studies, are shown in Table VI.

Table VI shows, for exposure levels DNL 55, 60 and 65 dB, the percentages of people in residential neighborhoods found in various studies and syntheses of attitude survey data to be "highly or more annoyed" by street traffic, railway and aircraft noise. The bottom line of Table VI shows that the differences between section B minus A, averaged over DNL 55, 60 and 65 dB, are: +11% points for aircraft, +1% points for street, and −4 points for railway noise.

C. Adverse nonauditory effects associated with annoyance

1. Emotional stress

It was found in a series of epidemiological medical field studies that with increases in exposure levels to aircraft noise above DNL 50 dB there are progressive increases in hypertensive blood pressure, prescription drug usage and physi-

TABLE VII. Nonauditory effects, aircraft and street traffic noise.

	% People high blood pressure			% House Depreciation			
	Aircraft	Street traffic		Aircraft: House Price range			
DNL	Knipschild	Lercher and Kofler		DNL	Low	Med	High
50	6%	DNL		50		.	0%
55	9%	<50	0%	55		0%	5%
60	12%	59	0%	60	0%	4%	10%
65	15%	64	0%	65	4%	7%	15%
70	19%	>65	0%	70	6%	11%	20%

cian contacts for general "stress" disorders (Knipschild, 1980). Hypertensive blood pressures were not found in a similar research study (Lercher and Kofler, 1993), with street traffic noise of comparable and greater exposure levels than in the Knipschild studies of aircraft noise, see left-hand section of Table VII. Also, see review by Babisch (1998) of other European studies that also show no consistent adverse cardiovascular effects associated with exposures to street traffic noise at DNL levels equal to levels of aircraft noise that are associated with physiological and psychological stress reactions.

A longitudinal study by Ohrstrom (2004) revealed that the reduction of average 67–55 dB $L_{A24h,eq}$ of street/road traffic noise significantly reduced annoyance and disturbance ratings and feelings of lower psycho-social well-being. ($L_{A24h,eq}$=1 s average of A-weighted levels over the 24 hour day). Note that the application of evening and nighttime decibel penalties would translate the cited $L_{A24h,eq}$ values into somewhat higher DNL and DENL values.

The noises associated with hypertensive blood pressure and other physiological stress responses, when they occur, are putatively not caused by exposure to the sounds per se, but are aftermaths of emotional resentment and anger triggered by the unwanted annoyance/disturbance effects of the noises. These stress responses are normal homeostatic cardiovascular, and autonomic nervous-glandular-system responses of the body to emotions and startle.

2. Depreciation of real estate values

An economic consequence of annoyance from aircraft noise is that people are apparently loath to purchase houses in environments that have DNL levels that exceed ~55 dB, for "high" priced, ~60 dB for average/medium priced, and ~65 dB for low priced house, as shown in the right-hand section of section of Table VIII. These findings represent averages of government sponsored analyses conducted in: USA, Nelson (1979); in Canada, Mieszkowski and Saper (1978); and in England, the Commission on the Third London Airport (1971).

The adverse effects cited above are based on objective data obtained by professional medical personnel and real estate appraisers. As such, these data lend practical meaning to and supportive validations of the subjective annoyance attitude survey data as related to noise level measurements.

D. Relations between EDNL and annoyance/disturbance from transportation vehicle noise

Table VIII shows that the average differences over the range of 50–70 dB between equally annoying EDNL levels of aircraft compared to street traffic noise are: (a) 0 dB by Schultz (1978); Fidell *et al.* (1991); and Finegold *et al.* (1994), (b) 10 dB by: Ollerhead (1980); Hall *et al.* (1981); and Kryter (1982), (c) 6 dB by Miedema and Oudshoorn (2001). The overall average differences between aircraft compared to railway noise are 0 dB according to Schultz and cohorts, and 13 dB according to Miedema and Oudshoorn (2001).

The average, 8 dB, of the Ollerhead (1980), Kryter (1982), Hall *et al.* (1981), and Miedema and Oudshoorn (2001) syntheses plus, again, the Miedema and Oudshoorn (2001) synthesis is herein taken to be the best available acoustical and psychological research data for predicting annoyance and associated adverse effects of street traffic compared to aircraft noise, see Sec. II of Table VIII. The double inclusion of Miedema and Oudshoorn (2001) synthesis justifiably gives it extra weight, but the other syntheses also have merit.

TABLE VIII. Sections I through III: differences in DNL between street, railway, and aircraft noises when ~30% people are equally annoyed. Section IV, differences between calculated sums of indoor/outdoor attenuations for listeners' positions at home.

Section:	1		2					3		4		
	[a]Schultz, '78;			Street-Aircraft				Rail-Airc.	Rail-Str.	Str.-Airc.	Rail.-Airc.	Rail-Str.
	Fidell *et al.* '91;		Kryter,[b]	[c]Av., Hall,	Miedema and		[c]Aver.	Miedema and		Diff. Calc. on-site attenu.-		
DNL	Finegold *et al.* '94	Ollerhead, '80	1982	Fidell, '81	Oudshoorn, '01		Sect. 2	Oudshoorn, 2001		measure diff. DNL-DNL		
50	0[a]	10	10	5	6	6	7	15	9	9	14	10
55	0[a]	10	10	9	7	7	9	15	9	9	14	10
60	0[a]	10	10	10	6	6	8	13	10	9	14	10
65	0[a]	10	13	10	6	6	9	12	10	9	14	10
70	0[a]	10	10	12	5	5	8	10	10	9	14	10
Av.	0[a]	10	10	9	6	6	8	13	10	9 [7][d]	14 [15][d]	10 [8][d]

[a]No differences: Street, Railway, and Aircraft. Schultz and cohorts data not included in section 2.
[b]Average of differences between aircraft-street traffic noise annoyance. Based on Hall *et al.* '81 and Fidell *et al.* '81.
[c]Average Ollerhead thru Miedema and Oudshoorn, see text.
[d][] Av. differences in EDNL between surveyed annoyance for the different noise sources per calculated atten. of noises judged to be equally annoying as aircraft noise at equal DNL, see 70% indoor/30% outdoor, Table V: Diffs. of (2,−1,2)=av.1 dB.

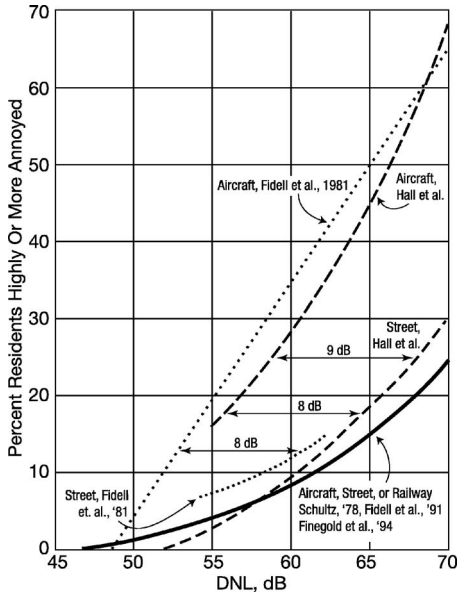


FIG. 6. As function of DNL, decibel differences between percentages of residents highly or more annoyed by aircraft compared to street noise as found by Fidell *et al.* (1981) (dotted curves) and by Hall *et al.* (1981) (dashed curves).

The studies by Fidell *et al.* (1981), and Hall *et al.* (1981) can claim special validity in that the data avoided experimental errors possibly present in other annoyance/disturbance surveys. Unlike most attitude surveys, in these two studies the same residents rated annoyance or disturbance in response to the same attitude questionnaires for both street traffic and aircraft noise, see Fig. 6.

Note that: (a) in the study by Fidell *et al.* (1981) the respondents falling in the top 40% of the annoyance scale were considered to be "highly annoyed," and (b) a bi-modal annoyance questionnaire was used in the study by Hall *et al.* (1981). As a result, the absolute percentages of people considered to be highly or more annoyed in both of those studies could be expected to be somewhat greater than found when ~30%, the usual criterion limit for people being highly or more annoyed, is used. However, the relative differences (~8 to ~9 DNL) found between the annoyance/disturbance data for aircraft vs. street traffic noise should be valid for both studies.

1. Relation between percent people highly annoyed and on-site attenuation magnitudes

Table VIII summarizes functions of DNL 50–70 dB: (a) when ~30% of respondents in attitude surveys are highly or more annoyed/disturbed because of noises from street, railway and aircraft traffic (Secs. I–III), and (b) measured and calculated decibel differences in on-site attenuations of the street, railway and aircraft traffic noises (Sec. IV). Note that as the formula defined, there are no differences between calculated on-site attenuations as a per se function of EDNL-street or railway DNLaircraft noise and further, as noted before, the measured and calculated on-site attenuations differed on average by less than 1 dB.

Data in Table VIII:

Sec. I. The Schultz and up dating, "cohorts," syntheses functions show no differences in DNL exposure levels when ~30% of exposed people were highly or more annoyed by each of the three types of transportation noises.

Sec. II. Shows that, for four other syntheses functions, over the 50–70 DNL range the average decibel difference between street traffic DNL compared to aircraft noise DNL when ~30% of respondents were highly or more annoyed is 8 dB (with Miedema and Oudshoorn (2001) twice weighted).

Sec. III. Shows that over the 50–70 DNL range when ~30% of respondents were highly or more annoyed the average decibel difference between railway DNL compared to aircraft noise DNL is 13 dB, and the average difference between railway DNL and street traffic DNL is 10 dB.

Sec. IV. Shows that over the DNL 50–70 dB range the average decibel differences between on-site attenuations are calculated to be: (a) 9 dB for street traffic-aircraft noise; (b) 14 dB for railway-aircraft noise; and (c) 10 dB railway-street noise. Also shown in brackets [ ] are the decibel differences between the DNLs of the street traffic and railway noises compared to the DNL of aircraft noise when ~30% of respondents judge each of the noises to be highly or more annoying/disturbing. The average of these differences is 1 dB, (2, −1, 2)/3).

Schultz, Fidell *et al.* (1991) and Finegold *et al.* (1994) syntheses curves are implicitly based on the premise that the magnitude of subjective annoyance in residential neighborhoods from different vehicles of transportation is proportional to their SENEL/DNL levels measured at, or estimated for, a common point outdoors near the closest-to-source façade of houses. The differences in on-site attenuations found for the noise from the different transportation sources of noise indicate that this premise is untenable for purely physical acoustic reasons.

Also note that Schultz, Fidell *et al.* (1991) and Finegold *et al.* (1994) did not, as was done for the other studies and syntheses in Table VIII, plot separately for each type of transportation noise the percentages of people who were highly or more annoyed as a function of DNL. Instead, Schultz and cohorts each plotted the annoyance data for street traffic, railway and aircraft noise on a single set of coordinates and fitted a single mathematical-formula curve to the scatter of all the data points: Schultz, a third-order polynomial; Fidell *et al.*, a least-squares quadratic; Finegold *et al.* (1994), a logest. The best-fitting curves for these three syntheses are nearly identical. However, those procedures possibly obfuscated the presence of separate clusterings for each type of noise source.

E. Generality and reliability of EDNL street, railway noise—DNL aircraft noise

Most of the research on annoyance and other adverse effects of transportation noise on people has been centered in urban neighborhoods near commercial jet airports. These neighborhoods, by and large, consist of one- and two-story houses located on lots of approximately one-sixth of an acre

(~640 sq. meters). The setbacks of the front façade of houses to the center of the street are typically ~15 m (50 ft), and 3.2 m (10 ft) between houses. Such distances are promulgated in international building codes for residential housing.

The proposed EDNL procedures can be applied to urban and suburban neighborhoods having different lot sizes, number of stories or building heights, house construction codes, etc.. The procedure would be to calculate, in accordance with formulas specified in Table IV, on-site house attenuations (Adiv, Abar, Agro, Aacc/trans and Aadapt) as needed for the noise from different sources. However, note that Adiv, Agro and Aacc/trans attentions of aircraft noise should generally be zero because of FAA "fixed" minimum allowable altitude of aircraft when flying over residential areas.

On-house-site attenuations for noises having broad frequency bands, e.g., noises of transportation vehicles, should, within limits, be approximately the same regardless of the materials, architectural design, or building codes used in the construction of houses. The presumed reason is that whatever perturbations may be imposed by the per se house structures on the spectrum of the noise of such sources will also be applied to the reference aircraft noise used in the calculation of EDNL. That practically this might be the general case is attested to by the close similarity of decibels of on-site attenuations and decibel equivalents for equal psychological annoyance/disturbance from transportation vehicles (within ~2 dB, average 1 dB, see bottom line Table VIII).

F. Predicting annoyance of "special" environmental noises

The Federal Interagency Committee on Urban Noise (1980) and the Federal Interagency Committee documents use A-weighted spectra sound level meters, or equivalent, when measuring exposure levels of noise from typical vehicles of transportations. Most of the sound energy in the noises from these vehicles of transportation have spectra that fall above ~315 Hz, and are in a range where their judged perceived noisiness is fairly proportional to their integrated one-third octave band levels. However, auditory research shows that the A-weighting has limitations when used in the measurement of noises having significant energy in sound frequencies below ~315 Hz. A "D"-weighted sound level meter gives more relative weight than does A weighting for frequencies below 315 Hz. The lesser weight is consistent with the "critical" bandwidth theory (Fletcher, 1940) and measurement of the perceived loudness and noisiness of sounds containing significant energy in the lower frequencies (Kryter, 1970, 1994).

For example, the "backblast" noise from jet aircraft during start of take-off has exceedingly more intense components in the low frequencies (below ~315 Hz) than are present in the noise after take-off. As a result, people in residential areas about two miles behind certain take-off runways of the San Francisco Airport report feelings of annoyance greater than is predicted by A weighted DNL, see Pearsons *et al.* (2000). Similar cases of the lack of accuracy of the A weighting for estimating the annoyance impact in residential areas of more general environmental noises have been reported (Goodfriend, 1991, 2006).

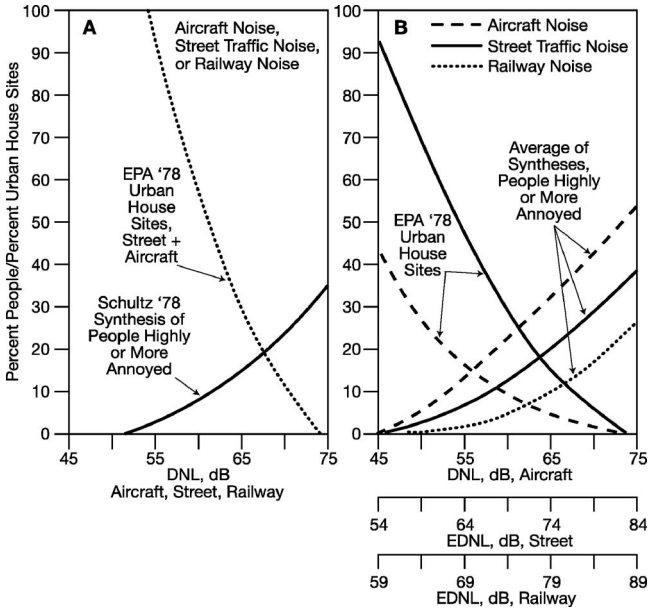


FIG. 7. Section A: Showing, as function of DNL: (1) percent of residents highly or more annoyed by street traffic, railway or aircraft noise (solid curve) according to syntheses by Schultz and cohorts, and (2) sum of percent of USA. urban house sites thusly exposed to street and aircraft traffic (dotted curve) according to Fig. 6, EPA. Section B: Showing as function of DNL: (1) percent of residents highly or more annoyed by aircraft (dashed curve) and, as function of EDNL, street traffic (solid curve) and railway (dotted curve) noise according to average of syntheses other than those of Schultz and cohorts, and (2) percent of USA. urban house sites thusly exposed to street (solid curve) and aircraft (dashed curve) traffic.

V. EDNL NOISE LEVELS COMPATIBILE WITH RESIDENTAIL LAND USAGE

Schultz (1978) prepared a "decision-maker's tool" for comparing the "costs" of: (a) increases above 15% of exposed populations being highly or more annoyed by aircraft, street traffic or railway noise versus (b) the loss in value of urban land that is identified as being in areas where environmental noise levels exceed limits deemed incompatible with typical residential living. The Federal Interagency Committee on Urban Noise (1980) and Federal Interagency Committee on Noise (1992) recommended criterion of incompatibility for residential-land usage is that no more than 15% of the people should be highly or more annoyed by transportation noise. FICUN and FICON do not reveal a criterion of what percentage of transportation noise-incompatible housing sites might be considered as acceptable by governmental authorities.

Schultz's 'tool-plot,' see Fig. 7(a), provides similar predictions for residential land-use zoning in regards to aircraft, street traffic, and railway noise. Presently available research data plotted on the same coordinates show differences between the different sources of noise, see Fig. 7(b).

Because of the averaging of the data for the subject three sources of noise, the Schultz curve: (a) reduces the apparent prevalence of people highly or more annoyed by aircraft noise by averaging it with the lesser percentage-effective exposure level of street and railway traffic noise as a cause of

TABLE IX. Annoyance and housing-sites impacts per DNL, EDNL, and EDNLm.

	Exposure Level			Percent people highly or more annoyed			
	DNL[a]	EDbNL[b]	EDNLm[c]	DNL[a]	EDbNL[b]	EDNLm[c]	EDNL(m)-DNL
Vehicle	dB	dB	dB	Percent	Percent	Percent	Percent
Aircraft	65	65	63	15%	35%	35%	20%
Street	65	74	72	15%	20%	20%	5%
Railway	65	79	77	15%	10%	10%	−5%
				Percent urban sites incompatible housing			
Aircraft	65	65	63	30%	4%	4%	−26%
Street	65	74	72	30%	15%	15%	−15%
Railway	65	79	77	30%	No data available		

[a]Data from Fig. 7A.
[b]Data from Fig. 7B.
[c]EDNLm = Annualized monthly av. EDNL.

annoyance, and (b) conversely increases the apparent percentage of urban sites to be considered as incompatible for residential housing because of aircraft noise. An additional reason for the lesser differences of noise condemnation of housing from street traffic noise compared to aircraft noise for a given DNL exposure level, is that by a factor of approximately 4, there are more urban house sites exposed to street traffic than to aircraft noise, see Fig. 6 in EPA report.

Table IX shows that the DNL (Schultz curve, Fig. 7(a)) leads, in comparison to the EDNL functions (Fig. 7(b)) to: (a) the underestimation by 20% points the percentage of people to be highly or more annoyed by aircraft noise, 5% points by street traffic noise and −5% points by railway noise; (b) overestimation, by 26% points, the percentage of urban sites to be considered as incompatible for residential housing because of aircraft noise, and 15% points by street traffic noise.

On-site attenuation data, in conjunction with state-of-art analyses of surveys of noise annoyance and associated effects, indicate that to meet the Federal Interagency Committee on Urban Noise (1980) and the Federal Interagency Committee on Noise (1992) criterion of compatibility in residential areas the maximum allowable exposure levels would be:

(a) Aircraft noise, 58 dB DNL (annualized), 56 dB DNL, (monthly average) for moderate climates;

(b) Street traffic noise, 66 dB EDNL (annualized), 64 dB EDNL (monthly average) for moderate climates;

(c) Railway noise, 72 dB EDNL (annualized), 70 dB EDNL (monthly average) for moderate climates.

VI. SUMMARY AND CONCLUSIONS

1. On-site acoustical measurements of aircraft noise and of street traffic noise in the yards and indoor rooms of one-story typical houses revealed that:

(a) Between the closest- and farthest-from source house facades, the outdoor levels were on average attenuated by: (a) 18 dB SENEL, 19 MaxdB for street traffic noise, and (b) 3 dB SENEL, 2 MaxdB for aircraft noise.

(b) Between the closest-to source house façade and the indoor room locations, the outdoor levels were on average attenuated by: (a) 27 dB SENEL, 27 MaxdB for street traffic noise; and (b) 24 dB SENEL, 24 MaxdB for aircraft noise.

2. For the assessment of on-site attenuation purposes, it is concluded that attenuation values in either dB SENEL or MaxdB will, on average, differ by <1 dB.

3. Application of acoustical-principled mathematical formulations of sound attenuations to be expected from the on-site physical conditions present for the measurements made in the present study revealed that compared to aircraft noise street traffic noise is on average ~9 dB, and railway noise ~14 dB less attenuated, and railway compared to street traffic noise, is ~4 dB less attenuated.

4. Excluding those of Schultz, Fidell *et al.* (1991) and Finegold *et al.* (1994), major studies and syntheses of surveys of annoyance from equal DNL exposures show an average measured difference of ~8 dB DNL between aircraft and street traffic noise, ~13 dB DNL between aircraft and railway noise, and ~10 dB between railway and street traffic noise.

5. Laboratory and field research findings of annoyance and associated emotion-related effects are uniformly congruent with measured and calculated house attenuation of street traffic noise compared to aircraft noise.

6. It is concluded that the implicit premise of the DNL methodology for the measurement or estimation of environmental noise dosage that on-site attenuations, if any, of environmental noises are equal, or of no significance, is in error.

7. The differences found in that regard from acoustical theory, formula and measurements are proportional to the subjectively scaled attitudes of noise annoyance and associated adverse nonauditory effects.

8. A new metric, EDNL, is proposed for estimating the percentages of people reporting "high, or more, annoyance" and other adverse effects from exposures to street traffic and railway noise in comparison to percentages from DNL exposures to aircraft noise.

9. A calculated "test" of EDNL, street traffic, railway and aircraft noise for predicting noise-annoyance surveys in moderate-climate urban areas indicated that, on average, the respondents in such areas spend approximately, ~30% of the day time outdoors and ~70% indoors of their homes on an annualized basis. It is estimated that the percentage of indoor listening will be significantly greater than 70% during the colder months of the year.

10. A decision maker's tool proposed by Schultz predicts, using DNL exposure data, that a smaller percentage of people will be highly or more annoyed, and a greater number of urban residential building sites will be noise condemnable due to aircraft noise than by street and railway traffic noise. The reverse relations are predicted by findings for exposure to street traffic and railway noise based on EDNL.

11. Further data are needed on the on-site noise attenuations characteristic of housing in typical urban neighborhoods exposed to aircraft, street traffic, railway and other environmental noises.

ACKNOWLEDGMENTS

The author is grateful to Jose Ortega and the Paul S. Veneklasen Research Foundation, Santa Monica, California, for funds to have recordings of aircraft and street traffic noise made and analyzed by Leo Amesuca of Veneklasen Associates, Santa Monica, California. The author is deeply indebted to Gilles Daigle, Canadian Research Council, Institute of Microstructural Science, Ottawa, Canada, for tutorial counsel on the use of mathematical formulas for the calculation of attenuations of environmental noises; any misuses herein are to be attributed to the author.

Andersen, T. V., Kuhl, K., and Restler, E. (**1983**). "Reactions to railway noise in Denmark," J. Sound Vib. **87**(2), 311–314.

Babisch, W. (**1998**). "Epidemiological studies of cardiovascular effects of traffic noise," In *7th International Congress on Noise as a Public Health Problem*, Noise Effects '98. P.T.Y. LTD, Sydney, Australia.

Beranek, L. L., Kryter, K. D., and Miller, L. N. (**1959**). "Reaction of people to exterior aircraft noise," Noise Control **5**(5), 23–31.

Committee on the Problem of Noise (1963). Noise Final Report, HM Stationery Office, London.

Commission on the Third London Airport (**1971**). Papers and Proceedings, Vol. **VII**, HM Stationery Office, London.

Environmental Protection Agency (**1978**). Protective Noise Levels, Condensed Version of EPA Levels. Doc. Report EPA 550/9-79-100. U.S. Environ. Protect. Agency, Wash., D.C. (Avail. From NTTS PB 82 138 827).

Federal Aviation Administration (**1991**). Section 91.119 of Title 14, Code of Regulation: Minimum safe altitudes, General, Amendment, Sept. 1993, Washington, D.C.

Federal Interagency Committee on Urban Noise (FICUN) (**1980**). Guidelines for Considering Noise in Land Use Planning and Control. Washington, D.C.

Federal Interagency Committee on Noise (FICON) (**1992**). Federal Agency Review of Selected Airport Noise Analysis Issues. Washington, D.C.

Fidell, S., Horornjeff, R., Mills, J., Baldwin, E., Teffeteller, S., and Pearsons, K. S. (**1981**). Community Sensitivity to Changes in Aircraft Noise Exposure, NASA Contractor Report 3490. NASA Langley Research Center, Langley, Virginia.

Fidell, S., Barber, D. S., Schultz, T. J. (**1991**). "Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233.

Fidell, S. (**2003**). "The Schultz curve 25 years later: A research perspective," J. Acoust. Soc. Am. **114**(8), Pt. 1, 3008–3015.

Fields, J. M., and Walker, J. G. (**1980**). "Comparing reactions to transportation noise from different surveys: A railway noise vs. aircraft and road traffic comparison," in *Noise as a Public Health Problem*, edited by J. V. Tobias, G. Jansen, and W. D. Ward, ASHA Report No. 10. Am. Speech-Lang.-Hearing. Assoc., Rockville, MD.

Fields, J. M., and Walker, J. G. (**1982**). "The response to railway noise in residential areas in Great Britian," J. Sound Vib. **85**(2), 177–255.

Finegold, L. S., Harris, S., and von Gierke, H. (**1994**). "Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people," Noise Control Eng. J. **42**, 25–30.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**(1), 47–65.

Goodfriend, L. S. (**1991**). "Re-evaluating the use of the A-weighted sound level meter for environmental noise impact assessment," in *Noise Control Proceedings*, 1991 National Conference on Noise Control Engineering: Twenty Years of Progress and Future Trends, Noise Control Foundation, New York.

Goodfriend, L. S. (**2006**). "Inappropriateness of the A-weighted level for environmental assessment," J. Sound Vib. **40**(6), 5–6.

Hall, F. L., Birnie, S. E., Taylor, S. M., and Palmer, J. E. (**1981**). "Direct comparison of community response to road traffic noise and to aircraft noise," J. Acoust. Soc. Am. **70**, 1690–1698.

Job, R. F. S. (**1988**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.

Knipschild, P. (**1980**). Aircraft noise and hypertension in Noise as a Public Health Problem, edited by J. V. Tobias, G. Jansen, and W. D. Ward, ASHA Report No. 10. Am. Speech-Lang.-Hearing. Assoc., Rockville, MD.

Kryter, K. D., and Pearsons, K. S. (**1963**). "Some effects of spectral content and duration on perceived noise level," J. Acoust. Soc. Am. **35**, 866–883.

Kryter, K. D. (**1970**). Possible Modifications to the Calculation of Perceived Noisiness. NASA CR 1936, NASA Langley Research Center, Hampton, VA.

Kryter, K. D. (**1982**). "Community annoyance from aircraft and ground vehicle noise," J. Acoust. Soc. Am. **72**, 1222–1242.

Kryter, K. D. (**1983**). "Rebuttal by Karl D. Kryter to Comments by T. J. Schultz," J. Acoust. Soc. Am. **72**, 1253–1257.

Kryter, K. D. (**1985**). *The Effects of Noise on Man*, 2nd ed. (Academic, New York).

Kryter, K. D. (**1994**). *The Handbook of Hearing and the Effects of Noise* (Academic, New York).

Kurze, U., and Beranek, L. L. (**1988**). Sound Propagation Outdoors, in *Noise and Vibration Control*, edited by L. L. Beranek (Institute of Noise Control Engineering, Washington, D.C.), Chap. 7.

Lercher, P., and Kofler, W. (**1993**). "Adaptive behavior to road traffic noise in Noise as a Public Health Problem," *Proceedings of the Sixth International Congress I'NRETS*, 94114 Arcuil Cedex, France, Vol. **2**, pp. 465–468.

McKennell, A. (**1963**). Psycho-Social Factors in Aircraft Noise Annoyance, International Congress on Noise as a Public Health Question. Dubrovnik. EPA Report No. 550/9-73-008 (Environmental Protection Agency, Washington, D.C.).

Miedema, H. (**1993**). Response functions for environmental noise in Noise as a Public Health, *Proceedings of the Sixth International Congress I'NRETS*, 94114, Arcuil Cedex, France, Vol. **3**, pp. 428–433.

Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.

Miedema, H. M. E., and Oudshoorn, C. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," J. National Institute of Environmental Health Sciences **109**, 409–416.

Miedema, H. M. E., Fields, J. M., and Vos, H. (**2005**). "Effect of season and meteorological conditions on community noise annoyance," J. Acoust. Soc. Am. **115**, 2853–2865.

Mieszkowski, P., and Saper, A. M. (**1978**). "An estimate of the effects of airport noise on property values," J. Urban Econ. **5**, 425–440.

National Academy of Sciences (**1977**). Guidelines for Preparing Environmental Impact Statements on Noise, Committee on Hearing and Bioacoustics, Washington, D.C.

Nelson, J. P. (**1979**). Aircraft Noise and the Market for Residential: Empirical Results for Seven Selected Airports. Rep. No. DOT/RSPA/DPB-50/78, Sept. 1978. (Available NTIS as PB 297 681).

Noise Control Act of 1972 (**1972**). Public Law 92-574, 92 Congress, HR 11021, October 27, **1972**.

Ohrstrom, E. (**2004**). "Longitudinal surveys on effects of changes in road traffic noise—annoyance, activity disturbances, and psycho-social well-being," J. Acoust. Soc. Am. **115**(2), 719–729.

Ollerhead, J. B. (**1980**). *Accounting time of day and mixed source effects in the assessment of community noise exposure in Noise as a Public Health Problem*, edited by J. V. Tobias, G. Jansen, W. D. Ward, ASHA Rep. No. 10. Am. Speech-Lang.-Hearing. Assoc., Rockville, MD.

Ortega, J. C., and Kryter, K. D. (**1982**). "Comparison of aircraft and ground vehicle noise levels in front and backyards of residences," J. Acoust. Soc. Am. **71**, 216–217.

Pearsons, K. S. (**1966**). The Effects of Duration and Background Noise Level on Perceived Noisiness. ADS-78. Fed. Aviation Admin., Washington, D.C.

Pearsons, K., Fidell, S., Silvati, L., Sneddon, M., and Howe, R. (**2000**). Study of the Levels, Annoyance and Potential Mitigation of Backblast Noise at San Francisco International Airport. BBN Report No. 8257, Cambridge, MA.

Piercy, J. E., and Daigle, D. A. (**1991**). "Sound propagation in the open air," in *Handbook of Acoustical Measurements and Noise Control*, C. M. Harris, Ed. (Acoustical Society of America, Melville, NY), Chap. 3.

Schomer, P. (**2002**). "On normalizing DNL to provide better correlation with response," J. Sound Vib. **36**(12), 14–23.

Schultz, T. J. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.

Schultz, T. J. (**1982**). Comments of K.D. Kryter's paper, "Community annoyance from aircraft and ground vehicle noise," J. Acoust. Soc. Am. **72**, 1243–1253.

Noise control in enclosures: Modeling and experiments with T-shaped acoustic resonators

D. Li, L. Cheng,[a] and G. H. Yu
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, SAR of China

J. S. Vipperman
Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

(Received 10 January 2007; revised 13 August 2007; accepted 14 August 2007)

This paper presents a theoretical and experimental study of noise control in enclosures using a T-shaped acoustic resonator array. A general model with multiple resonators is developed to predict the acoustic performance of small resonators placed in an acoustic enclosure. Analytical solutions for the sound pressure inside the enclosure and the volume velocity source strength out of the resonator aperture are derived when a single resonator is installed, which provides insight into the physics of acoustic interaction between the enclosure and the resonator. Based on the understanding of the coupling between the individual resonators and enclosure modes, both targeted and nontargeted, a sequential design methodology is proposed for noise control in the enclosure using an array of acoustic resonators. Design examples are given to illustrate the control performance at a specific or at several resonance peaks within a frequency band of interest. Experiments are conducted to systematically validate the theory and the design method. The agreement between the theoretical and experimental results shows that, with the help of the presented theory and design methodology, either single or multiple resonance peaks of the enclosure can be successfully controlled using an optimally located acoustic resonator array. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783122]

PACS number(s): 43.50.Gf, 43.50.Jh, 43.20.Ks [KA] Pages: 2615–2625

I. INTRODUCTION

Enclosures are widely used in civil and industrial systems. Typical examples, such as cabinets of ground vehicles, nacelles of helicopters, and payload fairings of launch vehicles, exhibit a commonality of having relatively compact interior dimensions. Low-frequency noise problems in such small enclosures are difficult to deal with since only compact noise control devices can satisfy the spatial limitations. When properly designed, a passive method using an acoustic resonator can effectively absorb acoustic energy from a targeted acoustic mode.[1–9] Ideally, it is desirable to integrate resonators into the host structure to save space, particularly for small enclosures. However, it is difficult to meet this requirement with a classical Helmholtz resonator because of its bulbous profile. In order to overcome the drawbacks of conventional resonators, considerable effort has been devoted to developing new ideas and solutions.[10] A good example is the recent work by Li and Vipperman,[7,8,11] who developed a multimodal design theory of a long T-shaped acoustic resonator (TAR) for the low-frequency noise transmission control in an expendable launch vehicle (ELV) payload fairing. A TAR consists of two mutually perpendicular tubes: a long closed-end tube and a short open-end tube. Compared with conventional Helmholtz resonators, one of the biggest advantages offered by the TAR is its large aspect ratio. This unique feature makes it possible to integrate a TAR into the host structure as a structural component, such as a beam, a pillar, or an enhanced rib, which relaxes the space requirement in the implementation. In the previous work, a TAR array has been structurally integrated into an ELV payload fairing, resulting in a 6–9 dB noise reduction in noise transmission control.[7,8]

Various theoretical and experimental studies on the acoustic interaction between an enclosure and conventional Helmhotz resonators have been reported in the literature. Fahy and Schofield presented a theoretical model involving acoustic interaction between an enclosure and a single Helmholtz resonator,[12] which provides early explicit insight into the fundamental physics of the acoustic interaction. Later, Cummings developed a multimodal theory to deal with the acoustic coupling problem when introducing a Helmholtz or quarter-wave resonator array into the enclosure.[13] Each resonator was modeled as a secondary point source with an unknown volume velocity. The problem was approached by equalizing the pressure at each resonator aperture, which was computed from knowledge of the acoustic impedance of the resonators, and the pressure at the same location, which was determined from the radiation of the primary and secondary sound sources. Thus, a singularity problem was reported when calculating the pressure at the resonator aperture using the volume velocity of the resonator itself.[13] In order to solve the problem, an average sound pressure at the surface of an equivalent small pulsating sphere was used as the pressure at

[a]Electronic mail: mmlcheng@polyu.edu.hk

the resonator aperture. This approximation can prevent the singularity problem but at the expense of introducing other problems. For example, it was found that the coupled frequencies predicted from the sphere-based model were different from the measured ones. Recently, a vibroacoustic coupling model was presented by Estève and Johnson to control noise transmitted into a composite cylinder using distributed vibration absorbers and Helmholtz resonators.[2,3] Noise control using long cylindrical tube-shaped resonators for fairing noise control has also been investigated.[4,5]

Despite all the effort, the design of acoustic resonators is still limited by a number of factors. The first one is the lack of theoretical modeling tools to facilitate and optimize the design of such devices for a given enclosure, which actually still involves intensive experimental measurements on a trial-and-error basis. Moreover, should more resonators be used to form a resonator array to increase the control performance inside a small enclosure, an effective design cannot be achieved without considering the acoustic coupling between the resonators and all acoustic modes of the enclosure, as suggested by the earlier observation on conventional Helmholtz resonators.[12]

This paper attempts to provide a theoretical model to predict the acoustic performance of multiple resonators placed in an acoustic enclosure, along with a reliable design methodology to maximize the control performance. In a broader perspective, the general model is applicable not only to the TAR, but also to classical Helmholtz resonators. The paper is divided into four sections. The general model is first presented in Sec. II. The formulation is based on the modal expansion of pressure in the enclosure and its coupling to the resonators. Acoustic output impedance is used to package the dynamic information of the TAR, which greatly simplifies the modeling process. A special case (with one resonator installed) is selected for detailed analysis, and a sequential design method (SDM) is also proposed in this section. Numerical simulations and experimental validations are performed in Sec. III. A detailed implementation procedure using a TAR array to control a single or multiple resonance peaks is demonstrated and discussed. Experimental validations are systematically carried out for each important stage of development, including the evaluation of dissipation characteristics of resonators, the performance assessment of the SDM with multiple resonators, and the validation of optimal locations of resonators. Finally, conclusions are drawn in the last section.

II. THEORY

In the following, a general model considering acoustic interaction between an enclosure and an acoustic resonator array is developed before the special case involving a single resonator is discussed. Throughout the paper, the superscripts and subscripts E, R, and S stand for variables associated with "enclosure," "resonator," and "primary source," respectively.

A. Acoustic interaction between enclosure and acoustic resonator array

The inhomogeneous wave equation governing the pressure fields inside the enclosure is

$$\nabla^2\phi(\mathbf{r},t) - \frac{1}{c^2}\ddot{\phi}(\mathbf{r},t) = q(\mathbf{r},t), \tag{1}$$

where $\phi(\mathbf{r},t)$ is the acoustic velocity potential; c the sound speed; and q the volume velocity source strength density distribution within the volume or on the surface of the enclosure. Assuming that a set of N harmonic sources with volume velocity source strength density $q_1^S, q_2^S, \ldots, q_N^S$ located at the points $\mathbf{r}_1^S, \mathbf{r}_2^S, \ldots, \mathbf{r}_N^S$ form the primary sound field, and M resonators with volume velocity source strength density $q_1^R, q_2^R, \ldots, q_M^R$ located at the points $\mathbf{r}_1^R, \mathbf{r}_2^R, \ldots, \mathbf{r}_M^R$ (centers of the resonator apertures) form the secondary sound field in the enclosure, Eq. (1) becomes

$$\nabla^2\phi(\mathbf{r},t) - \frac{1}{c^2}\ddot{\phi}(\mathbf{r},t) = \sum_{m=1}^{M} q_m^R(t)\,\delta(\mathbf{r}-\mathbf{r}_m^R) + \sum_{n=1}^{N} q_n^S(t)\,\delta(\mathbf{r}-\mathbf{r}_n^S), \tag{2}$$

where $\delta(\mathbf{r}-\mathbf{r}_0)$ is a three-dimensional Dirac delta function. Notice that the volume velocity out of the resonator has the same sign as that of the primary sound source, i.e., the positive sign is directed out of the source into the enclosure, which results in an opposite sign as compared to that defined in many textbooks.[14,15] The volume velocity source strength density out of the acoustic resonator can be computed from

$$Z_m = \frac{p(\mathbf{r},t)\,\delta(\mathbf{r}-\mathbf{r}_m^R)}{q_m^R(t)}, \tag{3}$$

where Z_m is defined as the acoustic output impedance at the aperture of the mth resonator, and $p(\mathbf{r}_m^R,t)$ is the sound pressure at $\mathbf{r}_m^R$ provided the largest dimension of the resonator aperture is smaller than the sound wavelength of interest. Substituting Eq. (3) into Eq. (2) and using $p(\mathbf{r},t) = -\rho_0\dot{\phi}(\mathbf{r},t)$, Eq. (2) becomes

$$\nabla^2\phi(\mathbf{r},t) - \frac{1}{c^2}\ddot{\phi}(\mathbf{r},t) = -\rho_0\sum_{m=1}^{M}\frac{\dot{\phi}(\mathbf{r},t)\,\delta(\mathbf{r}-\mathbf{r}_m^R)}{Z_m} + \sum_{n=1}^{N} q_n^S(t)\,\delta(\mathbf{r}-\mathbf{r}_n^S). \tag{4}$$

Acoustic velocity potential $\phi(\mathbf{r},t)$ can be decomposed on the basis of the respective mode-shape functions of the enclosure: $\phi(\mathbf{r},t) = \Sigma\psi_j(t)\varphi_j(\mathbf{r})$, where $\psi_j(t)$ is the jth modal response and $\varphi_j(\mathbf{r})$ is the jth eigenfunction given in Ref. 13. Substituting this modal expansion into Eq. (4) and applying orthogonality properties of the eigenfunctions yields an uncoupled acoustic equation,

$$\ddot{\psi}_j(t) - \frac{cz_0}{V^E}\sum_{h=1}^{J}\left\{\sum_{m=1}^{M}\left[\frac{1}{Z_m}\frac{\varphi_j(\mathbf{r}_m^R)\varphi_h(\mathbf{r}_m^R)}{\Lambda_j}\right]\dot{\psi}_h(t)\right\} + (\gamma_j^E)^2\psi_j(t) = -\frac{c^2}{V^E}\sum_{n=1}^{N}\frac{\tilde{\varphi}_j(\mathbf{r}_n^S)}{\Lambda_j}q_n^S(t), \tag{5}$$

where j and $h=1,2,3,\ldots,J$ are the indices of the modal response, $z_0=\rho_0 c$ is the characteristic acoustic impedance of the fluid, V^E is the volume of the enclosure, $\Lambda_j = \int_{V^E}[\varphi_j(\mathbf{r})]^2 dV/V^E$ is the modal normalization factor, $\tilde{\varphi}_j(\mathbf{r}_n^S)$ is the averaged $\varphi_j(\mathbf{r}_n^S)$ over the volume of the nth source, and γ_j^E is the jth complex eigenvalue of the enclosures, expressed as $\gamma_j^E=\omega_j^E+iC_j^E$, in which the real part is the angular frequency and the imaginary part is an equivalent *ad hoc* damping coefficient.

Assuming all time-dependent variables are harmonic, i.e., $\psi_j(t)=P_j e^{i\omega t}$ and $q_n^S(t)=Q_n^S e^{i\omega t}$, Eq. (5) becomes

$$\left[\frac{\omega^2-(\gamma_j^E)^2}{\omega^2}+i\frac{z_0}{kV^E}\sum_{m=1}^{M}\frac{1}{Z_m}\frac{[\varphi_j(\mathbf{r}_m^R)]^2}{\Lambda_j}\right]\frac{P_j}{\frac{Q^s}{k^2V^E}} + i\frac{z_0}{kV^E}\sum_{h\neq j}^{J}\left[\sum_{m=1}^{M}\frac{1}{Z_m}\frac{\varphi_j(\mathbf{r}_m^R)\varphi_h(\mathbf{r}_m^R)}{\Lambda_j}\right]\frac{P_h}{\frac{Q^s}{k^2V^E}} = \sum_{n=1}^{N}\frac{\tilde{\varphi}_j(\mathbf{r}_n^S)}{\Lambda_j}\frac{Q_n^S}{Q^s}, \tag{6}$$

where Q^S is the volume velocity source strength of any point source, and k is the wave number. When a finite number of enclosure modes are considered, Eq. (6) forms a set of linear equations. The set of linear equations can be numerically solved with the modal response coefficients $P_j/(Q^S/k^2V^E)$ as unknowns if the eigenfunctions and eigenvalues of the enclosure are given. A dimensionless amplitude of sound pressure inside the enclosure can be then computed from $p(\mathbf{r},t)=-\rho_0\dot{\phi}(\mathbf{r},t)$ as

$$\frac{p(\mathbf{r})}{\frac{i\omega\rho_0 Q^S}{k^2V^E}} = -\sum_{j=1}^{J}\left[\varphi_j(\mathbf{r})\left(\frac{P_j}{\frac{Q^S}{k^2V^E}}\right)\right]. \tag{7}$$

A dimensionless sound pressure level (SPL) is used to evaluate the pressure distribution inside the enclosure as

$$L_p(\mathbf{r}) = 20\log\left|\frac{p(\mathbf{r})}{\frac{i\omega\rho_0 Q^S}{k^2V^E}}\right|. \tag{8}$$

B. Acoustic interaction between an enclosure and a single TAR

In a previous study,[12] the targeted enclosure mode by a single resonator was assumed to be well separated in frequency from other neighboring modes, such that only the targeted mode was taken into account in the analysis of acoustic interaction. For enclosures with higher modal density, however, the neighboring acoustic modes are not well separated, such that a general model accounting for all enclosure modes is needed. In the following, such a model, namely the acoustic interaction between one single resonator and the enclosure with multiple modes, is introduced. This analysis results in analytical solutions of the sound pressure inside the enclosure and the volume velocity source strength from the resonator aperture, which provide insight into the fundamental physics of the acoustic interaction between the primary and the secondary sound sources.

With one acoustic resonator and N point sources in the enclosure, the acoustic equation in the enclosure can be obtained from Eq. (6) after eliminating subscript m,

$$[\omega^2-(\gamma_j^E)^2]P_j + i\frac{cz_0\omega}{V^E}\frac{1}{Z}\sum_{h=1}^{J}\frac{\varphi_j(\mathbf{r}^R)\varphi_h(\mathbf{r}^R)}{\Lambda_j}P_h = \frac{c^2}{V^E}\sum_{n=1}^{N}\frac{\tilde{\varphi}_j(\mathbf{r}_n^S)}{\Lambda_j}Q_n^S. \tag{9}$$

The modal response of P_j can be solved from Eq. (9) as

$$\frac{P_j}{\frac{Q^S}{k^2V^E}} = \underbrace{\frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\sum_{n=1}^{N}\frac{\tilde{\varphi}_j(\mathbf{r}_n^S)}{\Lambda_j}\frac{Q_n^S}{Q^S}}_{\text{Contribution of primary sound field}} - \underbrace{\frac{\left[\frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\frac{\varphi_j(\mathbf{r}^R)}{\Lambda_j}\right]\left(i\frac{z_0}{kZ}\frac{1}{V^E}\right)\sum_{h=1}^{J}\left[\frac{\omega^2}{\omega^2-(\gamma_h^E)^2}\frac{\varphi_h(\mathbf{r}^R)}{\Lambda_h}\sum_{n=1}^{N}\tilde{\varphi}_h(\mathbf{r}_n^S)\frac{Q_n^S}{Q^S}\right]}{1+i\frac{z_0}{kZ}\frac{1}{V^E}\sum_{h=1}^{J}\frac{\omega^2}{\omega^2-(\gamma_h^E)^2}\frac{[\varphi_h(\mathbf{r}^R)]^2}{\Lambda_h}}}_{\text{Contribution of acoustic resonator}}. \tag{10}$$

The first term on the right-hand side of Eq. (10) is the contribution of the primary sound field and the second term results from inserting an acoustic resonator into the enclosure. Note that a maximum $\varphi_j(\mathbf{r}^R)$ would not automatically warrant a maximum performance of the absorber. If the resonator is located in the node of the mode [i.e., $\varphi_j(\mathbf{r}^R)=0$], the value of the second term is systematically zero, suggesting zero contribution to the jth modal response. From Eq. (10) it can also be observed that the noise control is a result of the second part on the right-hand side of the equation, i.e., the acoustic interaction between the enclosure and the resonator. A strong acoustic interaction between the enclosure and the

resonator can provide good noise control, which depends on the location of $\mathbf{r}^R$ and the physical parameters of the resonator indicated in Z. Through optimizing the location and physical parameters of the resonator, we can maximize the acoustic interaction in the second term of Eq. (10), and thus obtain optimal noise control.

When only the targeted mode, i.e., the jth enclosure mode, is dominant, Eq. (10) can be simplified by ignoring all other minor enclosure modes as

$$\frac{P_j}{\frac{Q^S}{k^2V^E}} = \frac{1}{1+i\frac{z_0}{kZ}\frac{1}{V^E}\frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\frac{[\varphi_j(\mathbf{r}^R)]^2}{\Lambda_j}}\left(\frac{P_j}{\frac{Q^S}{k^2V^E}}\right)_{\text{without resonator}}, \tag{11}$$

where

$$\left(\frac{P_j}{\frac{Q^S}{k^2V^E}}\right)_{\text{without resonator}} = \frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\sum_{n=1}^{N}\frac{\tilde{\varphi}_j(\mathbf{r}_n^S)}{\Lambda_j}\frac{Q_n^S}{Q^S} \tag{12}$$

is the jth modal response of the enclosure in the absence of the resonators.

Equation (11) is consistent with that shown in Ref. 12. It is more clearly shown in Eq. (11) that when inserting an acoustic resonator at $\mathbf{r}^R$ that targets the jth enclosure mode, the jth modal response is reduced by the factor of

$$1/1+i\frac{z_0}{kZ}\frac{1}{V^E}\frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\frac{[\varphi_j(\mathbf{r}^R)]^2}{\Lambda_j}.$$

From this simplified case, it is seen that when the physical parameters of the resonator are already determined, the control result can be optimized only if the resonator is located at any of the antinodes of the targeted mode. However, the actual case is much more complex than the simplified one because the resonator is coupled to all enclosure modes, not just to the targeted or dominant one. This particular issue will be discussed later in this paper.

The volume velocity source strength Q^R directed outwards from the resonator into the enclosure can be derived based on Eq. (3), $p(\mathbf{r},t)=-\rho_0\dot{\phi}(\mathbf{r},t)$, $\phi(\mathbf{r},t)=\Sigma\psi_j(t)\varphi_j(\mathbf{r})$, and Eq. (10) as

$$\frac{Q^R}{Q^S} = -\frac{i\frac{z_0}{kZ}\frac{1}{V^E}\sum_{h=1}^{J}\left[\frac{\omega^2}{\omega^2-(\gamma_h^E)^2}\frac{\varphi_h(\mathbf{r}^R)}{\Lambda_h}\sum_{n=1}^{N}\tilde{\varphi}_h(\mathbf{r}_n^S)\frac{Q_n^S}{Q^S}\right]}{1+i\frac{z_0}{kZ}\frac{1}{V^E}\sum_{h=1}^{J}\frac{\omega^2}{\omega^2-(\gamma_h^E)^2}\frac{[\varphi_h(\mathbf{r}^R)]^2}{\Lambda_h}}. \tag{13}$$

It is seen that the volume velocity source strength out of the resonator aperture depends not only on the acoustic output impedance of the resonator itself, but also on all acoustic modes of the enclosure. In terms of volume velocity source strengths, the modal response shown in Eq. (10) can be rewritten as

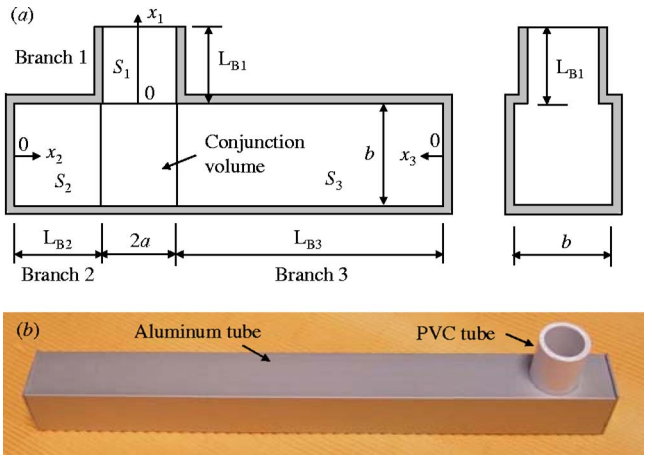


FIG. 1. (Color online) A typical T-shaped acoustic resonator: (a) Design geometries and (b) a fabricated TAR.

$$\frac{P_j}{\frac{Q^S}{k^2V^E}} = \underbrace{\frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\sum_{n=1}^{N}\left[\frac{\tilde{\varphi}_j(\mathbf{r}_n^S)}{\Lambda_j}\frac{Q_n^S}{Q^S}\right]}_{\text{Contribution of primary sound field}} + \underbrace{\frac{\omega^2}{\omega^2-(\gamma_j^E)^2}\frac{\varphi_j(\mathbf{r}^R)}{\Lambda_j}\frac{Q^R}{Q^S}}_{\text{Contribution of accoustic resonator}}. \tag{14}$$

Therefore, the analytical solution of the sound pressure inside the enclosure is represented as

$$\frac{p(\mathbf{r})}{\frac{i\rho_0\omega Q^S}{k^2V^E}} = \underbrace{-\sum_{h=1}^{J}\left[\frac{\omega^2}{\omega^2-(\gamma_h^E)^2}\frac{\varphi_h(\mathbf{r})}{\Lambda_h}\sum_{n=1}^{N}\tilde{\varphi}_h(\mathbf{r}_n^S)\frac{Q_n^S}{Q^S}\right]}_{\text{Contribution of primary sound field}} \underbrace{-\sum_{h=1}^{J}\left[\frac{\omega^2}{\omega^2-(\gamma_h^E)^2}\frac{\varphi_h(\mathbf{r})\varphi_h(\mathbf{r}^R)}{\Lambda_h}\frac{Q^R}{Q^S}\right]}_{\text{Contribution of acoustic resonator}}. \tag{15}$$

C. Acoustic output impedance of TAR

In the above-mentioned derivations, the acoustic output impedance of a resonator was used to characterize the dynamic behavior of the secondary sound source. In this section, the expression of the acoustic output impedance for T-shaped acoustic resonators is derived.

A typical TAR[7,8,11] is shown in Fig. 1. The resonator consists of three branches—1–3—as depicted in Fig. 1. Branch 1 is a circular tube having radius a and physical length L_{B1}, and Branches 2 and 3 are coaxial and have cross-sectional areas S_2 and S_3 (circular or rectangular) and physical lengths L_{B2} and L_{B3}, respectively. Branch 1 is perpendicular to Branches 2 and 3. It is assumed that the wave inside the TAR is a plane wave, which is satisfied in this study since the sound wavelength of interest is sufficiently larger than the largest cross-sectional dimension of the resonators. When ignoring the absorptive process within the fluid and at the walls of the resonator, the acoustic output impedance, Z, at the external aperture of the TAR can be obtained from Ref. 11 as

$$Z = iz_0 \frac{1 - \frac{S_2}{S_1}\tan(k_1L_1)\tan(k_2L_2) - \frac{S_3}{S_1}\tan(k_1L_1)\tan(k_3L_3)}{S_1\tan(k_1L_1) + S_2\tan(k_2L_2) + S_3\tan(k_3L_3)}, \tag{16}$$

where k_1, k_2, and k_3 are the wave numbers of three branches, composed of three tube segments (the three wave numbers are equal when the internal dissipation is not considered), and L_1, L_2, and L_3 are the effective lengths of Branch 1, Branch 2, and Branch 3, respectively, which can be computed using the physical lengths by adding end corrections. When the external aperture of Branch 1 is unflanged, the hybrid end corrections of Branches 1, 2, and 3 presented in Refs. 7 and 11 are used,

$$\Delta L_1 = \frac{8}{3\pi}a, \tag{17}$$

$$\Delta L_2 = 1.5\frac{8}{3\pi}a, \tag{18}$$

$$\Delta L_3 = 1.5\frac{8}{3\pi}a. \tag{19}$$

The absorptive process within the fluid and at the walls of a resonator is important because it provides a damping (internal resistance) mechanism for the resonator, which prevents the resonator from producing a mathematical singularity at its resonance frequencies. The absorptive process directly dissipates the input energy. The remaining, nondissipated energy is reradiated back to the enclosure from the resonator aperture as the secondary sound source, resulting in acoustic interaction with the primary source. When the internal resistance of the resonator is very small, the acoustic interaction between the enclosure and resonator sharply splits the targeted resonance peak of the enclosure into two parts, and the peak response is significantly attenuated within a very narrow frequency band.[4] By properly increasing the internal resistance of the coupled resonator mode, the working bandwidth of the resonator can be improved, while sacrificing some control performance due to decreased amplitude. However, if the resistance is too large, both the velocity amplitude of the fluid inside the resonator and the strength of volume velocity out of the resonator aperture become too small; such low velocities can only produce low dissipation in the resonators and slight acoustic interaction between the enclosure and resonators and hence result in insignificant control at the targeted enclosure resonances. A resonator with optimally designed internal resistance can provide a good control performance in a relatively large working frequency bandwidth, which will be addressed in another paper.

If the absorptive process within the fluid and at the walls of the resonator is considered, we can replace the lossless wave number k_i $(i=1,2,3)$ in Eq. (16) by a complex propagation constant $\mathbf{k}_i$, which can be approximately expressed with a dispersion relation[15]

$$\mathbf{k}_i = (k + \alpha_i) - i\alpha_i, \tag{20}$$

where $\alpha_i \ll k$ is the absorption coefficient of tube i defined in Ref. 15. For each tube segment, it can be computed by

$$\alpha_i = \sqrt{\frac{\omega\mu}{8cz_0}}\left(1 + \frac{\gamma - 1}{\sqrt{P_r}}\right)\frac{L_i^P}{S_i}, \tag{21}$$

where μ is the shear viscosity coefficient, γ is the specific heat ratio, $P_r = \mu C_p/\kappa$ is the Prandtl number of the medium, C_p is the specific heat at constant pressure, κ is the thermal conduction coefficient, and L_i^P and S_i are the inner-perimeter and cross-sectional area of tube i, respectively.

When considering absorption, the tangent function used in Eq. (16) is extended as[14]

$$\tan[(k + \alpha_i) - i\alpha_i]L_i = \frac{\sin^2(k + \alpha_i)L_i\cosh^2\alpha_iL_i + \cos^2(k + \alpha_i)L_i\sinh^2\alpha_iL_i}{\cos(k + \alpha_i)L_i\sin(k + \alpha_i)L_i + i\sinh\alpha_iL_i\cosh\alpha_iL_i}. \tag{22}$$

D. Sequential design method using acoustic resonators

When using acoustic resonators to control noise in a small enclosure, a SDM is proposed as follows.

(1) A TAR to target the desired enclosure mode at $\omega^R = ck^R$ is designed and fabricated using[11]

$$L_3 = \frac{\operatorname{atan}\left(\frac{S_1}{S_3}\cot(k^RL_1) - \frac{S_2}{S_3}\tan(k^RL_2)\right) + (i-1)\pi}{k^R}, \tag{23}$$

where L_1, L_2, S_1, S_2, and S_3 are given, and $i(=1,2,3,4,\ldots)$ are the harmonics of the acoustic resonator. Any resonator harmonics can be used for control.

(2) The optimal location of TAR is determined while accounting for multiple enclosure modes.

When there is a single resonator inside the enclosure and only the targeted enclosure mode φ_j is dominant, Eq. (11), it is known that a maximal noise control can be obtained only if the resonator is located at any antinode of the targeted mode.[12] However, this result is not true when multiple modes are taken into account in the acoustic coupling model. Actually, the resonator does couple with all acoustic modes of the enclosure.[12] Therefore, there exists an optimal location for the resonator among antinodal points, in which the best control performance of the resonator can be obtained. For controlling sound pressure at a specific point, the optimal location can be determined through comparing the SPL obtained from different resonator locations. The optimal location corresponds to the largest SPL reduction in the vicinity of the targeted frequency.

(3) After coupling the lightly damped TAR with the lightly damped enclosure, two new resonance peaks appear to replace the original peak, and a frequency shift at other nontargeted resonances also occurs. In order to improve the control performance, new resonators can be added to tackle the newly emerged peaks by repeating steps (1) and (2).

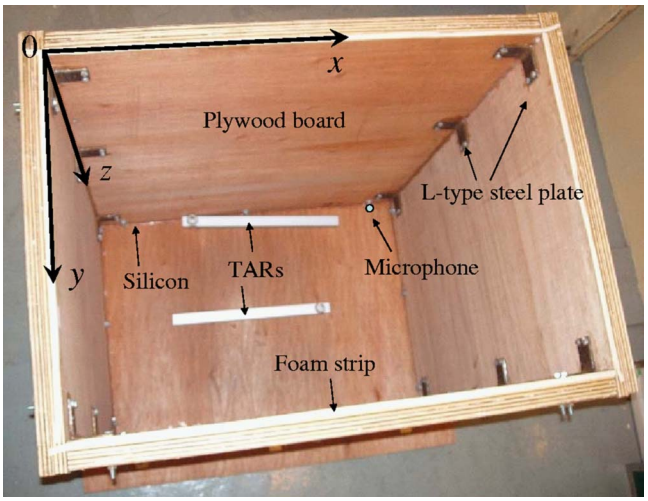


FIG. 2. (Color online) Right parallepiped reverberant enclosure with a coordinate system.

Theoretically, provided that enough space is available to host additional resonators, the sound pressure level at the targeted resonance can be reduced to a desired level after several iterations.

III. COMPARISONS BETWEEN THEORY AND EXPERIMENT

The noise control of an enclosure using multiple TARs was numerically investigated and experimentally validated. To this end, one right parallelpiped reverberant enclosure having geometric dimensions $l_x=0.976$ m, $l_y=0.695$ m, and $l_z=1.188$ m was used and shown in Fig. 2. The enclosure was fabricated using six plywood boards assembled with L-type steel plates, bolts, and nuts (see Fig. 2). Each of the six boards was comprised of two 1-in. boards, which were tightly screwed together. Silicon sealant and foam strips were also applied to prevent air leakage from the enclosure. The upper enclosure wall was removable and an enclosed loudspeaker was located on the upper surface at (100,59, −51) mm to excite the primary sound field in the enclosure through a square hole of 100×100 mm. The SPL at an arbitrary point (816,70,1028) mm in the enclosure was predicted and measured. In all simulations, a total of 216 enclosure modes were used [$(l,m,n)=(0\sim5,0\sim5,0\sim5)$, where l,m,n are the node number in x, y, and z direction, respectively.]. The eigenfunctions $\varphi_j(\mathbf{r}^R)$ of the enclosure accounting for the thermalviscous boundary conditions were obtained from Ref. 13, and the eigenvalues $\omega_j^E=\omega_j+iC_j$ were assembled using the undamped natural frequency of the jth enclosure mode ω_j and the jth *ad hoc* damping coefficient $C_j=\omega_j/2Q_j$, obtained from the measured quality or Q factors listed in Table I. All physical parameters of air, tabulated in Table II, were used in the simulations. The T-shaped acoustic resonators used in the simulations and experiments had a circular cross-sectional Branch 1, 21 mm in internal diameter, and square cross-sectional Braches 2 and 3 that were 29.5×29.5 mm. The physical lengths of Branch 1 and Branch 2 were $L_{B1}=30$ mm and $L_{B2}=20$ mm, respectively, and the physical length of Branch 3 was calculated with Eq. (23). In this study, the desired working frequency of all TARs was chosen to be their fundamental or Helmholtz frequency [$i=1$ in Eq. (23)]. No damping materials were used inside the resonator, and only the thermalviscous boundary conditions were considered.

TABLE I. Computed natural frequencies and measured Q factors of the enclosure.

Index	Mode number (lmn)	Natural frequency (Hz)	Q factor
1	000	0	⋯
2	001	145.0	32
3	100	176.5	42
4	101	228.4	44
5	010	247.8	46
6	011	287.1	84
7	002	290.0	42
8	110	304.3	51
9	111	337.0	79
10	102	339.5	60
11	200	353.0	80
12	012	381.5	60
13	201	381.6	45
⋯	⋯	>400.0	45

For the simulations, the convergence of the model size was achieved by examining the volume velocity at the resonator aperture of TARs and the pressure inside the enclosure as the number of enclosure modes was gradually increasing. When the number of the enclosure modes was varied from 216 ($l=m=n=0-5$) to 1000 ($l=m=n=0-9$), neither the volume velocity nor the pressure at (816,70,1028) mm in the enclosure, as predicted by Eq. (6), were found to change (not shown). As a result, 216 enclosure modes were determined to be for the model size.

A. Evaluation of dissipation characteristics of resonators

Equation (16) was used to calculate the acoustic impedance of TARs, based on the absorption coefficient of each tube segment using Eq. (21). It is known that the damping of a resonator is an important parameter, which significantly affects both the coupled amplitude at the targeted resonance peak of the enclosure and the working frequency bandwidth of the resonator itself. Therefore, the validity of Eq. (21)

TABLE II. Air properties.

Physical parameter	Used value
Ambient temperature, T (°C)	20
Humidity (%)	90
Speed of sound, c (m/s)	344.5
Density of air, ρ_0 (kg/m^3)	1.205
Specific heat ratio of air, γ	1.402
Thermal conductivity of air, κ [W/(m K)]	0.0257
Specific heat at constant pressure of the air, C_p [J/(kg K)]	1.005×10^3
Coefficient of shear viscosity, μ (Pa s)	1.88×10^{-5}

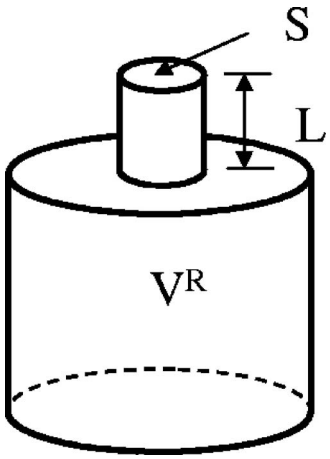


FIG. 3. A classical Helmholtz resonator.

should first be assessed. To this end, a classical Helmoholtz resonator shown in Fig. 3 is used, since a simple relationship exists between the absorption coefficient and the measurable Q factor. Therefore, it provides an alternative way to validate the absorption coefficient predicted by Eq. (21) using measured data. This relationship is derived as follows.

The acoustic output impedance of a Helmholtz resonator can be derived as

$$Z = \frac{z_0}{ikV^R A^R}, \tag{24}$$

where

$$A^R = \frac{(k^R)^2}{k^2 - iRk - (k^R)^2} \tag{25}$$

is the acoustic parameter of the Helmholtz resonator, V^R the volume of the resonator body, $R = R_w/\rho_0 cL$, L the effective neck length counting the exterior and interior end correction, and R_w the internal resistance in the resonator neck excluding the external radiation resistance, which can be computed from the measured Q factor, Q_R, and the resonance frequency ω^R of the resonator by $R_w = \omega^R \rho_0 L/Q_R$.[12]

Using the complex propagation constant $\mathbf{k}$, A^R can be also represented as

$$A^R = \frac{(k^R)^2}{\mathbf{k}^2 - (k^R)^2}. \tag{26}$$

When the medium is air, the absorption coefficient of an empty-necked resonator is very low at low frequencies, thus the complex propagation constant can be approximately expressed as $\mathbf{k} = k - i\alpha$.[14] Substituting this result into Eq. (26), expanding, and ignoring the higher-order term α^2 $(\alpha \ll 1)$ gives

$$A^R \approx \frac{(k^R)^2}{k^2 - i2\alpha k - (k^R)^2}. \tag{27}$$

Comparing Eq. (27) with Eq. (25), the internal resistance can be estimated as

$$R_w \approx 2\rho_0 cL\alpha. \tag{28}$$

Substituting $R_w = \omega^R \rho_0 L/Q_R$ into Eq. (28), the absorption coefficient of the Helmholtz resonator at its resonance frequency can be expressed in terms of the measured Q factor and resonance frequency ω^R as

$$\alpha(\omega = \omega^R) \approx \frac{\omega^R}{2cQ_R}. \tag{29}$$

For experimental validation, three resonators named HR1, HR2, and HR3 were fabricated using circular cross-sectional PVC tubes. The geometric dimensions are listed in Table III. The measured resonance frequencies, Q factors, and the predicted absorption coefficients using Eqs. (21) and (29) are also shown in Table III. It can be seen that the predicted absorption coefficients using Eq. (21) are slightly smaller than the experimentally estimated ones, with errors ranging from 6% to 9%. Generally speaking, the agreement between the model and experimental data is satisfactory. Therefore, it is concluded that Eq. (16) can be used to calculate the acoustic impedance of the TAR using Eq. (21).

TABLE III. Measured and predicted absorption coefficients.

Helmholtz resonator	Dimensions (mm)	Measured resonance freq. (Hz)	Measured Q factor	Measured absorption coefficient Eq. (29)	Predicted absorption coefficient Eq. (21)	Error (%)
HR1	Neck diameter 21 Neck length 60.0 Body diameter 77.5 Body length 155.3	136.0	35.3	0.035	0.033	6
HR2	Neck diameter 21 Neck length 79.2 Body diameter 79.2 Body length 65.3	222.0	43.5	0.047	0.043	9
HR3	Neck diameter 21 Neck length 50.4 Body diameter 77.5 Body length 57.5	245.6	47.1	0.048	0.045	6

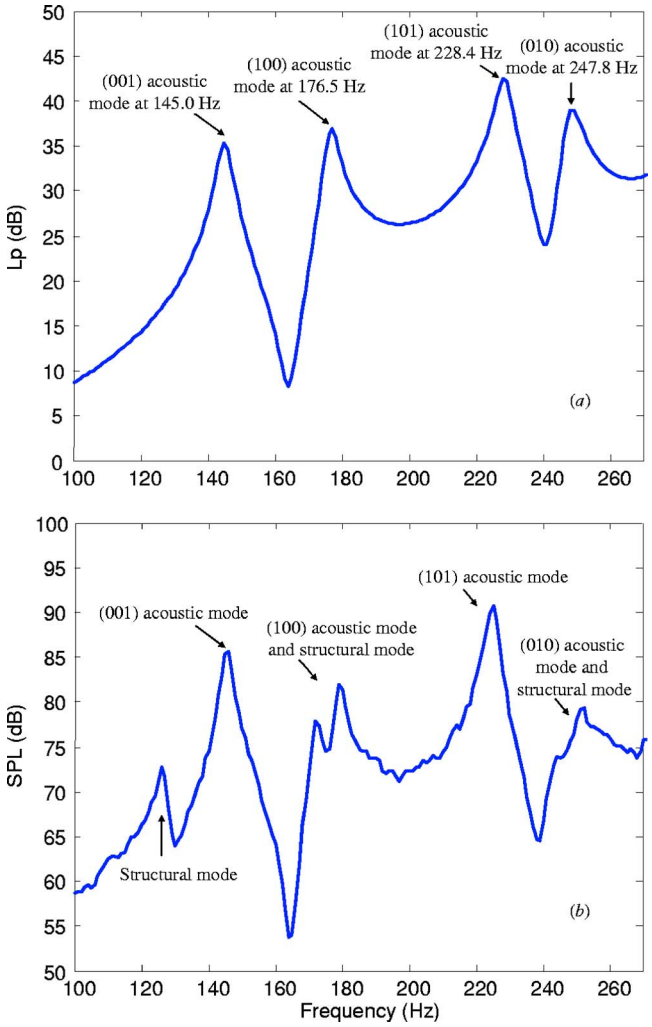


FIG. 4. (Color online) Predicted and measured baseline SPLs at (816,70,1028) mm: (a) Predicted and (b) measured.

B. Theoretical and experimental control results

In this section, the capability of the proposed SDM to tackle SPL reduction using a TAR array comprised of six TARs at two resonance frequencies is theoretically and experimentally demonstrated. First, the baseline SPL without resonators was predicted and compared with the measured baseline to validate the current theory. Second, the implementation of the SDM for noise control was illustrated in detail. Finally, the optimal location predicted for a 145 Hz TAR was experimentally validated.

1. Predicted and measured baseline SPL

The predicted and measured baseline SPLs without TARs are shown in Figs. 4(a) and 4(b), respectively. Notice that due to the difficulty in determining the volume velocity source strength Q^S for the primary sound source, numerical results are obtained in terms of $p(\mathbf{r})/(i\omega\rho_0 Q^S/k^2V^E)$. Therefore, the predicted SPL cannot be directly compared with the measured SPL. Instead, comparison will focus on the general trends and frequency-dependent variations for the sound pressure inside the enclosure. This is a common practice widely used in previous studies such as in Cummings.[13] The agreement between the predicted and measured results in this ideal condition (rigid walls) is obvious. Both curves undergo very similar global variations with respect to frequency. A more detailed inspection of the predicted SPL curve shows four major peaks at 145.0, 176.5, 228.4, and 247.8 Hz in the frequency band [100,270] Hz. The same peaks can also be found in the measured curve near these four locations. The experimental curve also shows two additional small peaks at 123.0 and 176.5 Hz. Since the wooden boards used to fabricate the enclosure walls are not fully rigid, the first peak at 123.0 Hz is a structural resonance that has been identified by substantial measurements, and the second peak at 176.5 Hz is a coupled structural-acoustic mode. This structural-acoustic coupling also affects the peak values around 247.8 Hz. Therefore, only two significant peaks were clearly identified from the measured baseline data: one at 145 Hz, having SPL 86 dB, which is dominated by the acoustic mode (001), and another at 224 Hz with SPL 91 dB, which is governed by the acoustic mode (101). While the measured (001) mode exactly matches the prediction, the predicted resonance frequency of 228.4 Hz for the (101) mode differs slightly from the measured peak at 224 Hz. The two peaks at 145 and 228 Hz will be targeted in the following noise control analyses.

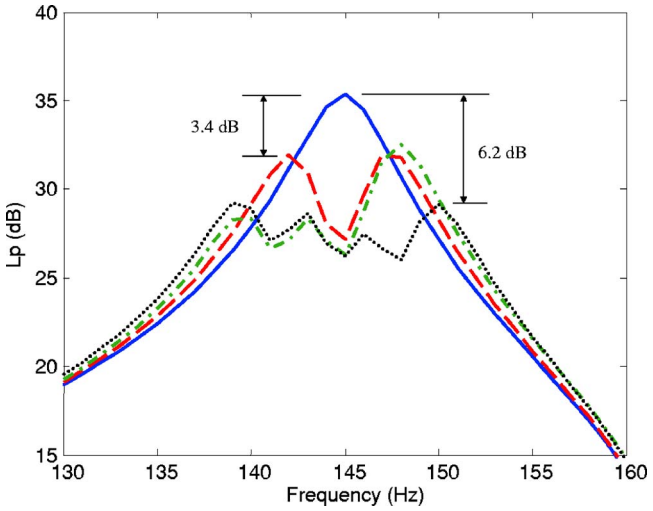


FIG. 5. (Color online) Predicted SPLs at (816,70,1028) mm:(—) Without resonator; (---) with TAR_145; (-·-) with TAR_145 and TAR_142; and (···) with TAR_145, TAR_142, and TAR_148.

2. Predicted and measured control results

The SDM procedure was implemented to control noise at the two resonance peaks at 145 and 228 Hz as follows. First, to reduce the noise level at the peak of 145 Hz, a TAR named TAR_145 was designed using Eq. (23). Note that the resonator was named using a "TAR_" and an integer indicating the resonator's Helmholtz frequency. The optimal location of TAR_145 was determined according to a procedure that will be addressed in Sec. III B 3. The theoretical SPL curve at (816,70,1028) mm with TAR_145 is shown in Fig. 5 by a dashed line. After introducing TAR_145, a 3.4 dB SPL reduction was achieved, and the original resonance peak at 145 Hz was replaced by two new peaks corresponding to two new-coupled frequencies at 142 and 147 Hz due to

TABLE IV. Parameters and optimal locations of TARs.

Resonator	Helmholtz frequency (Hz)	Branch 1 diameter (mm)	Branches 2 and 3 width × height (mm)	L_{B1} (mm)	L_{B2} (mm)	L_{B3} (mm)	Predicted location (mm)
TAR_145	145	21	29.5×29.5	30	20	482.4	(700, 400, 1128)
TAR_142	142	21	29.5×29.5	30	20	495.0	(300, 0, 1128)
TAR_148	148	21	29.5×29.5	30	20	470.4	(200, 200, 60)
TAR_229	228	21	29.5×29.5	30	20	265.7	(0, 300, 60)
TAR_224	224	21	29.5×29.5	30	20	272.5	(900, 500, 1128)
TAR_233	233	21	29.5×29.5	30	20	257.5	(800, 200, 1128)
TAR_177	177	21	29.5×29.5	30	20	374.9	(0, 0, 1128)
TAR_173	173	21	29.5×29.5	30	20	386.1	(916, 0, 110)
TAR_180	179.5	21	29.5×29.5	30	20	368.1	(60,0,0)
TAR_249	249	21	29.5×29.5	30	20	233.7	(100,60,0)

acoustic interaction. In order to further reduce the SPL around 145 Hz, another TAR named TAR_142 was designed to target the new peak at 142 Hz. Its optimal location was determined after TAR_145 was installed. The computed SPL with installed TAR_145 and TAR_142 is also shown in Fig. 5 by a dot-dashed line. From the curve, it is observed that the old peak at 147 Hz shifts to 148 Hz due to the new acoustic coupling with TAR_142. Thus, a TAR with Helmholtz frequency 148 Hz and named TAR_148 was inserted. The predicted SPL with a total of three TARs are again shown in Fig. 5 by a dotted line. This leads to a total of 6.2 dB SPL reduction in the enclosure mode at 145 Hz. The geometric dimensions of the three TARs and predicted optimal locations are listed in Table IV. These TARs were also fabricated and tuned for experimental validations. The measured SPL curves with and without TARs are shown in Fig. 6, showing 5.0 dB SPL reduction with only TAR_145 installed and 7.8 dB reduction with three resonators. Comparisons between Figs. 5 and 6 show very similar SPL variation for each configuration. Quantitatively speaking, the predicted SPL reductions are smaller than the measured ones and the predicted peaks are also sharper than the measured peaks. One plausible explanation may be that the real absorption coefficients of the resonators are slightly larger than the predicted ones using Eq. (21), which is in agreement with the observations made in Sec. III A and summarized in Table III.

To further control the peak at 228 Hz, another three TARs were designed and implemented by following the same methodology. The three resonators, TAR_228, TAR_224, and TAR_233, were added one after another at their corresponding optimal locations determined by the model. The final resonator array was comprised of six TARs that simultaneously control the two resonance peaks at 145 and 228 Hz. Their geometric dimensions, working frequency, and optimal locations are also listed in Table IV. Figure 7 shows the predicted SPLs with and without the resonator array. A 5.9 and 7.4 dB reduction in SPL around frequencies 145 and 228 Hz is predicted, respectively. Notice that the total volume occupied by all six resonators is 0.0022 m^3, which is only 0.3% of the enclosure volume 0.833 m^3. Therefore the effect of the physical presence of the resonators inside the enclosure is negligible.

The final configuration using all six resonators was also experimentally tested. Resonators (TAR_145, TAR_142, TAR_148, TAR_224, TAR_220, and TAR_228) were installed

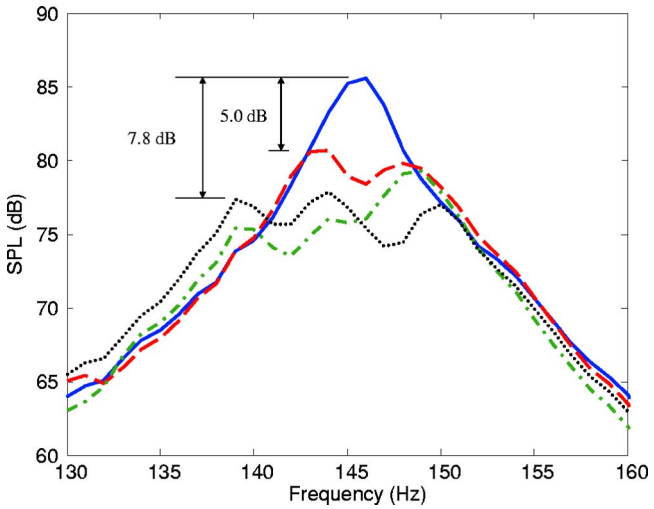


FIG. 6. (Color online) Measured SPLs at (816,70,1028) mm:(—) Without resonator; (---) with TAR_145; (---) with TAR_145 and TAR_142; and (···) with TAR_145, TAR_142, and TAR_148.

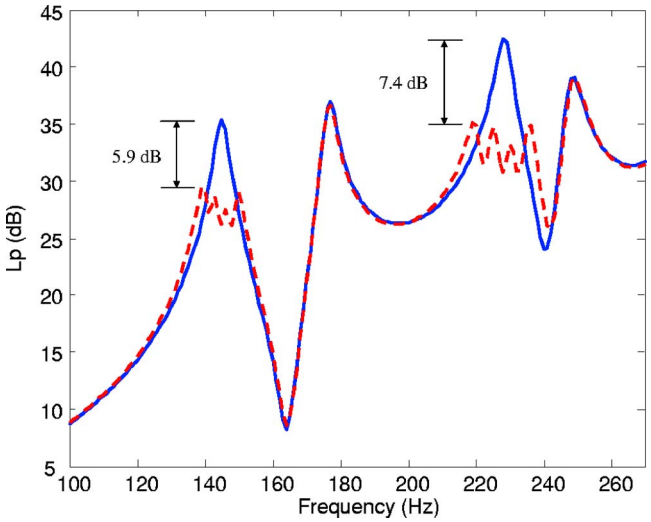


FIG. 7. (Color online) Predicted SPLs at (816,70,1028) mm:(—)Without resonator; and (---) with TAR_145, TAR_142, TAR_148, TAR_228, TAR_224, and TAR_233.

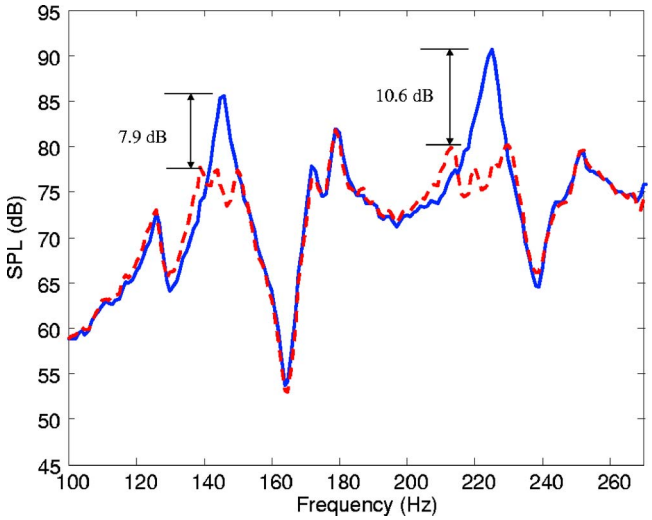


FIG. 8. (Color online) Measured SPLs at (816,70,1028) mm: (—) Without resonator; and (---) with TAR_145, TAR_142, TAR_148, TAR_224, TAR_220, and TAR_228.

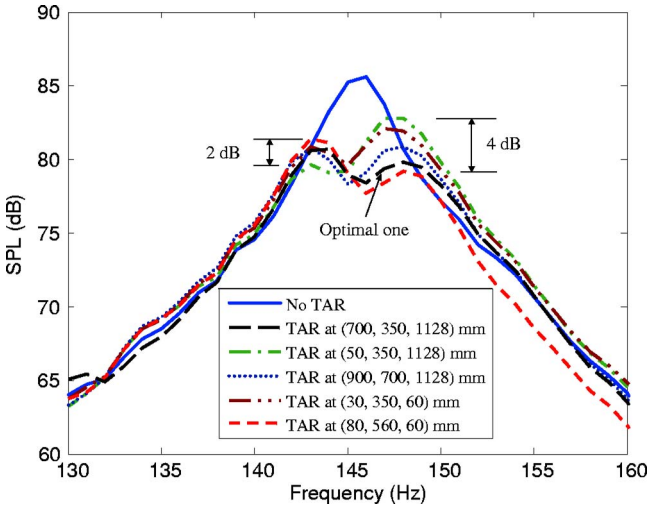


FIG. 9. (Color online) Measured representative SPL curves for identifying optimal location of the resonator TAR_145.

at locations determined by the model. The measured SPL with and without the TAR array is shown in Fig. 8. SPL reductions of 7.9 and 10.6 dB were observed around 145 and 224 Hz, respectively. Compared to the predicted results shown in Fig. 7, a remarkable resemblance in the control results can be observed, testifying to the validity of the proposed design model and methodology.

3. Predicted and measured optimal locations of TARs

Theoretical analysis in Sec. II suggests that the optimal location of a TAR is not an arbitrary point among all possible antinode points because each resonator couples with all enclosure modes. When determining the optimal location for a TAR, the interaction of the resonator with multiple enclosure modes, in addition to the targeted mode, must be considered. The proposed model actually provides such a tool. Assuming that possible locations for installing resonators are constrained near the six-wall surfaces of the enclosure, which are: $x=60$ mm (equal to the physical length of resonator Branch 1 plus the external cross-sectional height of the resonator Branch 2 or 3), $x=l_x-60$ mm, $y=60$ mm, $y=l_y-60$ mm, $z=60$ mm, and $z=l_z-60$ mm, a grid with a step of 100 mm in each direction was created for each surface, the frequency step is 1 Hz, and the SPL curves were computed for each TAR location at each grid point. After comparing the SPLs for all locations, the optimal one that provides the maximum SPL reduction in the vicinity of the targeted resonance frequency was chosen. Applying this procedure to a TAR_145, the optimization location was determined to be (700,400,1128) mm, with a corresponding SPL reduction of 3.4 dB around the targeted frequency of 145 Hz.

An experimental validation on this particular point was carried out. A series of measurements were conducted when the resonator was located at different locations. Five representative measured SPL curves are shown in Fig. 9, which shows a variation range of roughly 4 dB in terms of peak values among the five configurations. The experimentally determined optimal location at (700,400,1128) nm matches exactly the one predicted by simulations.

C. Theoretical design of a wide range control

This last example illustrates the capability of the proposed design approach in achieving wide band (or multiresonance) control using a resonator array, which consisted of ten TARs, including the six TARs used earlier and other four new TARs, named TAR_177, TAR_173, TAR_180, and TAR_249. Based on the control achieved in Sec. III B 2, the four new resonators were respectively designed and optimally positioned using SDM. The geometric dimensions of the four resonators and their predicted optimal locations are again listed in Table IV. The predicted SPL curves with and without the ten resonators are shown in Fig. 10. It can be seen that all four major resonances at 145, 177, 229, and 249 Hz can be simultaneously targeted, resulting in sound

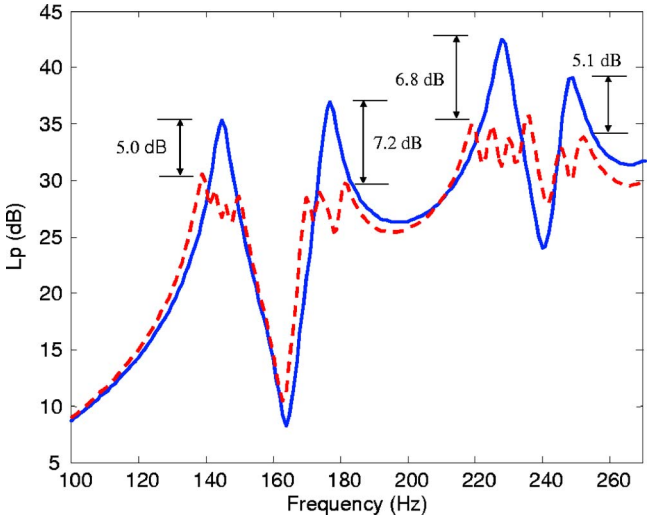


FIG. 10. (Color online) Predicted SPL curves at (816,70,1028) mm for demonstrating wide range control in the first four resonance peaks using ten TARs: (—) Without resonator; and (---) with resonators.

pressure level reduction ranging from 5.0 to 7.2 dB. Therefore, at the expense of adding more resonators, significant wide band control can be achieved.

IV. CONCLUSIONS

A theoretical model to predict the acoustic performance of multiple resonators placed in an acoustic enclosure and a reliable design methodology to maximize the control performance have been presented and experimentally validated. The sound field inside the enclosure with integrated acoustic resonators is a result of primary sound sources and the secondary sound sources formed from reradiation of the resonators, where the latter is the resonator response to the excitation from the enclosure. All these phenomena occur as a result of the acoustic interaction as a whole, which is significantly affected by the location of the resonators. The optimal location of the resonator to ensure strong interaction between the enclosure and the resonator is no longer an arbitrary point in the antinodal surface, as suggested by previous studies that only considered the interaction between the single targeted mode and the resonator. With the consideration of all possible coupled modes, the optimal location of a resonator can be determined using the general approach proposed in this paper, which is based on a general coupling model. This technique becomes a necessity when multiple resonators are used either to increase the sound reduction performance or to achieve wide range control.

The acoustic interaction is also affected by the internal resistance or the absorption coefficient of the resonator. It is found that the internal resistance of the resonator not only determines the energy dissipation inside the resonator, but also the source strength of the volume velocity directed out of the resonator aperture, and further influences the acoustic coupling with the enclosure. An optimally designed internal resistance can increase energy dissipation and improve bandwidth of the resonator around its working frequency.

A simulation model and a systematic design approach are established and validated by experiments. It is shown that, with the help of the proposed design methodology, either single or multiple resonance peaks of the enclosure can be successfully controlled.

ACKNOWLEDGMENTS

The authors wish to acknowledge a grant from Research Grants Council of Hong Kong Special Administrative Region, China (Project No. PolyU 5137/06E) and support by the Central Research Grant of The Hong Kong Polytechnic University through Grant No. G-YE67.

[1] J. M. Mason and F. J. Fahy, "The use of acoustically tuned resonators to improve the sound transmission loss of double-panel partitions," J. Sound Vib. **124**, 367–379 (1988).
[2] S. J. Estève and M. E. Johnson, "Reduction of sound transmitted into a circular cylindrical shell using distributed vibration absorbers and Helmholtz resonators," J. Acoust. Soc. Am. **112**, 2840–2848 (2002).
[3] S. J. Estève and M. E. Johnson, "Reduction of sound transmitted into a composite cylinder using distributed vibration absorbers and Helmholtz resonators," J. Vibr. Acoust. **112**, 2040–2048 (2002).
[4] S. Griffin, S. A. Lane, and S. Huybrechts, "Coupled Helmholtz resonators for acoustic attenuation," J. Vibr. Acoust. **123**, 11–17 (2001).
[5] S. A. Lane, R. E. Richard, and S. J. Kennedy, "Fairing noise control using tube-shaped resonators," J. Spacecr. Rockets **42**, 640–646 (2005).
[6] D. Li and J. S. Vipperman, "On the noise transmission and control for a cylindrical ChamberCore composite structure," J. Sound Vib. **288**, 235–254 (2005).
[7] D. Li and J. S. Vipperman, "Noise control of mock-scale ChamberCore payload fairing using integrated acoustic resonators," J. Spacecr. Rockets **43**, 877–882 (2006).
[8] D. Li, "Vibroacoustic behavior and noise control studies of advanced composite structures," Ph.D. dissertation, School of Engineering, University of Pittsburgh, Pittsburgh, PA, 2003.
[9] E. Giampaoli and S. N. Y. Gerges, "Low frequency sound absorption by cavity resonator masonry blocks," Noise Control Eng. J. **33**, 131–138 (1989).
[10] V. H. Fuchs, "Alternative fibreless absorber- new tools and materials for noise control and acoustic comfort," Acta Acust. **87**, 414–422 (2001).
[11] D. Li and J. S. Vipperman, "On the design of long T-shaped acoustic resonators," J. Acoust. Soc. Am. **116**, 2785–2792 (2004).
[12] F. J. Fahy and C. Schofield, "A note on the interaction between a Helmholtz resonator and an acoustic mode of an enclosure," J. Sound Vib. **72**, 365–378 (1980).
[13] A. Cummings, "The effects of a resonator array on the sound field in a cavity," J. Sound Vib. **154**, 25–44 (1992).
[14] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).
[15] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Melville, NY, 1989).

Porous metal absorbers for underwater sound

Xiaolin Wang[a)]
Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China

(Received 5 July 2006; revised 22 August 2007; accepted 26 August 2007)

Rubber has traditionally been used for underwater sound absorption. Porous metal is a relatively lightweight material and also has higher strength than rubber. However, exactly how porous metals can be used as effective underwater sound absorbers remains unclear. This paper shows how to use porous metal absorbers so that they work well under water, even under fairly constrained conditions. A method of nondimensional analysis is proposed that allows identification of vital characteristics. This means that such characteristics can be varied and the absorbers themselves filled with different types of viscous fluids. Such analysis suggests that the sound absorption coefficient of porous metals does not always increase when there are either increases in porosity or decreases in average pore size. The same method of analysis can show how, by choice of the right characteristics to choose a suitable viscous fluid, a porous metal absorber can be built that takes up little space but still effectively absorbs underwater sounds at low frequencies. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2785041]

PACS number(s): 43.50.Gf, 43.30.Ky, 43.55.Ev [KA] Pages: 2626–2635

I. INTRODUCTION

Porous materials have long been used to absorb sound and reduce noise. Mineral wool, fiber glass, and perforated panel materials have proven to be effective absorbers. But these materials are mostly used in the medium of air. They are not effective under water where pressure and temperature variations are more pronounced. In addition, the impedance mismatch on the surface between these materials and water is usually greater than that between the same materials and air. The impedance of rubber is similar to that of water. But the material deforms under high underwater pressure, and this leads to changes in absorbing frequencies. There is yet another, more general, problem that affects underwater sound absorbers made of homogeneous material. The wavelength of low-frequency underwater sound is usually on the order of meters. For underwater absorbers to achieve a high enough sound absorption coefficient, they also need to be correspondingly thick. This can be a significant problem where space is limited.

This paper outlines a way to address these different problems. First, porous metal is proposed as a possible alternative to more traditional materials. Porous metals have higher strength than rubber, and have impedance levels that are close to that of water. In fact, metal foams have already been used to control noise in air.[1] But few studies look at how metal foams perform in the medium of water. Cheng *et al.* investigated the underwater sound-absorbing qualities of porous aluminum and concluded that absorption was best with a porosity range of 0.75–0.8.[2] Another result was that a sample thickness of less than 15 mm hardly absorbed any sound. They also concluded that the sound absorption coefficient of porous aluminum increased as pore size decreased. Wang obtained similar conclusions, adding that the absorption coefficient also increases with increased porosity.[3] These results pertain only to experiments, and it is not clear if they have more general applications.

The second focus of this paper is to find a more general, unified way to investigate the individual observations listed earlier. A nondimensional analysis approach is proposed which identifies key controlling parameters, or nondimensional characteristics in porous metals used to absorb underwater sound. It shows how, by varying selected characteristics, it is possible to build a porous metal absorber that is effective for underwater sound at low frequencies even when space is very limited.

Zwikker and Kosten[4] introduced an approach, based on Kirchhoff's theory, that deals with thermal and viscous effects separately in the study of porous sound absorbing materials. They also based their analysis on Rayleigh's model of porous materials consisting of parallel cylindrical capillary pores. The capillary pore approach has since been employed in numerous studies concerning porous materials.[5]

Usually it is assumed that the porous frame is rigid and the fluid enclosed in the pores is homogeneous with a complex density and compressibility. For simple pore geometries like perforated holes in a thin plate, Maa's work has simplified the sound absorption formula for an airflow duct by linking solutions both at low and high frequencies.[6] A more sophisticated model developed by Wilson was based on the concept of modal wave modes and the dominance of a particular mode in the description of complex density and compressibility.[7] In the case of common porous materials which do not necessarily consist of cylindrical pores, Johnson *et al.* have developed the concepts of frequency-dependent tortuosity and viscous characteristic dimension.[8] By making use of these parameters, the complex density and compressibility are simplified in a form similar to that for porous materials with cylindrical pores.[9]

When it is assumed in the analysis that the porous frame is elastic, the most cited work is Biot's theory, in which a

[a)]Electronic mail: wangxl@mail.ioa.ac.cn

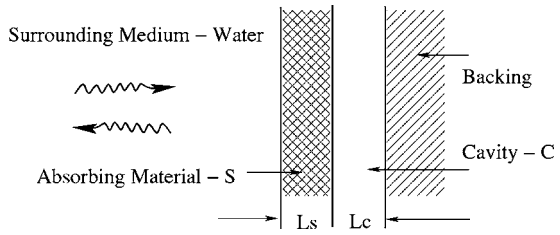


FIG. 1. Schematic description of an absorber. The porous plate S is saturated with medium s. The plate is sandwiched by two sound-transparent layers. Between the plate and a rigid wall there is a cavity C filled-in by medium c.

continuum approach is adopted.[10,11] Both the solid frame and the fluid in the pores are considered as two homogeneous phases interacting with each other. Recently, Göransson[12] published results based on Biot's theory for the overall governing equations, Johnson's model for the enclosed fluid, and Gibson and Ashby's scaling law[13] for the solid frame. A finite element method is employed to solve the basic equations.

In order to concentrate on the properties of nondimensional controlling parameters, this paper assumes the porous materials have rigid frames. Although this assumption is less valid for water-filled materials than it is for air-filled materials, taking into account the frame elasticity will be more likely to further increase the effective thickness of a porous absorbing plate. If a rigid framed porous absorber can be effective for underwater sound, so could it be more effective for an elastic one to absorb the sound. Parallel cylindrical pores are first assumed for the modeling of perforated panels, followed by the use of equivalent tortuosity and flow resistivity for the modeling of fibrous metal plates. Based on a domain-match method[14] in which both viscous and thermal effects are considered, this paper extends the method by including more variables. By analyzing these variables and the absorbing mechanisms, the possibility of applying porous metals in underwater sound absorption at low frequencies is explored.

II. BASIC PROBLEM AND THEORY

Figure 1 shows the structure of the underwater sound absorber used in this paper. A porous plate S is filled with medium s and the plate is sandwiched between two sound-transparent layers which prevent medium s from leaking into the surrounding medium. Just behind the sandwiched plate lies a cavity C, and then a rigid wall. Cavity C is filled with another fluid, medium c.

Sound waves propagate through the surrounding water before impinging normally on the porous plate. Sound energy is dissipated via viscous friction and damping. The viscous friction is generated between pore walls and the enclosed medium s. There is, in principle, also internal friction or damping within the porous skeleton itself. But this analysis assumes that the plate is rigid so that there is no damping effect from the porous skeleton of the plate, i.e., the plate is assumed to be rigid. The remaining sound energy enters the cavity. The cavity is filled with fluid, which bounces the sound back into the porous plate. The cavity behind the plate acts like an elastic spring that lowers the fundamental and higher resonance frequencies of the absorbing system. In short, the whole structure behaves like a viscoelastic system.

This paper focuses on two questions: how to identify relevant controlling parameters, and how these parameters affect sound absorption. Answering these questions helps to clarify related issues such as whether porous solids absorb sound better with higher porosity levels and smaller pore sizes. Another concern is how to decide on values for the controlling parameters and choose the fluid with which to fill the absorber, if the sound absorption coefficient has to be above α_0 for sound around a specific frequency f_0?

There are two characteristics which are vital for the determination of sound absorption in a continuum medium. One is the characteristic impedance $Z=\sqrt{\rho/C}$, where ρ and C are complex density and complex compressibility of the medium. The other is the complex propagation constant $m=i\omega\sqrt{\rho C}$, where ω is the angular frequency of the impinging sound. Let subscript 0 denote variables in the surrounding water, such as pressure p_0, density ρ_0, sound velocity c_0, characteristic impedance Z_0, and density and sound velocity at the ambient temperature ρ_0^*, c_0^*. Variables of medium s are denoted by subscript s; the surface impedance between water and the plate is z_1, while the surface impedance between the back of the plate and the cavity is z_2. The plate has porosity Ω. For a layered absorber, the sound absorption coefficient α and the surface impedance z_1 are[4]

$$\alpha = \left| \frac{z_1 - Z_0}{z_1 + Z_0} \right|^2,$$

$$z_1 = Z_s \frac{z_2 \cosh m_s L_s + Z_s \sinh m_s L_s}{z_2 \sinh m_s L_s + Z_s \cosh m_s L_s}, \tag{1}$$

where L_s is the thickness of the porous plate. In order to obtain certain values of α, variables such as z_2, Z_s, $m_s L_s$, need to be solved first.

AQ: Ok?Suppose a porous substance with tortuosity χ, which characterizes the complexity of the pores, is saturated with a liquid with the Prandtl number N_{Pr} and the parameter of nonlinearity B/A. The complex density and compressibility of the filled-in medium in the pores are then[14]

$$\rho_s(\omega) = \frac{i\chi^2}{\omega\langle v\rangle\Omega}\frac{dp}{dz},$$

$$C_s(\omega) = \frac{\Omega}{(1+B/A)p_s^*}\left[1 + \frac{B}{A}[1 - \rho_s^*/\rho_s(N_{\mathrm{Pr}}\omega)]\right], \tag{2}$$

where v denotes the fluid particle velocity along the axis z of a cylindrical pore, while $\langle v\rangle$ refers to the average of the particle velocity in the z direction across the pore. If the porous material is filled with gas, Eq. (2) can be converted to the commonly cited relation for the compressibility of gases. Based on the second-order approximation to the constitutive relations for an ideal fluid,[15]

$$p = p_s^* + A\frac{\rho - \rho_s^*}{\rho_s^*} + \frac{B}{2!}\left(\frac{\rho - \rho_s^*}{\rho_s^*}\right)^2, \tag{3}$$

the parameter of nonlinearity B/A for gases has a relation to the specific heat capacity ratio γ,[16]

$$\gamma \equiv \frac{B}{A} + 1, \tag{4}$$

where

$$A = \rho_s^* \left[\left(\frac{\partial p}{\partial \rho} \right)_e \right]_{\rho=\rho_s^*},$$

$$B = \rho_s^{*2} \left[\left(\frac{\partial^2 p}{\partial \rho^2} \right)_e \right]_{\rho=\rho_s^*}, \tag{5}$$

and the subscript e denotes a constant entropy, while for liquids γ is the polytropic index. A porous material can be simply modeled as a solid frame consisting of parallel tubes with equal diameters. If all the tubes are parallel to the direction of sound propagation, the tortuosity χ is 1. For a single tube, the governing equation for fluid flow inside the tube is[14]

$$\nabla^2 v - \frac{i\rho_s^* \omega}{\mu} v = \frac{1}{\mu} \frac{dp}{dz}, \tag{6}$$

where μ is the dynamic viscosity of the fluid in the tube, and p is the pressure perturbation. After nondimensionalization, Eq. (6) becomes

$$\bar{\nabla}^2 \bar{v} - i\beta^2 \bar{v} = 1, \tag{7}$$

where β is the acoustic Reynolds number $\beta \equiv r\sqrt{\omega \rho_s^*/\mu}$, r is the radius of the tube. $\bar{v} \equiv v\mu/(r^2 dp/dz)$, $\bar{\nabla}^2 \equiv \partial^2/\partial \bar{x}^2 + \partial^2/\partial \bar{y}^2$, and $\bar{x} \equiv x/r$, $\bar{y} \equiv y/r$. Generally speaking, a bundle of parallel tubes has a flow resistivity of $\sigma_s = 8\chi\mu/r^2\Omega$, and an acoustic Reynolds number of $\beta = 4\chi\sqrt{\pi\rho_s^* f/\sigma_s \Omega}$.

III. INFLUENCE OF NONDIMENSIONAL CHARACTERISTICS

The above-presented analysis indicates that there is a set of nondimensional characteristics $\{\beta, N_{\mathrm{Pr}}, B/A, \Omega, \chi, k_s L_s, k_c L_c, \eta_1, \eta_2\}$, where k_s, k_c are wave numbers in media s and c, respectively, $k_s \equiv 1/\lambda_s$, $k_c \equiv 1/\lambda_c$, and $\eta_1 \equiv \rho_s^* c_s^*/\rho_0^* c_0^*$, $\eta_2 \equiv \rho_c^* c_c^*/\rho_s^* c_s^*$. The first of this set of characteristics is the acoustic Reynolds number, which is the ratio between viscous and inertia forces. The second and the third are physical constants of the saturating medium in the porous material. The Prandtl number is the ratio of thickness between the viscous and thermal boundary layers inside the tube. The parameter of nonlinearity B/A characterizes nonlinear terms in the constitutive relation of the saturating medium enclosed in the pores. The fourth and fifth characteristics are about the structure of the porous material. For simplicity, influences involving tortuosity χ are discussed later in Sec. IV B dealing with fibrous metal absorbers, so the sample discussed in this section is assumed to be a perforated plate with parallel tubes running normal to the plate surface, or, in other words, $\chi = 1$. The sixth and seventh are about wave propagation characteristics of the saturating media. The eighth and the ninth characterize wave propagation through interfaces between two media. How these nondimensional characteristics affect sound absorption are discussed further in the following.

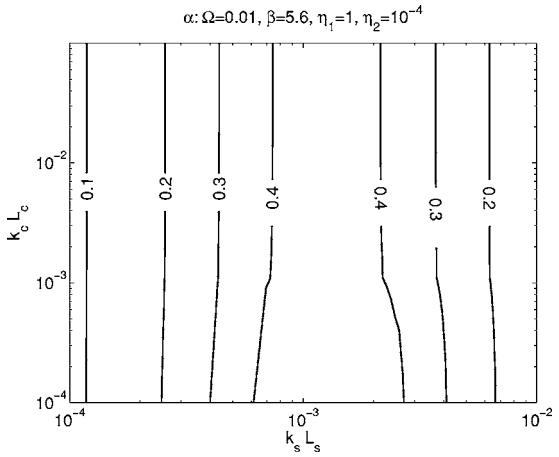


FIG. 2. Sound absorption coefficient contour as a function of $k_s L_s$ and $k_c L_c$. A perforated panel of $\Omega = 0.01$ and $\beta = 5.6$ is saturated with the surrounding medium—water. Air cavity is behind the plate.

Before proceeding to the analysis of these characteristics, Fig. 2 is given in order to illustrate why, apart from applying a cavity behind the porous material, there is a need to saturate the porous material with certain fluids rather than simply immerse the material into the surrounding medium. Sound absorption contours are shown as a function of $k_s L_s$, $k_c L_c$. If a porous plate is immersed directly with the surrounding medium, the extent to which the cavity alone can increase the effective thickness of the porous plate can be determined. In Fig. 2, a perforated plate of porosity $\Omega = 0.01$ is immersed directly into water with the plate itself backed by an air cavity. A perforated plate with parallel cylindrical pores of $r = 0.1$ mm diameter for absorbing underwater sound at 500 Hz is indicated by the acoustic Reynolds number $\beta = 5.6$. The reasons for choosing such porosity and pore diameter based on later characteristics analysis are to obtain as many favorable values as possible under given conditions. Figure 2 shows that the cavity is hardly enough for the absorber to effectively absorb low frequency underwater sound.

Now back to the analysis. The first parameter to be examined is the acoustic Reynolds number β. For perforated panels with parallel cylindrical pores, $\beta = 4\chi\sqrt{\pi\rho_s^* f/\sigma_s \Omega}$. For porosity near unity, the number is a surrogate for the quotient f/σ, which has been used extensively by Delany and Bazley[17] as a parameter to characterize porous materials. The acoustic Reynolds number therefore characterizes the effect of the flow enclosed in the porous material. In order to focus on this, the effect of variation between media in front of and inside the porous plate is ignored at this stage, $\eta_1 = 1$. The resonance effect of the cavity is also considered later in this section, i.e., for now $k_c L_c = 0$, $\eta_2 = 0$. A porous plate with high porosity $\Omega = 0.99$ is used as an example.

Figure 3 shows sound absorption coefficient contours when the perforated plate backed by a rigid wall is immersed in water. In Fig. 3, Λ is a characteristic length $\Lambda \equiv c_s \rho_s^* r_h^2/\mu = \beta^2 \lambda/2\pi$, which includes effects of both flow resistivity as well as the speed of sound in the fluid in the pores. A highly porous plate with a thickness exceeding $\Lambda/2$ has an absorption relationship commensurate with a plate of

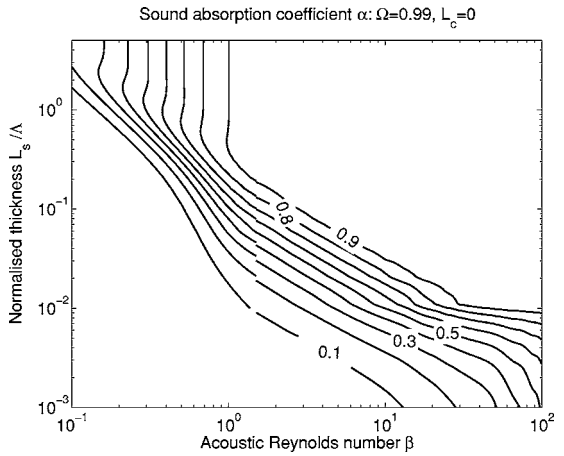


FIG. 3. Sound absorption coefficient contour as a function of the acoustic Reynolds number β and a normalized thickness L_s/Λ. The perforated plate, backed by a rigid wall without any cavity, is immersed directly into water.

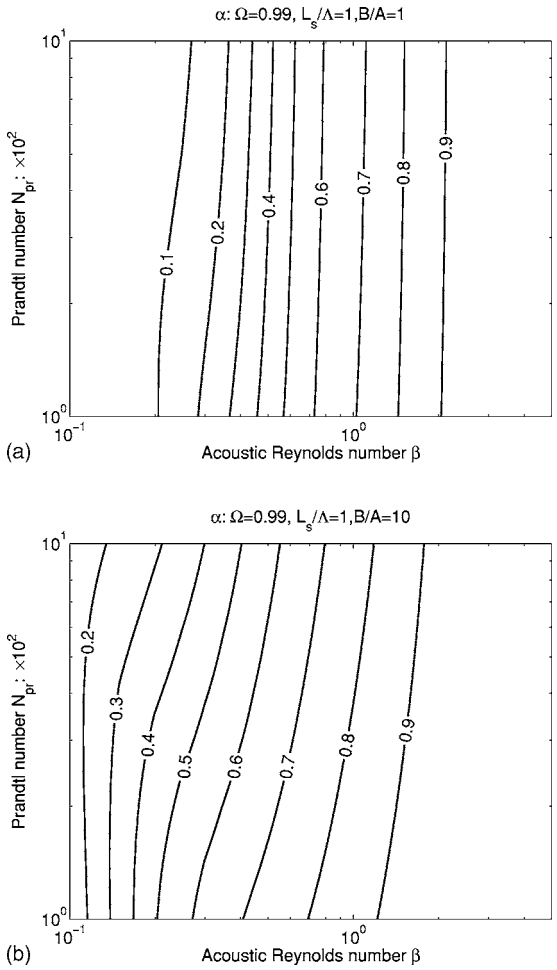


FIG. 4. Sound absorption coefficient contour as a function of β and N_{Pr} at various B/A. No cavity behind the plate. (a) $B/A=1$ and (b) $B/A=10$.

infinite thickness. Figure 3 indicates that for a plate with high porosity, the sound absorption coefficient tends to grow with increases in the acoustic Reynolds number while the ratio L_s/Λ remains constant.

The next parameters for discussion are the nondimensional physical constants N_{Pr} and B/A. Both constants reflect thermal effects on the compressibility of the fluid inside the porous plate. Taking highly porous plates as examples, the ratio L_s/Λ is assumed to be 1 in order to exclude the effect of plate thickness. The Prandtl number for gases is usually of the order $\sim O(1)$, and that for liquids varies across a wide range, $\sim O(10-10^3)$. The parameter of nonlinearity for gases is usually <1, while the parameter for liquids is around 5–10. Two typical values, 1 and 10, of the parameter of nonlinearity are taken as examples in Fig. 4, which shows how contours of sound absorption coefficients are influenced by β and N_{Pr}. Figure 4 shows that for gases, contours above 0.8 are little affected by the Prandtl number. For liquids, the Prandtl number has a slightly stronger, but still insignificant, effect on the same contours. This indicates that both the Prandtl number and the parameter of nonlinearity have no significant impact on sound absorption. This is also in agreement with what has been observed by other authors, e.g., Albert,[18] when comparing acoustical properties of air- and water-filled materials, that compressibility effects are less important in liquids than in gases. For simplicity, the following analysis takes the values of both constants to be 1.

Next discussed is the influence of porosity Ω. As in the previous analysis, it is still assumed that there is no cavity behind the plate, and that the media both inside and in front of the porous plate are the same, i.e., $k_cL_c=0$, $\eta_1=1$, $\eta_2=0$. Figure 5 shows how the acoustic Reynolds number β, and the propagation parameter k_sL_s affect sound absorption at various porosities. It indicates that when porosity increases, thinner plates with smaller Reynolds numbers attain higher absorption coefficients. A higher porosity usually represents a higher specific surface area. This means a greater area of friction between pore walls and the enclosed fluid. More sound energy is thus dissipated.

The next two characteristics to be studied are the impedance ratio η_1, and the wave propagation parameter k_sL_s. The discussion at this stage still excludes possible effects from the cavity. But the effect of variations between fluids inside and in front of the porous plate is now taken into account. In order to concentrate on the effect of the impedance ratio η_1, the plate is first assumed to be infinitely thick. Figure 6 shows the absorption coefficient contour as a function of the acoustic Reynolds number β and the ratio η_1 at various porosities. Figure 6 indicates that in order to have sound absorption coefficients above 0.8, for high porosity plates, the impedance ratio η_1 needs to be near unity. For low porosity plates, this impedance ratio becomes smaller. A complete absorption of impinging sound requires the fluid in front of the porous plate to have the same impedance as that at the surface of the plate. It is observed when porosity increases, plates with higher impedance ratios η_1 and smaller Reynolds numbers attain higher absorption coefficients. This suggests that an effective impedance ratio can be attributed to both the impedance of the fluid inside the porous plate and the flow resistivity of the plate. Generally speaking, the impinging sound is partly reflected by the solid surface of the porous plate. A low porosity yields a high flow resistivity. A low porosity plate then requires a lower ratio η_1 so that the impedance of the saturated plate surface matches the impedance of the fluid in front of the plate.

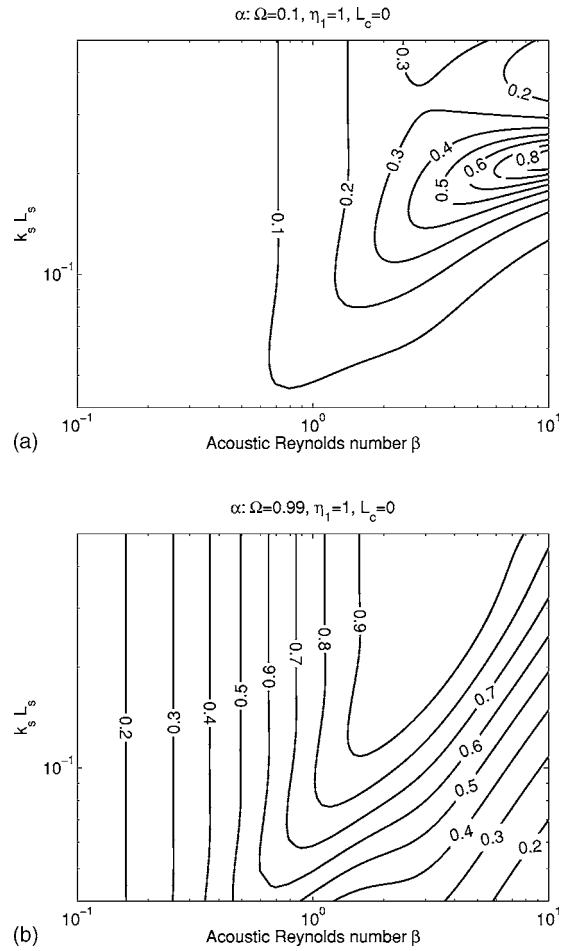


FIG. 5. Sound absorption coefficient contour as a function of β and k_sL_s at various Ω. No cavity behind the plate. (a) $\Omega=0.1$ and (b) $\Omega=0.99$.

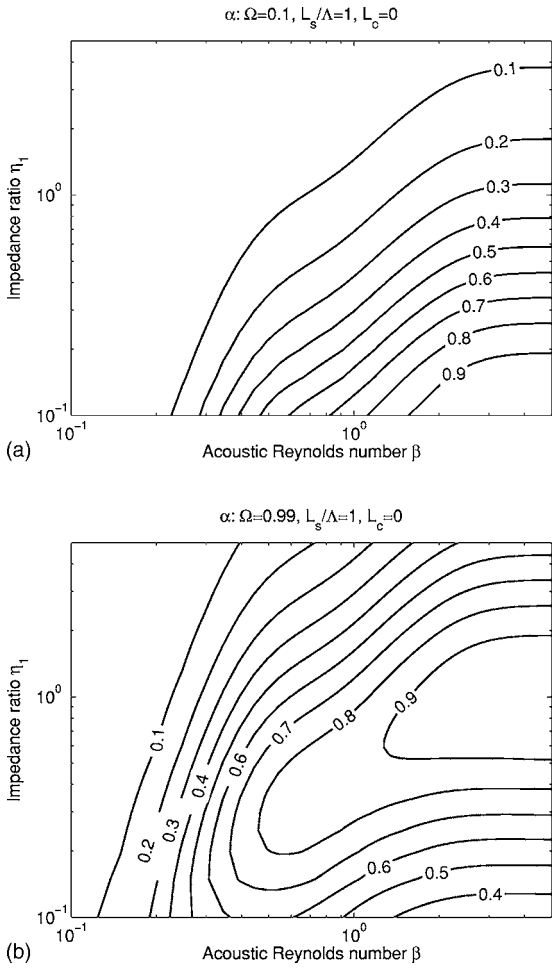


FIG. 6. Sound absorption coefficient contour as a function of β and η_1 at various Ω. No cavity behind the plate. (a) $\Omega=0.1$ and (b) $\Omega=0.99$.

The discussion now assumes that the thickness of the plates is finite. The influence of the impedance ratio η_1 is shown in Fig. 7, where the absorption coefficient contour is taken as a function of the acoustic Reynolds number β and the propagation number k_sL_s. When the ratio η_1 is around 0.5–1, high absorption can be obtained with the thinnest possible plate and a low acoustic Reynolds number. This is consistent with the trend when the plate is almost infinitely thick. Figures here indicate that if the impedance ratio is in the range $\eta_1<10^{-1}$, there is unlikely a sound absorption coefficient higher than 0.8. For example, the impedance ratio of air to water is of the order $\sim O(10^{-4})$. It is almost impossible for a porous plate filled with air to absorb underwater sound at low frequencies. This is because the effective impedance of the plate cannot match the impedance of water.

Parameters such as the impedance ratio η_2 and the propagation number k_cL_c are related to the cavity behind the porous plate. The following discussions take into account the existence of the cavity, in three groups. The first group deals with the effect of impedance ratio η_2 (Fig. 8), the second with the effect of porosity (Fig. 9), while the third deals with the effect of the acoustic Reynolds number (Fig. 10).

The graphs in Fig. 8 indicate that for higher sound absorption coefficients, the smaller the impedance ratio η_2, the thinner and shallower the plate and cavity need to be. In this way, the cavity is like a spring. A smaller η_2 represents a lower ratio between the impedance of the cavity medium and that of the saturating fluid enclosed in the pores. This is similar to a spring-mass system with a softer spring which has a lower fundamental resonance frequency. These graphs also show that an absorber with smaller η_2 tends not only to absorb sound at lower frequencies (corresponding to Reynolds numbers) but also to take up a smaller space (thickness). For most liquids and gases, the smallest possible η_2 is around 10^{-4}, i.e., a liquid saturated porous plate with a cavity of air. From this point on, all discussion assumes $\eta_2\sim10^{-4}$.

The graphs in Fig. 9 demonstrate that when porosity decreases, thinner plates with deeper cavities attain higher absorption coefficients. In other words, the absorption coefficient increases with falling porosity. This seems to contradict a common belief that the higher the porosity, the higher the sound absorption coefficient. This also seems to contradict the tendency indicated in Fig. 5. These seeming contradictions can be explained by the existence of the cavity. In most circumstances, high porosity represents a large specific surface area which tends to increase viscous dissipation. The existence of the cavity in the mean time provides resonance to enhance this dissipation. The reduction of specific surface area is therefore compensated by the dissipation enhancement. The analysis is thus required to take all factors into account. These include viscous dissipation, mass inertia of

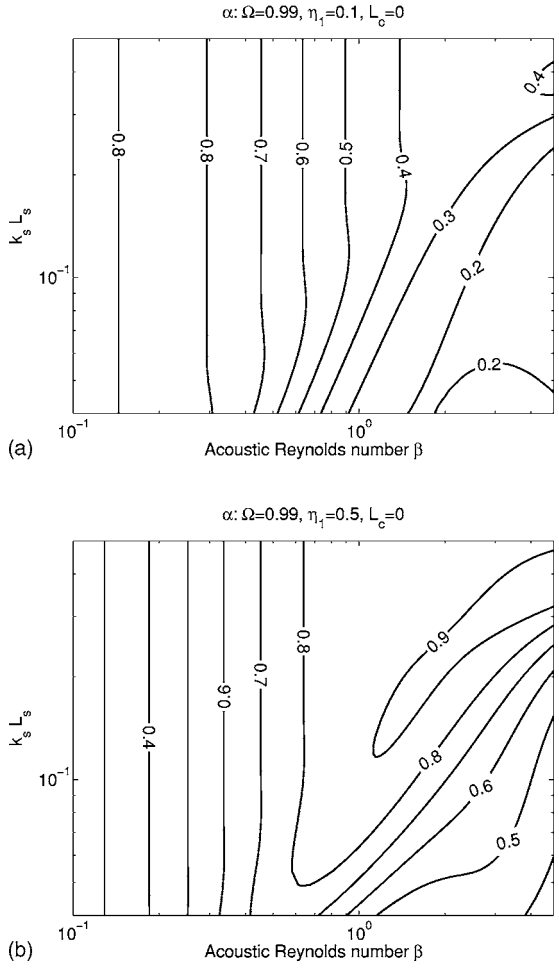


FIG. 7. Sound absorption coefficient contour as a function of β and k_sL_s at various η_1. No cavity behind the plate. (a) $\eta_1=0.1$ and (b) $\eta_1=0.5$.

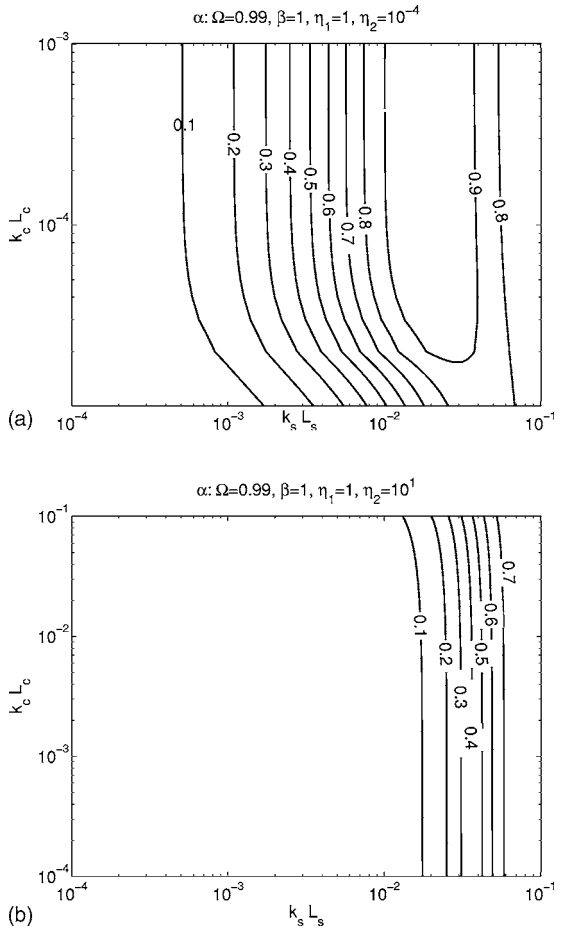


FIG. 8. Sound absorption coefficient contour as a function of k_sL_s and k_cL_c at various η_2. (a) $\eta_2=10^{-4}$ and (b) $\eta_2=10^{1}$.

the filled-in liquid, and the cavity resonance. These are discussed in Sec. IV in the context of a more specific absorber.

The graphs in Fig. 10 show that when the acoustic Reynolds number decreases, thinner plates with deeper cavities attain higher absorption coefficients. If the pore structure of a perforated panel is given, the acoustic Reynolds number is proportional to the square root of frequencies, $\beta \propto \sqrt{f/\sigma}$. However, according to theories about wave propagation in layered media, a thicker absorber tends to absorb lower frequencies. This seems to contradict the results calculated here. Again, this can be explained if all factors are taken into account. This is discussed further in Sec. IV.

IV. ABSORBER FOR SOUND AT LOW FREQUENCIES

The preceding section examined the effects of various nondimensional parameters on sound absorption. This section examines how to choose these parameters so that a small-sized absorber can achieve a high absorption coefficient for sounds at low frequencies. More specifically, examples of both perforated panels and fibrous metal plates are used to show how to construct an absorber with absorption coefficients above 0.8 for sounds around 500 Hz in liquids.

A. Perforated panels

Perforated panels are first used to determine possible plate thickness and pore size for optimum sound absorption in water. The panel is drilled with parallel holes running through the panel, perpendicular to the surface. Two panels, each drilled with holes of two different sizes, are used to show how thick an absorber might need to be if they are simply immersed in water without examination of optimal nondimensional parameters. If the holes of the panel have a diameter of 1 mm, the acoustic Reynolds number in water for sound at 500 Hz is around 56, and the characteristic length Λ is around 1500 m. According to Fig. 3, if there is no cavity, a highly porous panel must be at least $10^{-3}\Lambda$ thick, i.e., 1/3 of the wavelength of the underwater sound, or 1 m. Less porous panels need to be even thicker to absorb sound at the same level of absorption. If the diameter of the hole is 0.1 mm, the acoustic Reynolds number would be around 5.6. The thickness of the panel then needs to be $10^{-2}\Lambda$, i.e., 1/20 of the wavelength, or 0.15 m. In both cases working with such thick panels presents real practical difficulties. The discussion to follow shows how much thickness can be reduced by examining optimal nondimensional parameters.

According to previous discussions, the influence of both the Prandtl number N_{Pr} and the parameter of nonlinearity B/A can be simply set to an initial number, in this case, 1. The impedance ratio η_2 is assumed to be 10^{-4} for optimal

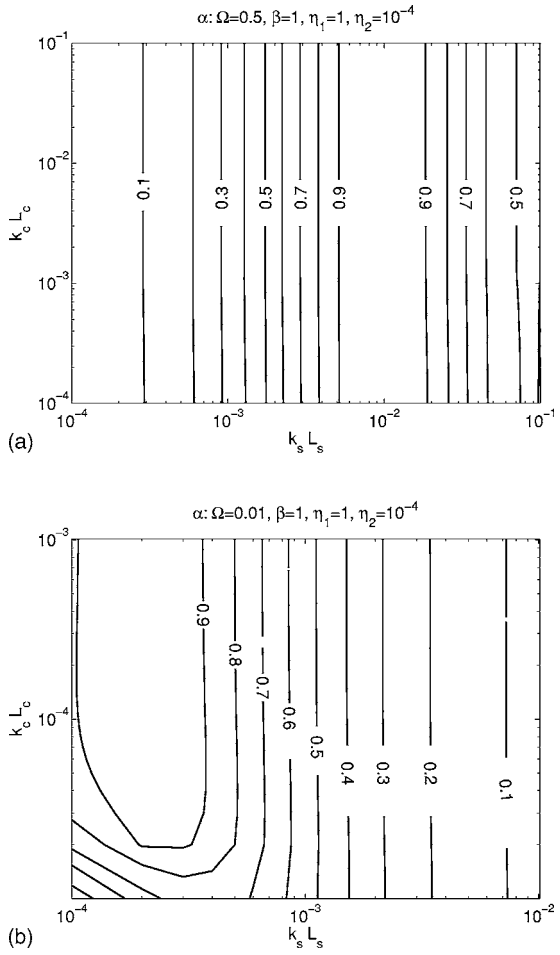


FIG. 9. Sound absorption coefficient contour as a function of k_sL_s and k_cL_c at various Ω. (a) $\Omega=0.5$ and (b) $\Omega=0.01$.

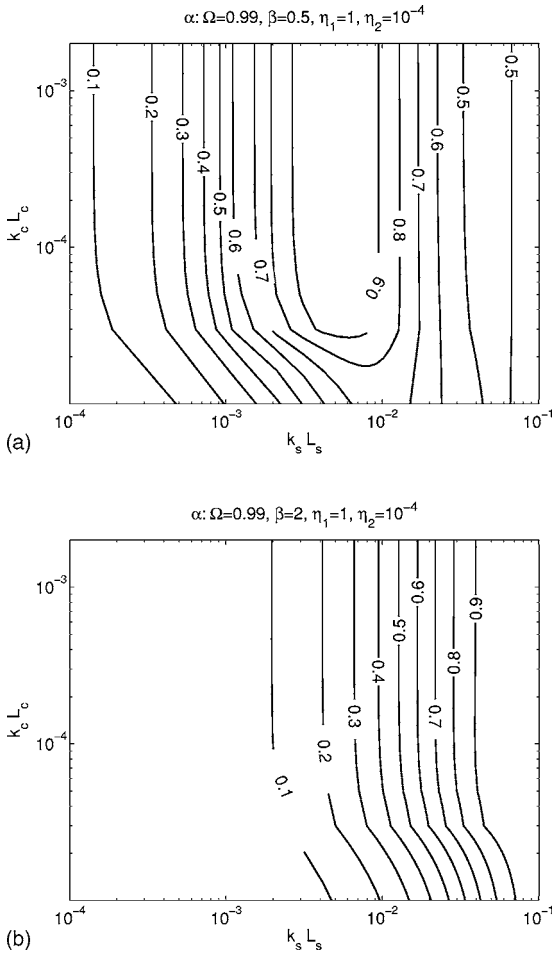


FIG. 10. Sound absorption coefficient contour as a function of k_sL_s and k_cL_c at various β. (a) $\beta=0.5$ and (b) $\beta=2$.

performance. For the same reason, the impedance ratio η_1 is initially set at 1, and is readjusted later in this section. The other adjustable parameters are β, Ω, k_sL_s, and k_cL_c. A liquid for saturating the panel can be selected by choosing the right β. The pore structure can be determined by Ω while the thickness of the plate and the depth of the cavity can be determined by k_sL_s and k_cL_c.

If there is no cavity, i.e., $L_c=0$, the impedance ratio is $\eta_2=0$. Figure 6 shows that when β is above 1 to 2 or just below 1, the sound absorption coefficient can reach 0.8. This discussion assumes, for a start, that the nondimensional parameters take the following values: $\beta=1$, $\eta_1=1$, and $\eta_2=10^{-4}$. In this case, according to Fig. 9, the optimal porosity range is $\Omega<0.1$, and the ratio between the thickness of the plate and the wavelength of the medium enclosed in the porous plate can be as small as 10^{-3}. Taking into account the possible range of η_1 shown in Fig. 7, the initial set of optimal nondimensional parameters can be $\beta>1$, $\Omega<0.1$, $\eta_1>0.1$, and $\eta_2=10^{-4}$. In other words, for sound absorption coefficients higher than 0.8, the thickness of the porous plate can be as thin as one-thousandth of the wavelength of the medium in the porous plate, and the cavity depth can be one ten-thousandth of the wavelength of the cavity medium.

Note that no mention has been made of either specific saturating fluids or specific geometric values of the absorber. Suppose the pores are now 1 mm in diameter, and the cavity is filled with air. To achieve a sound absorption coefficient of 0.8, saturating liquids need to meet criteria mentioned in the preceding paragraph. Liquids that meet these criteria include glycerine ($\beta=1.63$, $\eta_1=1.65$, $\eta_2=1.7\times10^{-4}$), castor oil ($\beta=1.76$, $\eta_1=1.00$, $\eta_2=2.8\times10^{-4}$), and silicone oil ($\beta=1.77$, $\eta_1=0.66$, $\eta_2=4.3\times10^{-4}$).

If castor oil is selected as the saturating liquid, a panel fitted with a cavity only needs to be less than 1 cm thick with a porosity of 0.05. The cavity itself only needs a depth of 1 mm, as shown in Fig. 11. But as the porous panel is assumed to be rigid, the panel needs to be at least 1 cm thick. Even so, the thickness of the whole absorber is no more than one-hundredth of the wavelength of the underwater sound. Filling the porous plate with viscous fluids means a mounting space for the absorber that is much smaller than that needed when a porous plate is simply placed underwater. Experiments were conducted to verify what this calculation suggests. According to the contour map in Fig. 11, an absorber in the experiment was a 2-cm-thick porous aluminum plate saturated with castor oil. A 2-cm-wide air cavity was set between the plate and a rigid wall. Figure 11 indicates that this absorber should theoretically achieve a sound absorption coefficient of 0.8 for underwater sound around 500 Hz. Falling into the same region enclosed by contour 0.8 in Fig. 11, there are another two absorbers which both have

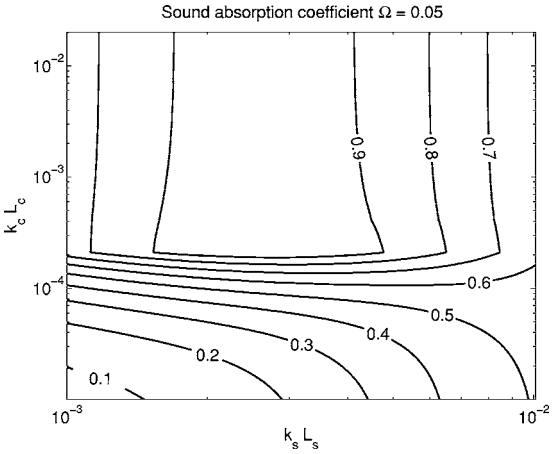


FIG. 11. Sound absorption coefficient contour as a function of k_sL_s and k_cL_c at $\Omega=0.05$. A perforated panel is saturated with castor oil.

3 cm air cavities. Similar to Fig. 11 for castor oil, we can also draw a graph for silicone oil, and expect that a 2-cm-thick perforated panel with 2 cm air cavity can absorb 500 Hz sound with an absorption coefficient above 0.8. Experimental results for all four absorbers shown in Fig.12 confirm our prediction.

B. Fibrous metal plates

This section looks at how to find a fibrous metal absorber with absorption coefficients above 0.8 for sound around 500 Hz in castor oil. Two types of porous metal fibers, CME and Bekipor, are used here because both have been tested by Albracht and Lotze.[19] Plates of these fibers have porosities of $\Omega=0.76-0.90$ and tortuosities of $\chi=1.3-3$. The last section dealt with absorption using perforated plates. This section extends the method to fibrous metal plates.

First, the extension of the modeling is verified by comparing calculations in this paper to experimental results by

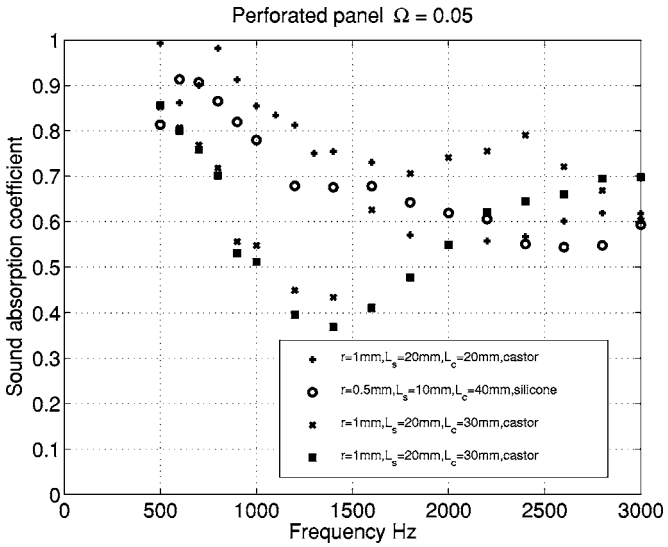


FIG. 12. Verification tests. In each test, a perforated panel is filled with either castor oil or silicone oil. All panels have through holes with a porosity of 0.05. An air cavity is behind each plate.

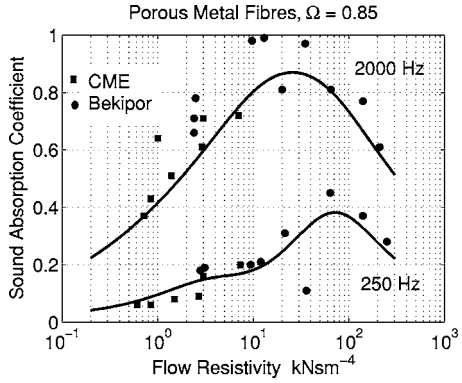


FIG. 13. Verification with published experimental results of CME and Bekipor (Ref. 19). Solid lines are theoretical calculations. Porosity $\Omega=0.85$ and tortuosity $\chi=1.3$.

Albracht and Lotze.[19] In their experiments, a 5-cm-thick fiber plate is backed by a rigid wall. Air sound absorption coefficients at various flow resistivities are given in Fig. 13. In Fig. 13 all the dots are their experimental results, while the solid lines are our calculations based on a porosity of $\Omega=0.85$ and tortuosity of $\chi=1.3$. We also conducted underwater experiments with fiber metal plates of similar porosity and static flow resistivity. In each experiment, one such plate of 1 mm thick, backed by a 2 cm air cavity, is saturated with silicone oil of viscosity 1000 cS. Underwater sound absorption coefficients are given in Fig. 14. Both experimental and theoretical results agree with each other.

This investigation uses a fibrous metal with a porosity of $\Omega=0.76$ and tortuosity of $\chi=3$. In order to achieve the optimal absorption, it is necessary to choose the right saturating fluid and decide on flow resistivity, plate thickness, and cavity depth. As in the earlier discussion of perforated panels, nondimensional parameters η_1, β, Ω, k_sL_s, and k_cL_c need to be examined.

The absence of a cavity has the same effect on sound absorption as that shown in Fig. 7 for various η_1. This im-

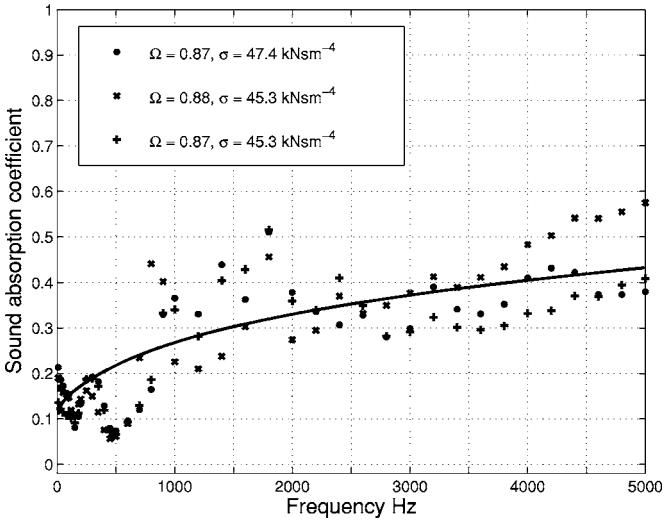


FIG. 14. Verification tests. A fiber metal plate 1 cm thick, backed by 2 cm air cavity, is filled with silicone oil. Plates are provided by H. P. Tang and J. L. Zhu of the Northwest Research Institute of Ferrous Metals. Solid lines are theoretical calculations with tortuosity $\chi=1.3$.

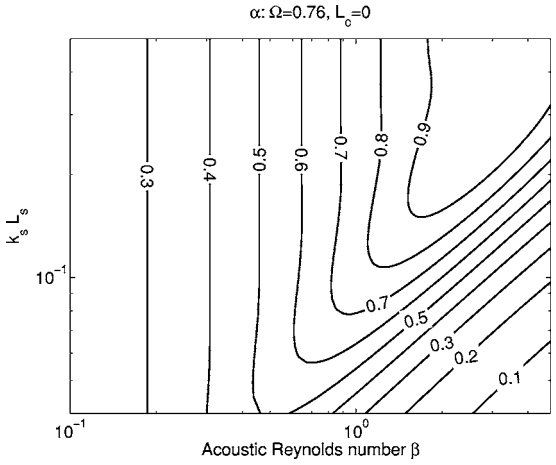


FIG. 15. Sound absorption coefficient contour as a function of β and k_sL_s. A porous panel of Ω=0.76 and χ=3 is saturated with silicone oil. No cavity behind the plate.

plies that a saturating fluid with shorter wavelengths is preferred. However, if η_1 is too small, which can happen when plates are filled with air, it is harder to achieve high absorption coefficients. If silicone oil is chosen as the saturating fluid, the impedance ratio η_1 is then around 0.66. Figure 15 shows that in order to have the thinnest possible plates without any cavity, the acoustic Reynolds number needs to be around 1 to 2. The plate thickness is then around one-tenth of the wavelength of silicone oil, or 20 cm. If silicone oil of viscosity 1000 cS is chosen and the porous plate is kept around 2 kN s m^{-4}, a valid β is 1.6 for frequencies around 500 Hz. So β=1.6 is taken as an initial value for cases where there is a cavity. Figure 16 shows the effects of both the plate thickness and the cavity depth. It suggests that a 6-mm-thick plate with a 1-mm-deep cavity can achieve an absorption coefficient of 0.8 for sound in castor oil at 500 Hz. This has reduced the mounting space by at least 95%.

The foregoing analysis has shown that it is possible to build an effective sound absorber for low frequencies by making use of viscous fluids in porous metals. The graph in Fig. 17 shows the absorption coefficient of a porous metal absorber saturated with silicone oil and fitted with an air

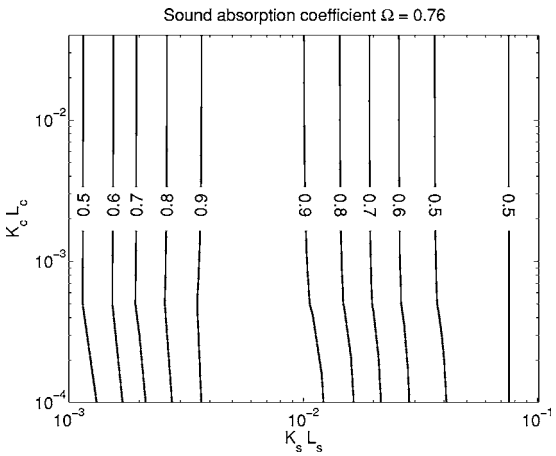


FIG. 16. Sound absorption coefficient contour as a function of k_sL_s and k_cL_c. A porous panel of Ω=0.76 and χ=3 is saturated with silicone oil. Cavity is behind the plate.

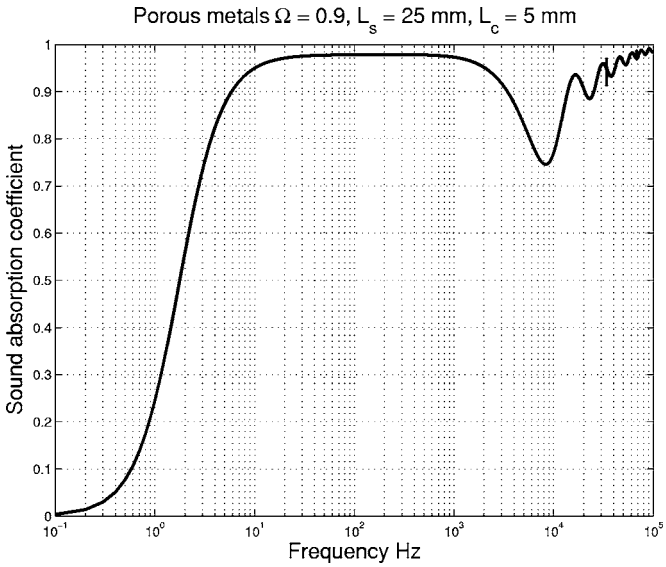


FIG. 17. The absorption coefficient curve of a porous plate 25 mm thick saturated with silicone oil with a 5-mm-deep air cavity behind.

cavity. It shows that a 3-cm-thick absorber can absorb underwater sound with coefficients above 0.8 over a wide range of frequencies, from as low as 4 Hz to as high as 100 kHz. The thickness of the absorber is only around a hundredth of the wavelength of the underwater sound.

V. CONCLUSION

This paper shows how effective a porous metal absorber filled with a specific viscous fluid and fitted with a cavity can be for the absorption of underwater sound at low frequencies. The viscous fluid serves to increase viscous dissipation, while the cavity decreases the fundamental resonance frequency, so that the mounting space can be significantly reduced. Nondimensional analysis can help to identify vital characteristics. This analysis proposes a way to build underwater absorbers to meet specific requirements. An absorber constructed in this way can reduce the thickness of the whole structure by at least 80%, compared to a perforated panel simply placed in water. Moreover, the absorber can cover a wide range of frequencies. This analysis also shows that higher porosity and smaller pore size does not always produce a higher sound absorption coefficient. Furthermore, what is discussed in this paper is not limited to underwater sound absorption. Findings here can be applied to the design of sound absorbers in other fluids.

ACKNOWLEDGMENTS

This work was supported by the National Basic Research Programs of China (Grant Nos. 2006CB601204 and 5131503). It is also listed as one of the Director-approved and funded projects in the Institute of Acoustics (Grant No. GS07SJJ17). The author wishes to thank X.P. Tong of the Institute of Acoustics for his help in obtaining the experimental results in Figs. 12 and 14, and H. P. Tang and J. L. Zhu of the Northwest Institute for Non-Ferrous Metal Research for providing fibrous metal samples in Fig. 14. The inspiration and help from W. S. Loke is greatly appreciated.

The author is also grateful for the support from J. Tian and K. Liu of the Institute of Acoustics.

[1]J. Banhart, "Characterization and application of cellular metals and metal foams," Prog. Mater. Sci. **46**, 559–632 (2001).
[2]G. Cheng, D. He, and G. Shu, "Underwater sound absorption property of porous aluminum," Colloids Surf., A **179**, 191–194 (2001).
[3]Y. Wang, "Influence of pore structures on the underwater sound absorbing properties of open-pore foamed aluminum," Development and Application of Materials **16**, 16–18 (2001) (in Chinese).
[4]C. Zwikker and C. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).
[5]K. Attenborough, "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–227 (1982).
[6]D.-Y. Maa, "Microperforated-panel wideband absorbers," Noise Control Eng. J. **29**, 77–84 (1987).
[7]D. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[8]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[9]J. Allard, *Propagation of Sound in Porous Media* (Elsevier, London, 1993).
[10]M. A. 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).
[11]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[12]P. Göransson, "Acoustic and vibrational damping in porous solids," Philos. Trans. R. Soc. London, Ser. A **364**, 89–108 (2006).
[13]L. Gibson and M. Ashby, *Cellular Solids* (Cambridge University Press, Cambridge, UK, 1997).
[14]X. Wang and T. Lu, "Optimized acoustic properties of cellular solids," J. Acoust. Soc. Am. **106**, 756–765 (1999).
[15]B. Enflo and C. Hedberg, *Theory of Nonlinear Acoustics in Fluids* (Kluwer Academic, Dordrecht, 2002).
[16]R. Beyer, *Nonlinear Acoustics* (U.S. Naval Sea Systems Command, Washington, DC, 1974).
[17]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1969).
[18]D. Albert, "A comparison between wave propagation in water-saturated and air-saturated porous materials," J. Appl. Phys. **73**, 28–36 (1993).
[19]F. Albracht and G. Lotze, in *Acoustical Investigation of Sintered Highly Porous Metal Fibre Structures as Sound Absorbers*, edited by J. Banhart, M. Ashby, and N. Fleck (MIT, Berlin, 1999), pp. 409–412.

Monitoring road surfaces by close proximity noise of the tire/road interaction

S. E. Paje,[a] M. Bueno,[b] F. Terán,[c] and U. Viñuela[d]
Laboratory of Acoustics Applied to Civil Engineering (LA²IC), Department of Applied Physics, E.T.S.I. Caminos and E.S. Informática, Universidad de Castilla-La Mancha (UCLM), 13071 Ciudad Real, Spain

(Received 26 January 2007; revised 2 July 2007; accepted 3 July 2007)

Applied acoustics is becoming an important field for civil infrastructure and environmental assessment, and road maintenance or rehabilitation strategies. In this research LA²IC has developed a GPS-based measurement techniques and apparatus on a test vehicle, for monitoring the acoustical properties of different road pavement surfaces with a reference tire. A field test on PA-12 Spanish porous pavement found in Ciudad Real is developed. The test procedure, a modification based upon the close-proximity method (CPX), relies on the use of three standard microphones situated very close to the tire/road contact patch. This procedure allows the simultaneous measurement of the sound emission synchronized to a GPS receiver, which permits tracking of the position of the sound emission. Geo-referenced sound spectra for every 10 m during individual passes of the test vehicle are analyzed to determine the tire/road noise emissions from tire/PA-12 pavement interaction. Noise levels of around 102 dB(A), with a variability of approximately 0.6 dB(A), are found at a reference vehicle speed of 85 km/h. The frequency spectrum analysis over the test section shows noticeable differences for frequencies above 1 kHz, where the tire/road noise generation mechanisms are dominated by air pumping.
 [DOI: 10.1121/1.2766777]

PACS number(s): 43.50.Lj, 43.50.Ed, 43.50.Rq [KA] Pages: 2636–2641

I. INTRODUCTION

Acoustic evaluations of field performance of road surfaces, pavement construction techniques, pavement rehabilitation, and maintenance treatments are a growing necessity for road and environmental administrations. Pavement monitoring and evaluation are of great interest in Castilla-La Mancha, Spain, especially in pavement maintenance management because of the amount of vial infrastructure construction in this region. The in-service performance of the road surface depends on consistent, cost-effective, and accurate monitoring of condition for early scheduling of repair and maintenance. Nondestructive methods, such as wave propagation, vibration, ultrasonic, electromagnetic, or electrical resistivity methods, have played a major role in pavement condition monitoring assessments and evaluation in the past. In the present day, the satisfactory joint use of global positioning systems (GPS) and acoustic methods in pavement evaluation could allow the integration of acoustic parameters in the pavement monitoring systems. Acoustic parameters like A-weighted sound pressure levels, measured in close proximity to the tire/road contact patch, or dynamic absorption measurements could be used to evaluate the state of maintenance of a road surface, to check compliance with a noise specification for a pavement surface, or to assess the acoustic performance of road surfaces.[1–4] Sound emission and absorption could reflect the real state of maintenance and conservation of road surfaces from the acoustics point of view. Therefore, all of these aspects could be crucial for an efficient road management.

The aims of this research are to contribute through geo-acoustic characterization of road surfaces to the control of traffic noise and to the assessment of road surfaces.

This research for monitoring acoustical characteristics of road surfaces is inspired in some of the measurements and techniques proposed in the European Framework Project SILVIA. It verifies that some of the procedures proposed in the project give interesting and valuable information about road surfaces.

Preliminary investigations were made to develop a methodology for geo-referenced measurements over road pavements.[5]

In this study we use a test vehicle and monitoring techniques based on continuous tire/road close proximity measurements developed in LA²IC. The mobile monitoring system consists of a four-channel sound level meter, a computer program to manage the measured data, and a GPS receiver. Close proximity sound levels (L_{CPtr}), together with the corresponding vehicle speed and global position coordinates registered from the GPS receiver module, are continuously recorded for a reference tire running in the nearside wheel track. The L_{CPtr} are normalized to the selected reference speed following the procedure described by Sandberg and Ejsmont[4] or in ISO/CD 11819-2.[6] The A-weighted sound levels are geo-referenced because of the need for road and environmental administrations for linking and correlating different variables such as pavement materials, mix formulation, acoustic parameters, surface characteristics, or environ-

[a]Author to whom correspondence should be addressed. Electronic mail: santiago.exposito@uclm.es
[b]Electronic mail: moises.bueno@uclm.es
[c]Electronic mail: fernando.teran@uclm.es
[d]Electronic mail: urbano.vinuela@uclm.es

mental noise. Thus, the geo-referenced sound pressure levels are suitable for their inclusion in a geographic information system (GIS).

II. DESCRIPTION OF THE EXPERIMENTS

A. Road surface and location of the test site

Geo-referenced close proximity tire/road noise tests on the road N-401 in Ciudad Real (National Road Network in Spain) were performed. The test sections for this study were approximately 5 km long, from N-401 kilometric post 181 to N-401 kilometric post 186. Part of the test section over which the acoustical measurements were made, as indicated in Fig. 1(a), consists of approximately three straight segments between 1000 (KP 186–185, KP 185–184, and KP 184–183) and 2000 m long (KP 183–181) and 9 m wide. The mobile monitoring was done at a speed of 85 km/h.

The test sections selected for the acoustical measurements were chosen so that no buildings, guard rails, barriers, or other acoustics reflecting objects were present in the surrounding area, as is shown in Fig. 1(a).

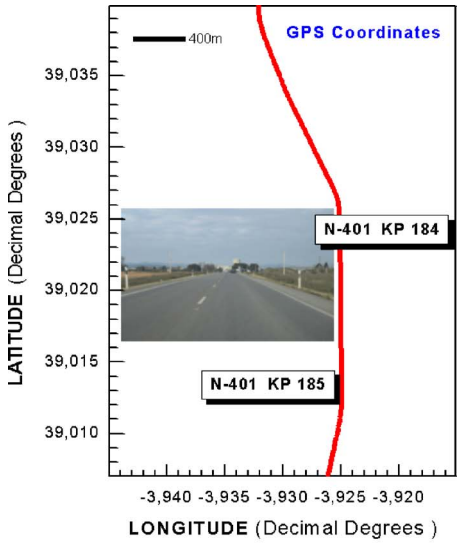
The surface of the test sections consisted of a porous bituminous layer, Spanish national designation PA-12, 4 cm thick on average, with a 10-mm maximum grain size and air void content of around 20%. It was laid 8 years before the measurements for this research. Figure 1(b) shows the tested surface at two different locations with their corresponding GPS coordinates.

B. Monitoring method of the close proximity noise

The operating principle was to continuously collect noise levels (spectral values $\frac{1}{3}$ oct of A-weighted sound pressure levels), instantaneous vehicle speed, and global positions of the tested surface, while the reference tire of our test vehicle, a standard passenger car Ford Mondeo, was rolling.

The sound pressure levels emitted by the rolling of a reference tire, which represents one of the acoustic characteristics of the surface, were measured over a distance of around 5 km by three $\frac{1}{2}$-in. microphones located close to the tire. The measurement setup involved the mounting on a self-powered vehicle with its rear right tire as the reference tire, three microphones such that the distance horizontally from the plane of the nearest tire sidewall is 20 cm, and at a height of 10 cm above the road pavement surface. The microphone positions were numbered from 1 to 3. Thus, microphone 3, M3, was situated at an angle of 45 deg to the rolling direction and microphones 2, M2, and 1, Ml, were at angles of 90 and 135 deg, respectively. A diagram of the microphone positions for the measurement test is shown in Fig. 2. The three microphones, with their windscreens, were fixed in cylindrical tube holders with the purpose of reducing microphone vibrations and flow turbulence noise. The choice of microphone positions, especially the 45 deg positions, was made on the basis of their high sensitivity to changes in the surface characteristics of the road.[4]

The acoustical measurements were carried out by using microphones, BSWA MP201, meeting the requirements of IEC 61672 Class 1 and a portable NI CompactRIO Control and Acquisition System with a four-channel dynamic signal acquisition module, located in the rear of the test vehicle. A cRIO Mobile Module (GPS) allowed global position determination and was used to store the location of geo-referenced noise levels at regular intervals in order to create a track file. These data could be used for future postprocessing in GIS software. The frequency analyses in third-octave bands and the sound pressure level measurements were car-



(a)

(b)

FIG. 1. (Color online) (a) View of the global position (GPS) coordinates of the test site located between kilometric posts K P 182 and K P 185 of the road N-401 near Ciudad Real City, and photograph of the test section. (b) Representative photographs of the tested road surface at Ciudad Real, Spain, road N-401, with PA-12 porous asphalt pavement.

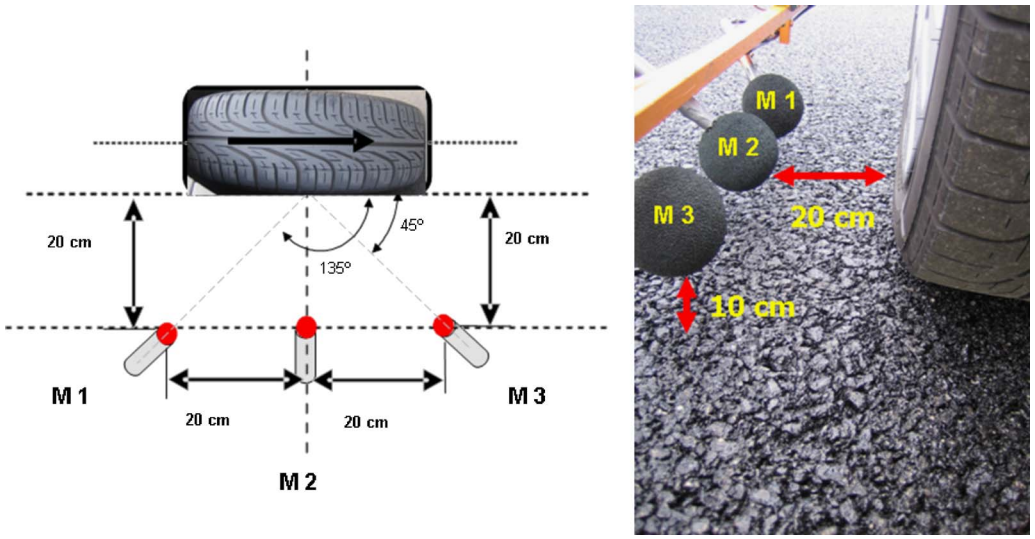


FIG. 2. (Color online) Microphone setup used for the test measurements. The tread pattern of the reference tire Pirelli P6000 205/55 R16 is also shown.

ried out with frequency weighting A, in the range from 150 Hz to 16 kHz. The sensitivity of the whole acoustic measurement setup was checked with an acoustic calibrator 4231 B&K, meeting the requirements of IEC 60942 Class 1, before and after the measurement over the tested PA-12 road surface. A digital tachometer was used to continuously measure the vehicle speed. This system was equipped with an infrared ray optical detector situated near the rolling tire. In addition, the ground positioning satellite (GPS) system was used to continuously record the vehicle speed. During the measurements, the vehicle speed was kept as close as possible to the chosen reference speed. Other parameters that were measured during the test were the air temperature and the wind speed. The average temperature of the air and the wind speed, representative of the entire road test section during the measurements, were around 26 °C and below 5 m/s, respectively.

In this study the reference tire was a regular Pirelli P6000 205/55 R16 V (see Fig. 2). It was used because it represents an important type of tire that is currently in use in standard cars in Spain and because it could be sensitive to absorption variations and texture or other surface characteristics changes. The inflation pressure of the test tire is 240 kPa in cold condition and the tire had run less than 40 000 km.

The measurements were carried out on a time interval in which the influence of other vehicles over the acoustical measurements and disturbances from passing vehicles was practically nonexistent. Before the test measurement, the reference tire was warmed up by driving for more than 20 min, and the fourth gear was used during the measurement in order to assure that the influence of the power unit of the vehicle was the lowest possible.

III. FIELD MEASUREMENTS AND DISCUSSION

As the noise emission from tire/road interaction is speed dependent, deviations of the test vehicle from the reference speed lead to variations of the noise emission levels for a nominal speed. Because deviations in speed are inevitable and with the aim of normalizing the results to the selected reference speed, fixed in this research to 85 km/h, the sound pressure levels were speed corrected, L_{corr}, following the procedure described, for example, by Sandberg and Ejsmont:[4]

$$L_{\mathrm{corr}}(t) = L_{\mathrm{meas}}(t) - B \log_{10}\left(\frac{V(t)}{V_{ref}}\right), \tag{1}$$

where L_{meas} is the averaged tire/road sound level (mics 1 and 3), V_{ref} is the selected reference speed (85 km/h), $V(t)$ is the current speed, and B is the speed constant (noise-speed slope).

The speed constant B, for the test surface PA-12 of road N-401 and for the reference tire, was obtained from the test measurements, calculating the slope of the logarithmic regression of the sound pressure levels and speeds (see Fig. 3). A road segment where the speed values cover a spread of 20 km/h was selected for the measurements. The coefficients A and B, obtained from the linear relationship

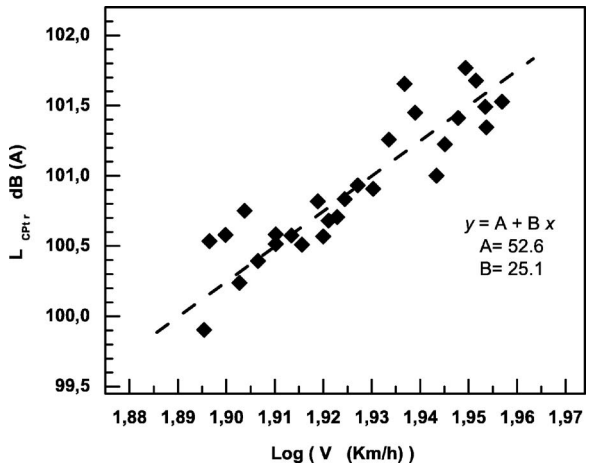


FIG. 3. Noise vs. logarithm of speed characteristics and linear regression of averaged sound pressure levels (mic 1 and mic 3) for the reference tire rolling on PA-12 porous layer. The analysis gives the following relationship: $L=A+B\log_{10}(V\ (\mathrm{km/h}))$, with coefficients $B=25.1$ and $A=52.6$.

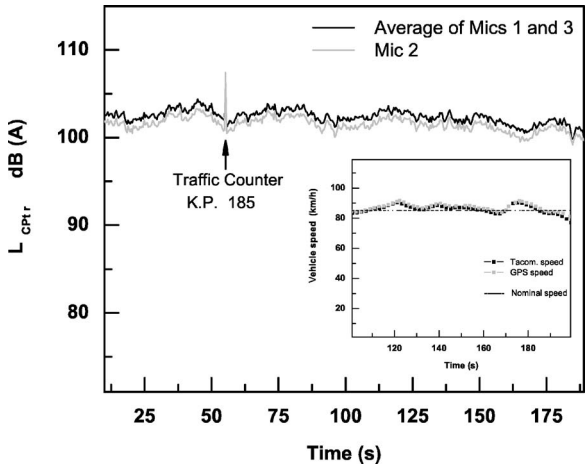


FIG. 4. Fluctuations in time of averaged noise levels from microphones 1 and 3 over the test sections at road N-401 Ciudad Real, expressed as SPL dB(A) and rolling at around 85 km/h. Inset shows variations of speeds from the digital tachometer and by the GPS over a road segment defined by global positions (Latitude$_1$=39.0248/Longitude$_1$=−3.9251) and (Latitude$_2$=39.0456/Longitude$_2$=−3.9319).

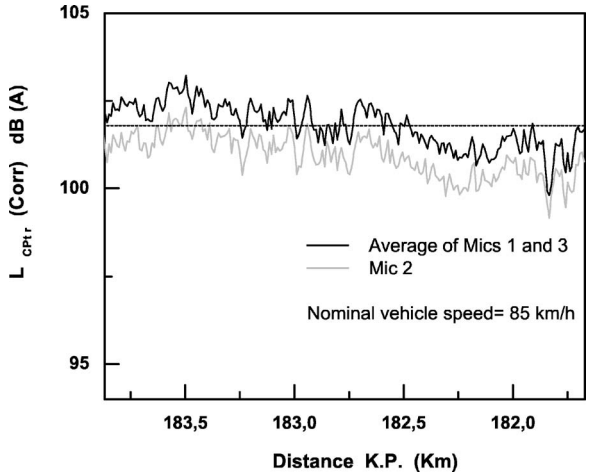


FIG. 5. Variations with distance (referenced distance to Kilometric Point) of the averaged noise level from microphones 1 and 3, expressed as speed corrected close proximity sound levels dB(A), for a reference speed of 85 km/h. Dashed lines show the mean value 102 dB(A).

$$L_{\mathrm{CPtr}} = A + B \log_{10}(V), \tag{2}$$

are about 53 and 25 dB(A), respectively. This result agrees well with those obtained from laboratory drum measurements, at the Technical University of Gdansk.[4] The experimental data show that the relation between the coefficients A and B is linear, independent of the type of surface studied. In these laboratory drum measurements, the microphone distance is very much like the microphone distances fixed in this research. This similarity could explain why coefficient A is so closed for a coefficient B of around 25.

A. Geo-referenced global close proximity noise fluctuations

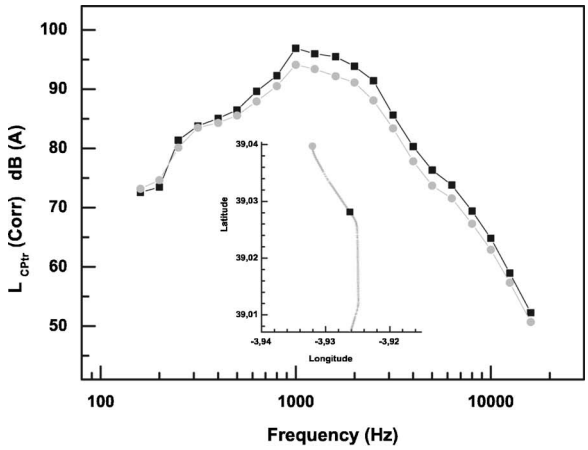
Figure 4 shows the fluctuation of the typical tire/road noise levels from arithmetically averaged microphones 1 and 3, continuously measured on the PA-12 porous pavement lengthways test segments of the road defined by global position coordinates (GPS) shown in Fig. 1. The evolution of the level of the central microphone, M2, is also shown. The tests were conducted using the Pirelli P6000 tire at a vehicle speed of around 85 km/h. The vehicle speed, monitored by the digital tachometer and by the GPS system over a segment of the test section, is shown in the inset of Fig. 4. This test section is defined by the latitude/longitude coordinates system of 39.0248/−3.9251 and 39.0456/−3.9319, in decimal degrees. The tests were performed directly in the wheel track and were developed without the presence of road traffic, which could disturb the measurements. As can be seen in the inset of Fig. 4, there are deviations between the current speed and the nominal test speed, unavoidable in this kind of measurement.

The results of the measurements in Fig. 4 indicate that PA-12 porous pavement, in road N401-Ciudad Real, provide continuous tire/road noise levels, L_{CPtr}, between 100 and 104 dB(A) at speeds of between 85 and 90 km/h. These kinds of geo-referenced acoustic data, which show the tire/road sound level profile of a given road section, could be used for assessing the acoustic performance of the road surface, especially near populated areas by environmental administrations.

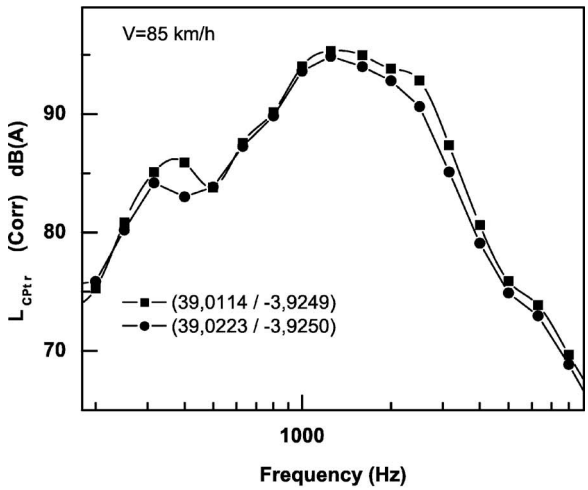
With the intention of reporting the acoustic characteristics of the PA-12 pavement near Ciudad Real, at 85 km/h nominal reference speed, the above-mentioned method of correcting for speed deviations is applied. Thus, Fig. 5 shows the averaged (mics 1 and 3) tire/road noise fluctuations of the test segment defined by coordinates 39.0248/−3.9251 and 39.0456/−3.9319, after speed correction by Eq. (1), as a function of the distance. The acoustic data are referenced by means of distance between kilometer posts. These averaged global noise levels [dB(A)] are slightly higher than those for typical porous asphalt pavement surface CPX levels reported by other authors. This could be for two reasons: the aging and the wear of the PA-12 pavement and, also, because the reference tire is not a new tire. The tread depth was reduced in our reference tire, which decreases the air drainage between the tire and the road surface. On the other hand, the acoustical monitoring test indicates that the highest noise levels correspond to the first section of the measurements. In spite of that, the evolution of noise levels of the PA-12 pavement surface reflects a relative surface homogeneity of road N-401 between kilometric posts 181 and 184.

B. Geo-referenced spectral analysis

All tire/road noise emissions from tire/PA-12 interaction were analyzed in each third octave band between 150 Hz and 16 kHz over the test section at 85 km/h reference speed. An example of tire/road spectrum, measured on the porous asphalt PA-12 with the reference tire, is shown in Fig. 6(a) for different localizations. These locations of the test section are indicated in the inset of figure. Each third octave band level derives from the average of microphones 1 and 3, speed corrected following Eq. (1). Spectral differences are noticeable above 1 kHz, where the tire/road noise emission from tire/PA12 interactions is mainly due to aerodynamic effects over the surface such as air pumping.[4,7–9] In the range of



(a)



(b)

FIG. 6. (a) Close proximity noise spectrum at 85 km/h on porous asphalt PA-12 for Pirelli P6000 205/55/R16 V tire at different locations inside the test section (see inset of figure). (b) Noise spectra, in the same conditions, at positions (Latitude$_1$=39.0114/Longitude$_1$=-3.9249) and (Latitude$_2$=39.0223/Longitude$_2$=-3.9250).

frequencies between 150 and 500 Hz, sound pressure levels show good similarity. At these frequencies, the noise emission is mainly due to the surface roughness or texture pattern. For this reason, this behavior could indicate similar roughness over the section tested, due to a very similar maximum stone size in the PA-12, or a similar surface texture pattern of the road. On the other hand, spectral differences above 1 kHz could be attributed to a different reduction/increment of air pumping noise by the degree of porosity. Figure 6(b) shows the spectrum at global position coordinates defined in Fig. 1(b). The photographs display two locations where the surfaces present a different degree of polish and clogging. In the two locations, the layers have the same maximum chipping size but different binder and void content. These parameters are reflected directly in the acoustic characteristics of road surfaces.[10,11] Due to the polishing, the sharpness of the chippings and the amplitude are reduced [see figures of surface defined by (39.0114/−3.9249)]. As Sandberg and Ejsmont[4] denote, this affects the air pumping noise generation mechanism. This led to an increase in levels at high frequencies, as Fig. 6(b) shows.

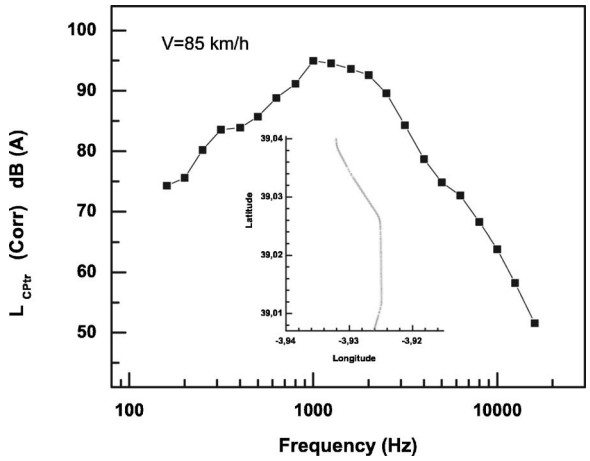


FIG. 7. Averaged close proximity noise spectrum, measured at 85 km/h on the porous asphalt PA-12 with Pirelli P6000, 205/55 R16 V tire, for the selected test section described in the inset of figure.

It is clear that spectral analysis reflects slight differences depending on the location over the test section. Figure 7 shows the averaging of the third-octave-band frequency spectra, measured over a number of road segments (see inset of figure), after speed corrections applied in a similar way to that for the overall A-weighted levels. This could be reported as the representative frequency spectrum of the PA-12 pavement over the tested road.

C. Variability of the sound emission

Another parameter that has to be considered in the acoustical characterization of road surfaces is the variability. In this research, the variability of the close proximity noise of the tire/road interaction is studied along a segment of 2 km of the road N-401 with a PA-12 surface, as is shown in Fig. 1. The 2-km test section is subdivided into segments, each of length 10 m. In each of these road segments, the A-weighted sound level is measured for the reference tire used in this research and for a nominal speed of 85 km/h. The variation of the noise level, after speed correction, as is shown in Fig. 5, could be considered to be an indication of the surface homogeneity or the state of maintenance of the porous pavement of the test section. The average sound level over the test section defined by latitude/longitude coordinates of (39.0248/−3.9251) and (39.0456/−3.9319) is 102 dB(A). The standard deviation around this mean, 0.6 dB(A), could be expressed as the A-weighted sound emission variability of the PA-12 surface over the test section of the road N-401. On the other hand, for the above-mentioned reference vehicle speed, the total range of noise level variation within the selected sections is around 1.7 dB(A). This result is supported by the existence of locations with a different state of maintenance, as is evident from Fig. 1(b).

IV. CONCLUSION

The measurement techniques and test vehicle described in this paper, whereby the functional and intrinsic acoustic qualities of the N-401 road surface are determined in a georeferenced way, offer road network management administra-

tion an important criterion for the judgment of the state of the road network. In addition, the geo-referenced acoustical information recorded by this system could allow us to create tire/road noise maps. These acoustics maps will contain the noise effects of the road surfaces. These results could be used for environmental purposes, for instance in road noise maps and for other road network administration purposes in pavement rehabilitation strategies.

Geo-referenced sound spectra, generated by tire/road interaction on porous asphalt (PA-12 Spanish denomination), show global values of around 102 dB(A) with a variability of 0.6 dB(A) over a distance of 2 km, between kilometer post 184 and 181. The reference nominal speed used in this research is 85 km/h. This variability is attributed to a different degree of the clogging of air voids and the state of maintenance in the pavement studied, which could be visually confirmed. This essentially affects the air pumping noise generation mechanism.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science (Junta de Comunidades de Castilla-La Mancha) through Grant No. PCI05-025. We thank Juan Antonio Mesones, head of *Unidad de Carreteras en Ciudad Real*, General Directorate of Roads of the Spanish Ministry of Infrastructure and Transport, AEPO, and Aglomancha for their valuable assistance.

[1]C. Padmos, P. A. Morgan, P. Abbott, G. van Blokland, M. S. Roovers, W. Bartolomaeus, and F. Anfosso-Lédée, "Classification scheme and COP method," SILVIA Project Deliverable SILVIA-dww-025-14-wp2-141005 (2005).

[2]J. A. Reyff and P. Donavan, "Studies of the effect of aging of quiet pavements on tire/pavement noise (A)," J. Acoust. Soc. Am. **118**(3), 1915 (2005).

[3]P. A. Morgan and G. R. Watts, "A novel approach to the acoustic characterization of porous road surfaces," Appl. Acoust. **64**(12), 1171–1186 (2003).

[4]U. Sandberg and J. A. Ejsmont, "Tyre/road noise reference book," Informex, Harg, SE-59040 Kisa, Sweden (2002).

[5]S. E. Paje, U. Viñuela, F. Terán, S. Lopéz-Querol, and A. Sanz, "Georeferenced near field noise assessment of road N-401 in Castilla-La Mancha," Tecniacústica 2006-37 Congreso Nacional de Acústica, Encuentro Ibérico de Acústica, EAA European Symposium on Hidroacoustics, Gandia (2006) (in Spanish).

[6]ISO/CD 11819-2, 2000-12-13 "Acoustics-measurement of the influence of road surfaces on traffic noise-Part 2: the close-proximity method," Draft standard of working group ISO TC43/SCI/WG33.

[7]P. M. Nelson, "Designing road surfaces to reduce traffic noise," Transport Research Laboratory, Annual Review, Crowthorne (1994).

[8]M. Heckl, "Tire noise generation," Wear **113**, 157–170 (1986).

[9]S. Kim, W. Jeong, Y. Park, and S. Lee, "Prediction method for tire air-pumping noise using a hybrid technique," J. Acoust. Soc. Am. **119**(6), 3799–3812 (2006).

[10]M. C. Berengier, M. R. Stinson, G. A. Daigle, and J. F. Hamet, "Porous road pavements: Acoustical characterization and propation effects," J. Acoust. Soc. Am. **101**(1), 155–162 (1997).

[11]W. K. Liu and K. M. Li, "A theoretical study for the propagation of rolling noise over a porous road pavement," J. Acoust. Soc. Am. **116**(1), 313–322 (2004).

Annoyance due to single and combined sound exposure from railway and road traffic

Evy Öhrström,[a] Lars Barregård, Eva Andersson, Annbritt Skånberg, Helena Svensson, and Pär Ängerheim
Department of Occupational and Environmental Medicine, The Sahlgrenska Academy at Göteborg University, Box 414, SE-405 30 Göteborg, Sweden

(Received 19 January 2007; revised 19 June 2007; accepted 27 August 2007)

Environmental noise is a growing and well recognized health problem. However, in many cases people are exposed not to a single noise source—for example, road, railway, or aircraft noise—but to a combination of noise exposures and there is only limited knowledge of the effects on health of exposure to combined noise sources. A socio-acoustic survey among 1953 persons aged 18–75 years was conducted in residential areas exposed to railway and road traffic noise with sound levels ranging from $L_{\mathrm{Aeq,24h}}$ 45–72 dB in a municipality east of Gothenburg, Sweden. The objectives were to assess various adverse health effects, including annoyance, and to elucidate the impact of exposure to single and combined noise sources. In areas exposed to both railway and road traffic, the proportion annoyed by the total traffic sound environment (*total annoyance*) was significantly higher than in areas with one dominant noise source (rail or road traffic) with the same *total sound exposure* ($L_{\mathrm{Aeq,24h,tot}}$). This interaction effect was significant from 59 dB and increased gradually with higher sound levels. Effects of the total sound exposure should be considered in risk assessments and in noise mitigation activities. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2785809]

PACS number(s): 43.50.Qp [BSF] Pages: 2642–2652

I. INTRODUCTION AND AIM

Community noise is an environmental stressor with documented short-term effects (e.g. annoyance, interference with speech communication, and disturbance of sleep and rest) and long-term consequences for health and well being (WHO, 2000; Öhrström, 2004). In contrast to many other environmental problems, noise pollution is still growing. In residential areas, road traffic is normally the dominant noise source; and road traffic in Sweden has been predicted to reach 120–130% of 1997 levels by 2010 (SIKA, 2000, 2002). A meta-analysis conducted by Miedema and Oudshoorn used results from large data sets to examine the relationship between annoyance and exposure to noise from a single source (road traffic, aircraft, or railways) (Miedema and Oudshoorn 2001); the analysis showed that road traffic noise causes more extensive annoyance reactions than railway noise, and so a " bonus" of 5 dB has been applied to railway noise in most European Union (EU) countries, e.g. Austria, Germany, France, Switzerland, and Sweden (EU Commission 2002). Some of the more recent studies on railway and road traffic noise, not included in the meta-analysis by Miedema and Oudshoorn (2001), show no support for a railway bonus (e.g., Kaku and Yamada, 1996; Yano *et al.* 1998; Lercher *et al.*, 1999; Kurra *et al.*, 1999; Morihara *et al.*, 2002) while other studies (Lambert *et al.*, 1998; Moehler *et al.*, 2000; Vos 2004) give support for a railway bonus. Although a large part of the population is exposed to environmental noise from more than one source, there is little knowledge of the relationship between annoyance and noise exposure when more than one noise source is present. There is no standardized or commonly accepted method for assessing the total noise impact from combined sources, although several models have been proposed, such as the energy summation model and the dominance model (Ronnenbaum *et al.*, 1996; Nilsson, 2001; Miedema, 2004).

Only a few larger studies have investigated annoyance from multiple noise sources. Ronnenbaum *et al.* (1996) found support for the dominance model in their analysis of data from a large German study involving 1740 respondents exposed to road traffic and railway noise. In areas with one of the two sources dominant, annoyance due to this source exceeded total annoyance, which in turn exceeded annoyance due to the other source. Joncour *et al.* (2000), Cremezi *et al.* (2001), and Champelovier *et al.* (2004) presented results from a French study performed in 1997 involving 62 study sites exposed to both road traffic and railway noise. In total, 664 people were interviewed: 126 in railway dominated sites, 282 in road traffic dominated sites, and 256 in sites with no single dominant noise source. Sites were defined as "railway dominant" or "road traffic dominant" if the difference in sound level between the two sources was larger than 5 dB. The results gave some support for the dominance model. The probability of being disturbed by railway noise increased with increased sound level. A potential interaction was observed between annoyance due to railway traffic and annoyance due to road traffic; however, this interaction was rather complex, making it difficult to draw any firm conclusions on interaction effects. There is, then, a need for further

[a]Electronic mail: evy.ohrstrom@amm.gu.se

studies on the health effects of multiple noise sources, and further consideration of ways to handle noise exposure from more than one source at a time.

A large socio-acoustic survey was conducted in 2004 among people ($n=1953$) living in residential areas exposed to road traffic noise, railway noise, and also to some extent aircraft noise (L_{Amax} levels below 70 dB) in the municipality of Lerum, east of Gothenburg, Sweden. The aims of this survey were to assess the prevalence of short-term and long-term health effects of noise exposures, to develop a method for determining individual noise exposure based on geographical information systems (GIS) techniques, and to examine and propose new indicators suitable for following up the prevalence of health-related effects of traffic noise. The main results (see main report Öhrström *et al.*, 2005a, 2005b) demonstrated extensive, dose-related effects of noise in terms of annoyance, activity, and sleep disturbances, as well as stress-related psycho-physiological symptoms. Among men, particularly those who had lived in their present homes for more than 10 years, the incidence of hypertension and use of hypertension medication increased with increasing road traffic noise.

This article presents results from this survey on general annoyance resulting from road traffic noise, railway noise, and the total traffic sound environment. Noise exposure was calculated for railway and road traffic separately, and the energy summation model was applied to assess the total noise exposure from road and railway traffic. Noise annoyance was assessed with separate questions for the two main noise sources and for the total traffic sound environment.

II. MATERIALS AND METHODS

A. Background and study area

At the request of the Municipality of Lerum, an acoustic consultant (Ingemansson Technology AB) performed a mapping of noise exposures from different sources (road traffic, railway, and aircraft noise, along with certain stationary noise sources) in Lerum in 2003. The calculations were made on a grid with a distance of 15 m between each calculation point. In the spring of 2004, a socio-acoustic study was performed to provide knowledge on the health effects of noise, as a basis for health-based actions against noise exposures (for details see main report Öhrström *et al.*, 2005a). The study was designed to cover a wide range of sound levels, from $L_{\mathrm{Aeq,24h}}$ 45 dB to more than 70 dB from both sound sources, with a large enough number of participants in various sound exposure categories to allow for detailed analyses of the effects of road traffic noise, railway noise, and a combination of the two noise sources. Participation was restricted to those living in residential buildings exposed to sound levels from railway and road traffic exceeding $L_{\mathrm{Aeq,24h}}$ 45 dB at the most exposed side.

The municipality of Lerum is situated east of Gothenburg, Sweden. A major part of the population lives along two major traffic routes and is exposed to both noise from road traffic on highway E20 (about 20 000 vehicles/24 h) and noise from a major railway (about 200 trains/24 h). A minor part of the study population lived in the area between the railway and the highway. The study area is approximately 20 km by 6 km (see Fig. 1). The residential buildings consist of detached, terraced, and apartment houses.

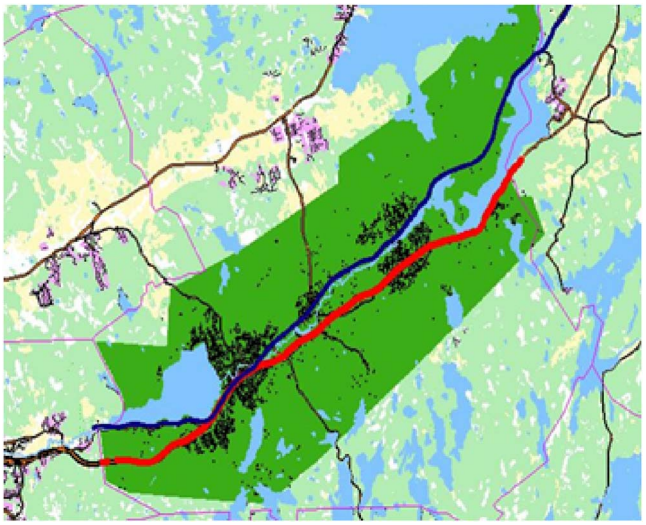


FIG. 1. (Color online) The study area with the railway (top) and the highway (bottom).

B. Assessment of sound levels

In the first phase of the study, a GIS-based method was used to determine the noise exposure adjacent to each residential building. Calculations of sound levels from road traffic (highway and all other roads) and railway noise at the most exposed side were provided for each residential building by an acoustic consultant using a validated model (Nordic Prediction Method, Jonasson *et al.*, 1996), the calculation program, Sound Plan, and the GIS program Arc View. In some cases (very small roads) for which no data on traffic volume were available, the calculated sound levels were based on a volume of 300 vehicles/24 h. All calculation points were determined at 2 m above the ground as free field values. This implies that reflections from all other façades in the area were included in the calculation for each building, excluding reflections from the façade of that building itself. In some cases (large buildings or buildings oriented with the gable facing the road or the railway) close examinations or site visits were performed in order to ensure that sound levels were calculated accurately for different apartments in the building. For each type of noise source, sound levels were calculated as $L_{\mathrm{Aeq,24h}}$ and L_{Aeq} for day (0600–1800 hrs), evening (1800–2200 hrs), and night (2200–0600 hrs), and also as L_{Amax} and L_{den}. In addition, the total sound exposure for road traffic and railway traffic was calculated as $L_{\mathrm{Aeq,24h}}$ ($L_{\mathrm{Aeq,24h,tot}}$) and L_{den}($L_{\mathrm{den,tot}}$). The assessment of the total sound exposure did not consider whether the sound levels from road and railway were obtained at the same side of the dwelling or at different sides. However, only 12% of the study population lived in dwellings situated between the railway and the highway. The majority of these residents (194 out of 250 respondents) lived in detached houses (one-family villas surrounded by gardens) with very little shielding from railway and road traffic noise. The accuracy of the noise calculations was assessed at ±3 dB.

TABLE I. Number of participants in the different road traffic and railway sound exposure categories.

		Railway noise $L_{Aeq,24h}$ (dB)				
		45–50 dB	51–55 dB	56–60 dB	61–72 dB	Total
Road traffic noise $L_{Aeq,24h}$	45–50 dB	455	192	88	27	762
	51–55 dB	294	158	89	49	590
	56–60 dB	134	126	108	66	434
	61–70 dB	42	31	46	48	167
Total	*N*	925	507	331	190	1953

C. Study population

After determining the sound levels for all residential buildings, we linked the data to the Land Registry to obtain noise exposure data for 7000 dwellings. Next, these data were linked to the local population register obtained from the Town Planning Office in Lerum municipality. Among this population, a selection was made of persons aged between 18 and 75 years who had resided at their present address for at least 6 months and who lived in dwellings with outdoor sound levels of $L_{Aeq,24h}$ 45 dB or more from both railway and road traffic. The population sample was then divided into 5 dB categories based on outdoor sound levels from separate sources ($L_{Aeq,24h}$ 45–50, 51–55, 56–60, and 61+ dB). The aim was to obtain at least 100 respondents within each exposure category; however, there were relatively few individuals in the highest noise categories, $L_{Aeq,24h}$ 61+ dB (see Table I). In all noise categories, except for 45–50 and 50–55 dB, all persons between 18 and 75 years of age were chosen. The 45–50 and 50–55 dB categories were twice as large as the categories with higher noise levels and therefore we selected only those who were born between the 1st and the 15th in each month.

In total, 2747 residents were selected for the study. Of these, 1953 responded to the postal questionnaire, giving a final response rate of 71%. The distribution of respondents among 16 combined road traffic and railway noise exposure categories in $L_{Aeq,24h}$ is shown in Table I, and the distribution of respondents among eight total noise exposure categories in $L_{Aeq,24,tot}$ is shown in Table II.

D. Evaluation of adverse effects of noise

The questionnaire was designed to evaluate perceived symptoms and health effects of traffic noise (e.g., annoyance, sleep disturbance, well being); it also included questions on psycho-physiological symptoms, hypertension, and antihypertensive drugs. The design of the questionnaire was based on previous research on the adverse health effects of noise (e.g., Öhrström, 1989, 2004) and it included batteries of questions on: (1) the living environment, residency, and various sources of nuisance; (2) annoyance and interference with various activities, both indoors and outdoors; (3) work situation, socio-demographic factors, and self-estimated noise sensitivity; (4) perceived sleep quality; and (5) general physical health and mental well being.

This article addresses only the results regarding annoyance. Annoyance caused by noise from: (i) road traffic; (ii) railway; and (iii) the total sound environment from road traffic, railway, and aircraft noise was evaluated with a five-point verbal category scale according to the ISO standardization of annoyance scales (ISO, 2003). The questions were phrased as follows: "Thinking about the last 12 months or so, when you are here at home, how much does noise from (source) bother, disturb, or annoy you?" (Alternatives: "not at all," "slightly," "moderately," "very," and "extremely"). In the presentation of the results, the "annoyed" category consists of those who were moderately, very, or extremely annoyed by a given noise source, and the "highly annoyed" category of those who were very or extremely annoyed. Additionally, annoyance caused by the total sound environment including road traffic, railway, and air traffic noise is referred to as "total annoyance" (consisting of those who were moderately, very, or extremely annoyed by the total traffic sound environment.

The postal questionnaire was distributed together with an introductory letter in February–March 2004. Those who did not return the questionnaire were sent two reminders at 10-day intervals.

E. Statistical analysis and treatment of data

The association between noise exposure and the degree of noise annoyance was analyzed using the Spearman rank order correlation coefficient (r_s). The possible differences between proportions (e.g., proportion annoyed) were tested using χ^2 tests. The McNemar test (for paired analysis) was used to analyze the level of annoyance by different sources of noise in a given individual. In the multivariate analyses of the relative risk, where several noise sources were considered as well as background covariates, logistic regression analysis (with a dichotomous response variable) was used. The annoyance due to a single source (road traffic noise and railway noise, respectively) was studied and a possible interaction between noise from railway and noise from road traffic was investigated. In the investigation of annoyance by road traffic noise, an interaction between railway noise and road traffic noise means that the effect of road traffic noise (on annoyance) is different, depending on the railway noise level. The same analysis was made for annoyance to railway noise. Another analysis was made (also using logistic regression), regarding the annoyance by the total traffic sound environment. We investigated whether the same total sound

TABLE II. Number of participants in the different total sound exposure categories.

	Total sound levels from railway and road traffic, $L_{Aeq,24h,tot}$ (dB)								
	48–50	51–53	54–56	57–59	60–62	63–65	66–68	69–72	Total
Total	142	526	403	377	260	146	66	33	1953

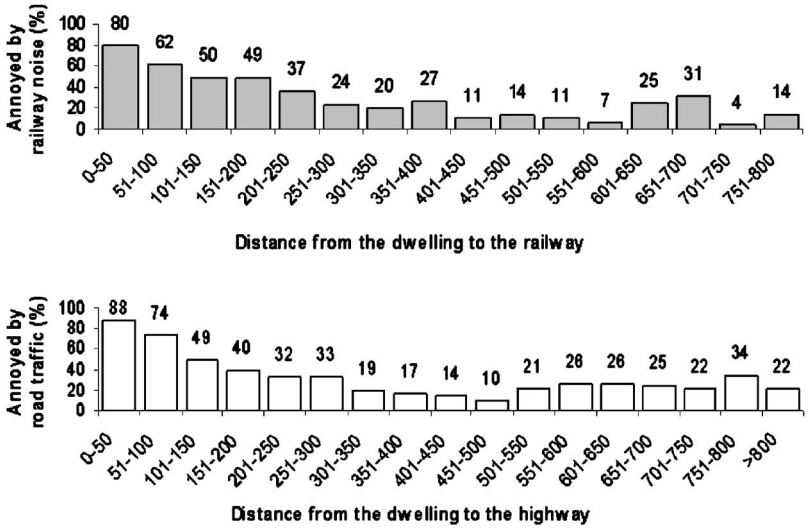


FIG. 2. Relationship between distance from the railway track and annoyance due to railway noise (top). Relationship between distance from highway and annoyance due to road traffic noise (bottom). "Annoyed" is defined as moderately + very + extremely.

exposure was perceived as more annoying when caused by two equally noisy sources (railway and road traffic) compared with one dominating source (either railway or road traffic). From the logistic regression models, the relative risk of being annoyed was estimated using the odds ratio and its 95% confidence interval. An ordinal logit model was used to estimate the dose-response relationship for different degrees of annoyance ("not at all," "slightly,…, "extremely"). Linear regression and intraclass correlation were used to analyze the associations between annoyance due to single sources and annoyance due to the total traffic sound environment. Data were analyzed using SPSS for Windows, Version 12.0.1, and SAS, Version 9.1. Differences associated with p values below 0.05 were considered statistically significant.

III. RESULTS

A. Annoyance due to single noise sources

1. Annoyance due to road traffic and railway noise in relation to distance to the railway and the highway

All respondents lived at most 800 m from the railway track, and the majority (78%) of them lived less than 800 m from the highway. Figure 2 illustrates the relationship between the distance to these main traffic noise sources and the prevalence of annoyance caused by each in isolation. Regarding railway noise (upper figure), there was a fairly good correlation between distance from the railway track and the degree of annoyance (five-point scale) caused by railway noise ($r_s=-0.40$, $p<0.001$, $n=1942$). No plausible explanation could be found for the somewhat higher prevalence of annoyance due to railway noise among residents living at a distance of 600–700 m from the railway track.

For road traffic noise (lower figure), a somewhat weaker correlation was found between annoyance and distance to the highway ($r_s=-0.26$; $p<0.001$; $n=1942$). Sound levels from highway traffic seemed to determine road traffic noise annoyance up to a distance of about 500 m from the highway, while for dwellings further away, the prevalence of annoyance due to road traffic noise was probably mainly determined by road traffic in local roads nearby. When the analysis was restricted to those living within 500 m of the highway, the correlation with road traffic noise annoyance was stronger ($r_s=-0.42$; $p<0.001$; $n=1109$). The corresponding correlation between distance and railway noise annoyance for those living within 500 m of the railway track was $r_s=-0.41$ ($p<0.001$; $n=1684$).

2. Annoyance due to road traffic and railway noise in relation to sound levels expressed in $L_{Aeq,24h}$

Figure 3 shows the dose-response relationships for different degrees of railway noise annoyance (left) and road traffic noise annoyance (right), as a function of the sound level ($L_{Aeq,24h}$). For example, we estimate that at 55 dB of railway noise, 10% were extremely annoyed by railway noise, 13% were very annoyed, 20% were moderately annoyed, 34% were slightly annoyed, and 23% were not annoyed at all. For road traffic at 55 dB, we estimate that 6% were extremely annoyed, 11% were very annoyed, 19% were moderately annoyed, 45% were slightly annoyed, and 19% were not annoyed at all. For the grouped category of "annoyed" (extremely, very, or moderately annoyed), we estimate that at 55 dB, 43% were annoyed by railway noise and 36% by road traffic noise, while at 65 dB, 77% were annoyed by railway noise and 75% by road traffic noise.

The prevalence, with 95% confidence intervals, of annoyance caused by road traffic and railway noise, respectively, is presented in Fig. 4 for different sound exposure categories. Annoyance at the lowest sound levels (45–50 dB) was similar for the two noise sources, 14%. At higher sound levels (51–55 and 56–60 dB), annoyance due to railway noise was reported significantly more frequently than annoyance due to road traffic noise ($p<0.001$). No significant difference in annoyance between the two noise sources was found at sound levels above 60 dB.

B. Annoyance and possible interaction with exposure to a second noise source

In the logistic regression models, with annoyance (from road or railway noise) as the response variable, the explanatory variables were road traffic and railway noise (catego-

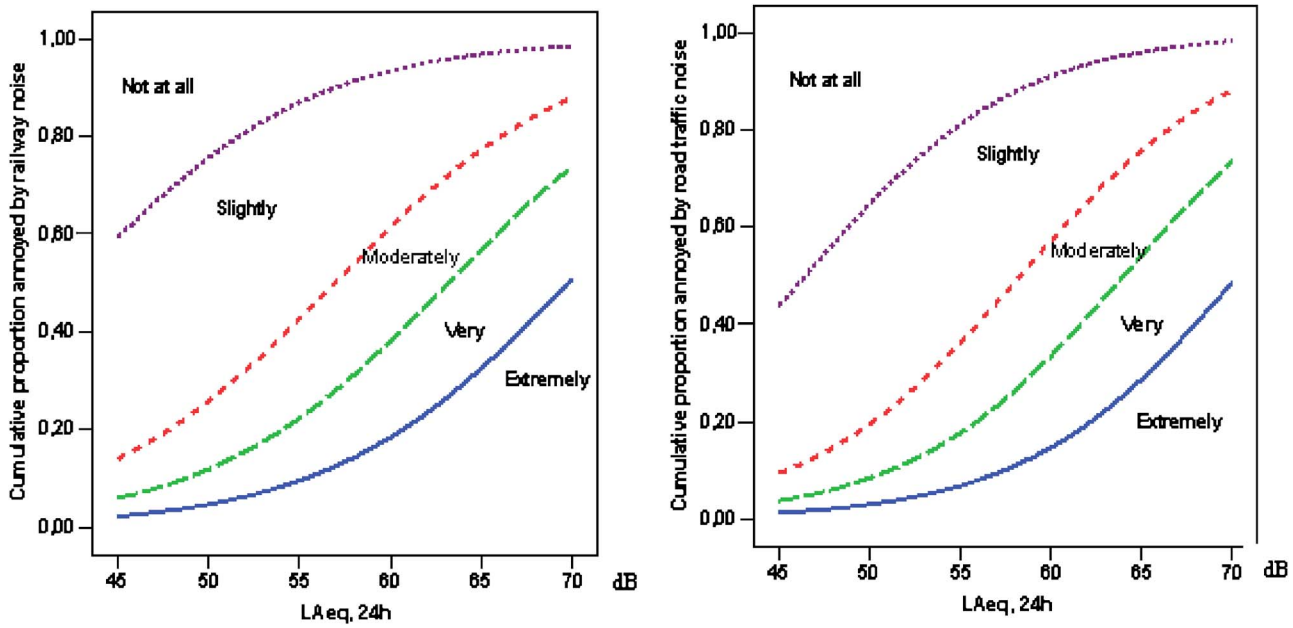


FIG. 3. (Color online) Estimated dose-response relationships for the different degrees of noise annoyance, as functions of the sound level in $L_{Aeq,24h}$ for railway noise (left) and road traffic noise (right). The area above the upper curve shows the proportion "not at all annoyed," the area between the curves shows the proportion "slightly," "moderately," or "very," annoyed, respectively, and the area below the lowest curve shows the proportion "extremely" annoyed.

rized as in Fig. 5) and interaction between road traffic noise and railway noise (operationalized by multiplying the road traffic noise variable by the railway noise variable), along with covariates (years spent in the dwelling, type of windows, bedroom window position, and noise sensitivity).

The analysis regarding annoyance due to *road traffic* noise showed a significant interaction effect for the highest road traffic exposure category ($p=0.025$). This means that the dwellings with very high road traffic noise exposure (61–70 dB) and, simultaneously, a high railway noise exposure (56–60 and 61–72 dB) show an amplified annoyance from road traffic noise. This amplified annoyance cannot be found for dwellings exposed to lower road traffic noise, e.g., the combination of road traffic noise 56–60 dB and railway noise 56–70 dB. The results are indicated from the descriptive values in Fig. 5 (left panel).

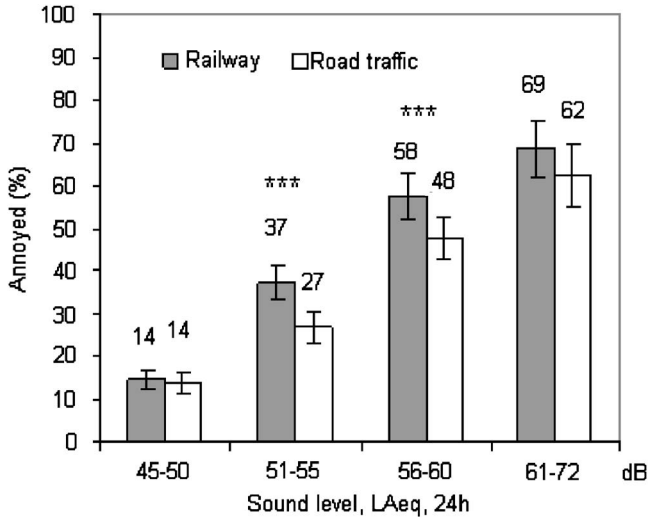


FIG. 4. Annoyance (% moderately, very, extremely annoyed, with 95% confidence intervals) due to road traffic and railway noise in relation to sound levels expressed in $L_{Aeq,24h}$.

A similar response pattern was seen for annoyance due to *railway* noise. For the two highest railway noise categories (56–60 and 61–72 dB), the prevalence of annoyance due to railway noise was amplified by high exposure to road traffic noise, with significant interaction effects in both the 56–60 dB ($p=0.036$) and the 61–72 dB ($p=0.005$) categories of railway noise. For dwellings in the 56–60 dB category of railway noise, annoyance due to railway noise was amplified by simultaneous exposure to high (56–60 dB) or very high (61–70 dB) levels of road traffic noise; while in the 61–72 dB category of railway noise, annoyance due to railway noise was amplified by simultaneous exposure to very high (61–70 dB) levels of road traffic noise. The effect of high road traffic noise could not be found for the other railway noise categories (45–50 and 51–55 dB). These results are illustraded in Fig. 5 (right panel).

Similar analyses were performed using the category of "highly annoyed," defined as "very or extremely annoyed," as the response variable. For high annoyance from road traffic noise, no interaction could be found. However, in the 56–60 and 61–72 dB categories of railway noise, the annoyance due to railway noise was more prevalent if there was simultaneous exposure to high (56–60 dB) or very high (61–70 dB) road traffic noise ($p<0.0005$ and 0.013, respectively).

C. Annoyance due to single noise sources and to the total traffic sound environment in relation to total sound exposure from railway and road traffic ($L_{Aeq,24h,tot}$)

The association between sound levels from road traffic and the degree of annoyance (five-point scale) due to road

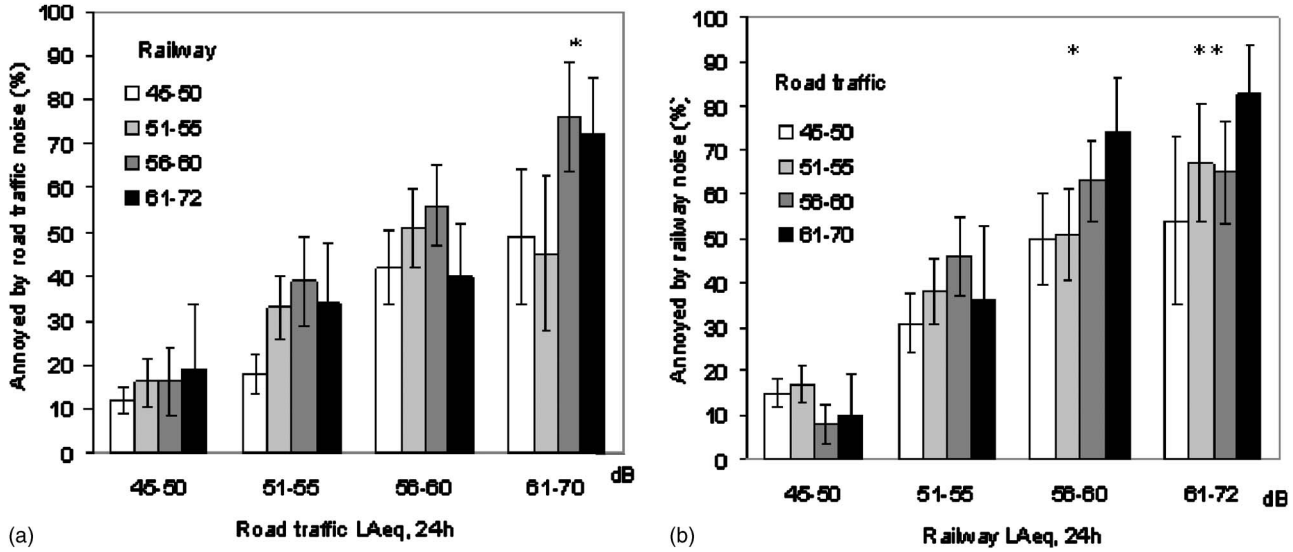


FIG. 5. Annoyance (% moderately, very, extremely annoyed, with 95% confidence intervals) due to road traffic noise in relation to sound levels from road traffic for each of the four different railway noise exposure categories (N=1,953) (left); annoyance (% moderately, very, extremely annoyed) due to railway noise in relation to sound levels from railway traffic for each of the four different road traffic noise exposure categories (N=1953) (right).

traffic noise was similar (r_s=0.40) to the association between sound levels from the railway and annoyance due to railway noise (r_s=0.43). The association between the total traffic sound exposure ($L_{\mathrm{Aeq,24h,tot}}$) and annoyance due to the total traffic sound environment was of about the same magnitude (r_s=0.37) as that for the single noise sources.

We also analyzed the association between annoyance due to the single sources (road traffic and railway) to the total traffic sound environment (total annoyance). Three possibilities were compared, namely the *sum* (the total annoyance is the sum of the annoyance due to each single source), the *average* (the total annoyance is the average of the annoyance due to each single source), and the *maximum* (the total annoyance is the maximum of the annoyance due to each single source). Linear regression and intraclass correlation were used to investigate the association between each measure and the assessed total annoyance. The results showed that the *average* had the strongest association with the total annoyance: the linear regression showed an intercept of 0.25 and a slope of 0.91, and the intraclass correlation was 0.88 (where 1.00 is maximum). Thus, when assessing annoyance due to the total traffic sound environment, respondents are likely to have used an approximate average of the annoyance due to the single sources, rather than the maximum or the sum.

For a more detailed analysis of the annoyance due to single sources and the total annoyance, the study material was divided into eight exposure categories, each 3 dB wide, ranging from $L_{\mathrm{Aeq,24h,tot}}$ 48–50 dB up to $L_{\mathrm{Aeq,24h,tot}}$ 69–72 dB. Figure 6 illustrates the observed prevalence of annoyance due to road traffic noise, railway noise, and the total traffic sound environment in relation to the total sound exposure expressed as $L_{\mathrm{Aeq,24h,tot}}$.

The proportions annoyed by single noise sources (road traffic or railway) and by the total traffic sound environment were very similar in the five lowest exposure categories, although for total sound levels between 54 and 56 dB, the prevalence of annoyance due to railway noise was significantly higher than annoyance due to road traffic noise (26% versus 20%, p=0.006) and to the total traffic sound environment (26% versus 22%, p=0.049), McNemar's test.

There were larger differences between annoyance due to the single noise sources and annoyance due to the total traffic sound environment in the three highest sound exposure categories. When the total sound level was between 63 and 68 dB, the prevalence of annoyance due to road traffic noise was lower than both the prevalence of annoyance due to railway noise (McNemar: p=0.041 at 63–65 dB, p=0.007 at 66–68 dB) and the prevalence of annoyance due to the total traffic environment (McNemar: p=0.029 at 63–65 dB, p=0.004 at 66–68 dB). However, there was no significant difference between total noise annoyance and railway noise annoyance at these sound levels. Additionally, the low number of observations in the highest exposure category (69–72 dB) meant that no significant differences could be shown in this category (see Tables II and III).

D. Maximum sound levels and noise annoyance

Further analyses were performed in order to determine why there were relatively large differences in annoyance due

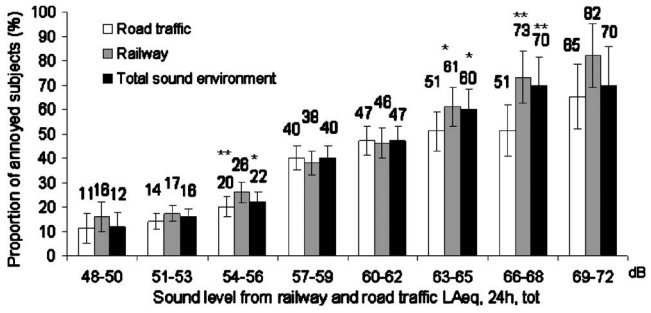


FIG. 6. Prevalence of annoyance (% moderately, very, extremely annoyed, with 95% confidence intervals) from road traffic noise, railway noise, and the total traffic sound environment in relation to the total sound exposure from railway and road traffic, $L_{\mathrm{Aeq,24h,tot}}$. N=1953.

TABLE III. Median sound levels in L_{Amax} for railway and road traffic in the different total sound exposure categories for $L_{Aeq,24h,tot}$.

	Total sound levels from railway and road traffic, $L_{Aeq,24h,tot}$ (dB)							
	48–50	51–53	54–56	57–59	60–62	63–65	66–68	69–72
Number of respondents	142	526	403	377	260	146	66	33
Railway L_{Amax}	65	66	69	73	75	81	86	86
Road traffic L_{Amax}	74	79	78	78	77	78	79	80
Median difference in L_{Amax} railway - road traffic	−9	−13	−9	−5	−2	3	7	6

to road traffic, railway traffic, and the total traffic sound environment in the three highest combined sound exposure categories but not in the lower exposure categories. The sound level in $L_{Aeq,24h}$ for road traffic and for railway noise was similar in the different total sound exposure categories ($L_{Aeq,24h,tot}$). There were, however, large differences in L_{Amax} levels (see Table III) between road traffic and railway noise.

The L_{Amax} level in the different sound exposure categories varied between 65 and 86 dB for railway traffic and between 74 and 80 dB for road traffic. In the five lowest sound exposure categories, the median L_{Amax} levels for railway noise were between 2 and 13 dB *lower* than those for road traffic noise. For the sound exposure categories of 63–65 dB and higher, there was a shift in the relative L_{Amax} levels leading to higher levels for railway noise than for road traffic noise. In these sound exposure categories, where railway noise caused more annoyance than road traffic noise, the median values for L_{Amax} for railway noise were 3–7 dB *higher* than the L_{Amax} levels for road traffic noise.

E. Annoyance due to the total traffic sound environment in relation to total sound exposure for situations with and without dominant noise sources

In order to compare those who were exposed to two noise sources of the same magnitude and those who were exposed to one dominant source of noise, but to the same total traffic sound exposure level, a multiple logistic regression analysis was performed, with the *total traffic sound exposure* ($L_{Aeq,24h,tot}$) as one explanatory variable (continuous variable, no categorization) and *total annoyance* as the response variable ("annoyed"/"not annoyed"). To capture the situation in which railway and road traffic produced equal noise, the difference between railway noise and road traffic noise was calculated, and dwellings were classified into three categories: "road traffic dominant" (road-rail difference >2 dB), "rail and road equally noisy" (rail-road difference between −2 and 2 dB), and "railway dominant" (railroad difference >2 dB). An interaction term between type of dwelling and noise exposure was added to the model (operationalized by multiplying the "type of dominance" variable by the "total traffic sound exposure" variable). The logistic regression analysis was based on the whole sample (rail and road equally noisy $n=683$, road traffic dominant $n=752$, and railway dominant $n=518$). This analysis showed no significant differences in total annoyance between "road traffic dominant" and "railway dominant" dwellings. Therefore the "road traffic dominant" and "railway dominant" dwellings were combined into one category; "one dominant source" dwellings ($n=1270$). The result from the logistic regression is shown in Table IV and also graphically in Fig. 7.

Figure 7 shows the results from the logistic regression models regarding the relationship between *total annoyance* and total traffic sound exposure for the two types of dwellings; "rail and road equally noisy" ($n=683$) and "one dominant source of noise" ($n=1270$), respectively.

The curves in Fig. 7 are estimated using the result in Table IV. The equation for the upper curve (exposed to equal noise from railway and road traffic noise) is: Probability of being annoyed$=\exp(-13.44+0.22x)/(1+\exp(-13.44+0.22x))$, where $x=L_{Aeq,24h,tot}$. The equation for the lower curve (exposed to one dominant noise source, either railway or road traffic noise) is: Probability of being annoyed$=\exp(-9.26+0.15x)/(1+\exp(-9.26+0.15x))$, where $x=L_{Aeq,24h,tot}$ in both cases).

Figure 7 shows that for a low total traffic sound exposure there was no difference in total annoyance between dwellings exposed to one dominant source of noise (lower

TABLE IV. Result from logistic regression model with response variable "annoyed by total traffic sound environment" and explanatory variables "total traffic sound exposure," type of dwelling ("equal noise," or "one dominant source") and the interaction between sound exposure and type of dwelling.

	Response variable: annoyed by total traffic sound environment			
	Estimate (B)	S.E	p	Exp(B)
Total sound exposure ($L_{Aeq,24h,tot}$)	0.149	0.014	<0.0005	1.160
Type of dwelling[a]	−4.180	1.474	0.005	0.015
Sound exposure * type of dwelling	0.076	0.026	0.003	1.079
Constant	−9.261	0.798	<0.0005	<0.0005

Type of dwelling=1 if equal noise from railway and road traffic, 0 otherwise.

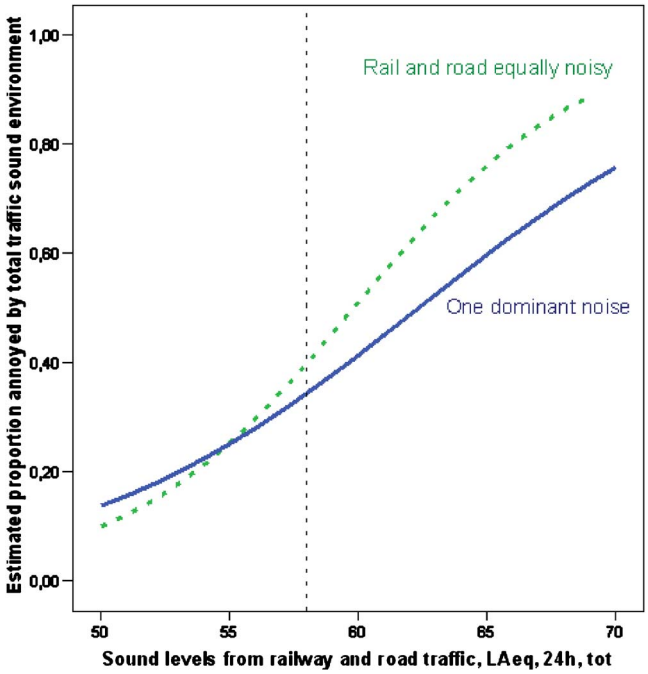


FIG. 7. (Color online) Estimated relationship between total annoyance (% moderately, very, and extremely annoyed) due to the total traffic sound environment and the total traffic sound exposure level ($L_{\mathrm{Aeq,24h,tot}}$) for persons equally exposed to railway and road traffic noise (upper curve) and for those exposed to one dominant noise source, whether railway or road traffic (lower curve).

curve) and those exposed to equal noise from railway and road traffic (upper curve). But as the total traffic sound exposure increased, the prevalence of total annoyance gradually became higher for dwellings where railway and road traffic contributed equally. Thus, there was an interaction between total traffic sound exposure and type of dwelling and this interaction was statistically significant ($p=0.003$). The difference in total annoyance between the two types of dwellings was statistically significant when total traffic sound exposure ($L_{\mathrm{Aeq,24h,tot}}$) was above 58 dB. We estimated that the expected difference in the proportion of total annoyance between a situation with one dominant noise source and a situation with two equally noisy sources was 0.2% at 55 dB, 10% at 60 dB (40% versus 50% annoyed), 17% at 65 dB (60% versus 77% annoyed), and 15% at 70 dB (75% versus 90% annoyed).

The same conclusion (a significant interaction with a higher proportion of annoyed with two equally noisy sources) was reached when a logistic regression was performed using another definition of annoyance due to the total traffic sound environment, namely "highly annoyed," defined as "very or extremely annoyed."

In the logistic regression model above, dwellings were categorized as being exposed to two equally noisy sources if the difference between the sources was 2 dB or smaller, and as being dominated by one source if the difference was 3 dB or larger. We also performed another analysis, based on a subsample ($n=1662$), where we defined the sources as being equally noisy if the difference was 1 dB or smaller, and as having one source dominant if the difference was 3 dB or larger. The conclusions from this analysis were the same as above: the total sound exposure from two equally noisy sources caused a higher prevalence of total annoyance at high total sound levels than the same total sound exposure in situations with one dominant source (either railway or road traffic). The difference in total annoyance was again significant for sound levels above $L_{\mathrm{Aeq,24h,tot}}$ 58 dB.

IV. DISCUSSION

A. Annoyance from single sources—railway and road traffic

Annoyance due to road traffic noise varied from 14% at sound levels of 45–50 dB up to 62% at sound levels between $L_{\mathrm{Aeq,24h}}$ 61 and 70 dB. For the lower sound levels, this is in accordance with results obtained in meta-analyses based on large data sets (Miedema and Oudshoorn, 2001), and recent results from Swedish studies which show that 3% of participants were annoyed by road traffic noise at sound levels around $L_{\mathrm{Aeq,24h}}$ 42–43 dB (Öhrström *et al.*, 2006). However, at higher sound levels, the prevalence of annoyance due to road traffic noise was considerably higher than would be expected from the dose-response relationships obtained by Miedema and Oudshoorn (2001) and by Öhrström *et al.* (2006); for example, 48% at 56–60 dB and 62% at 61–70 dB. To some extent, the high prevalence of annoyance due to road traffic noise in the present study might be explained by the simultaneous exposure to railway noise.

An overwhelming majority of studies have shown that railway noise causes less annoyance than road traffic and aircraft noise (e.g., Miedema and Vos, 1998; Miedema and Oudshoorn, 2001; and review by Öhrström and Skånberg, 2006) while only a few studies show no support for a railway bonus (e.g., Kaku and Yamada, 1996; Yano *et al.*, 1998). However, participants in the present study were more likely to be annoyed by railway noise than by road traffic noise at sound levels above 50 dB $L_{\mathrm{Aeq,24h}}$, with the percentage annoyed by railway noise being 7–10% higher than the corresponding percentage annoyed by road traffic noise. We have considered a number of possible explanations for this unexpected finding. It has been shown that railway noise is perceived as being more annoying in areas where there is simultaneous exposure to vibration from railway traffic (Öhrström and Skånberg, 1996; Ota *et al.*, 2006). In parts of the present study area, the ground consists of clay, and so the passage of trains creates vibrations which propagate through the buildings. This may have made some contribution to the relatively high annoyance due to railway noise, but spontaneous reports of annoyance due to vibrations from the trains were rare, and other explanations seem more plausible.

Changes in railway traffic have taken place over the last 5 years with the number of passenger trains increasing by 6%, and also more and heavily loaded freight trains in nighttime (4/night). In addition to this, there are plans to build additional railway tracks passing through the study area. When our survey was conducted, there were no concrete proposals concerning the localization of these new tracks, but nevertheless people may have been worried about the extension and feared increased disturbances from railway noise. Thus, while Lambert *et al.* (1998) found evidence for

a "railway bonus" of up to 5 dB for existing railway lines at higher sound levels (around 70 dB), when comparing data from French socio-acoustic surveys, annoyance regarding new railways (or roads) was considerably higher than annoyance regarding existing infrastructures. The plans for additional railway lines may have outweighed the expected 5 dB railway bonus found in studies of stationary exposure situations. In a review of studies on community response to changes in railway noise, Huybregts (2003) concluded that there is an exaggerated annoyance response to noise when railway noise exposure changes. In one of the longitudinal studies reviewed by Huybregts (van Dongen and van den Berg, 1983), annoyance due to noise from a new railway line in Zoetermeer, The Netherlands, was evaluated 4 and 18 months after its opening. The prevalence of high annoyance was much greater 4 months after the opening than 18 months after, especially at levels below level day night (DNL) 55 dB. At these levels, annoyance after 18 months was as expected from the steady-state dose-response curve for railway noise (Miedema and Vos 1998). At sound levels above DNL 55 dB, annoyance also decreased after 18 months, but only by 3–4% as compared with annoyance 4 months after the opening of the new railway. Increased annoyance when the new infrastructure was planned was also demonstrated by Öhrström and Skånberg (1996). In this study, a much higher prevalence of annoyance due to railway noise was found in one of the study areas compared to other study areas with similar noise levels. In this area, new railway tracks were planned for local passenger trains. In a later study performed in the same study area a few years later (Öhrström, 1997) after the railway tracks had been rebuilt and an additional 90 local passenger trains introduced, the prevalence of annoyance due to railway noise decreased from 37% to 23% as an average for the entire study area, similar to that in other study areas with stationary exposure situations. In summary, we believe that concern about possible changes in railway noise may have increased the annoyance reported in the present study.

B. Annoyance due to single noise sources versus total annoyance due to combined sources

Risk assessments of the health effects of noise are often based on known dose-response relationships for single noise sources. The most recent comprehensive meta-analysis presents synthesis curves for the relationships between exposures from separate transportation noise sources (road, railway, and aircraft) and the percentage of individuals who were highly annoyed by these exposures (Miedema and Vos 1998; Miedema and Oudshoorn 2001). As research on the effects of combined noise sources is rare, corresponding relationships for noise from two or more simultaneously occurring sources are not yet available. Different theories exist on how to best assess the total sound exposure (Miedema, 2004; Nilsson, 2001); two of the models that have been suggested are the energy summation model, which was applied for assessment of sound exposure in this study, and the dominance model.

There is no standard method for assessment of annoyance due to the total sound environment, whereas validated questions are available (ISO, 2003) for noise annoyance due to single sources. In addition, the validity of a total annoyance judgment has been questioned by Miedema (2004). In a recent study (Ota *et al.* 2006), total noise annoyance was evaluated using a question on "Quietness of surroundings;" the response alternatives were "satisfied," "partially satisfied," "neutral," "partially dissatisfied," and "dissatisfied," and the percentage of "dissatisfied" responses was used as an index of total noise annoyance. Since the results by Ota *et al.* (2006) may be based on a range of noise sources, and not only traffic noise as in the present study, comparisons are not possible. In the present study, annoyance from combined noise sources was evaluated with a specific question phrased to capture annoyance due to all types of noise from transport ("annoyance due to the total traffic sound environment"). The total annoyance was never higher than annoyance from the most annoying single noise source within the same exposure category (Fig. 6). Analyses indicate that the respondents assessed total annoyance as the average of their annoyance due to the single sources. This is in line with results from assessment of loudness in experimental studies (Nilsson, 2001). The results showed that the dose-response relationship between annoyance due to the total traffic sound environment (total annoyance) and the total sound level from railway and road traffic ($L_{\mathrm{Aeq,24h,tot}}$) was about the same (r_s =0.37) as the relationship between road traffic noise annoyance and sound levels for road traffic (r_s=0.40) and the relationship between railway noise annoyance and sound levels from the railway (r_s=0.43).

The results of this study imply that combined exposures to noise from road traffic and railway induce more extensive annoyance reactions than noise from single sources at the same sound level. This must be considered in risk assessments. When railway and road traffic were equally noisy (L_{Aeq} level), total annoyance was higher than when one of the two sources was dominant (rail or road) but the total sound level was the same. The difference in total annoyance between these two situations increased with higher sound levels (Fig. 7). A possible explanation is that the combination of road traffic and railway noise is a combination of a relatively high constant background noise (road traffic) and a more intermittent noise (railway). The difference in total annoyance between combined and single noise source situations was significant for total sound levels above 58 dB but not for lower sound levels. As shown in Table III, from about 63 dB ($L_{\mathrm{Aeq,24h,tot}}$) the L_{Amax} level from the railway was higher than the L_{Amax} from road traffic. Such a combination of constant and intermittent noise with very high maximum sound levels may be more annoying than constant or intermittent noise alone, even if the total sound level is the same in the three cases. Railway noise, especially that from the heavy freight trains with high sound levels, was more dominant than road traffic noise at night, and it is well known that intermittent noise disturbs sleep more than continuous steady-state noise (Öhrström, 2000; Öhrström and Skånberg, 2004). The research questions should also be addressed for the combination of high road traffic noise and (intermittent) aircraft noise.

C. Methodological considerations

As previous socio-acoustic surveys (e.g., Öhrström *et al.*, 2006) have demonstrated that adverse effects of noise are rare at low sound levels, this study was restricted to residential areas with dwellings exposed to sound levels from road traffic and railway exceeding $L_{Aeq,24h}$ 45 dB. The assessment of sound levels at the dwellings was thorough, and the accuracy of calculated sound levels in $L_{Aeq,24h}$ was estimated at ±3 dB. Calculated L_{Amax} levels for road traffic noise in areas further away from Highway E20 are more uncertain, since traffic counts were not available for small local roads, and so had to be estimated.

There are several aspects of this study which strengthen the internal validity, that is, how well the results reflect the true effects of noise in the population of the study area. First, subjects in the various sound exposure categories were selected randomly from the target population. Second, the response rate (71%) was higher than is usual in this type of study, minimizing the possible effect of a difference in the prevalence of annoyance reactions between respondents and nonrespondents. Third, the analyses were based on almost 2000 questionnaires, making the random error relatively small for annoyance and other common effects in the total sample. For example, the random error for the 29% of 1953 individuals who answered that they were (rather, very, or extremely) annoyed by road traffic noise was only ±2% (SE). For the analyses of subgroups, this uncertainty is larger. Among the 434 individuals in the 56–60 dB sound exposure category, 48% were annoyed by road traffic noise with a random error of ±5%. For the subgroup with only 167 individuals in the highest sound exposure category (62% annoyed), the random error was ±8% (Fig. 4).

V. CONCLUSIONS

Combined exposures to noise from two sources, road traffic and railway, induced more extensive annoyance reactions than noise from single sources at the same sound level. In situations where the total sound level was above 58 dB, the total annoyance was significantly higher for those exposed to two equally noisy sources (rail and road), compared to one dominant source. Therefore, effects of the total sound exposure should be considered in risk assessments and in noise mitigation activities.

The extent of railway noise annoyance was similar to that of road traffic noise annoyance at sound levels below 50 dB or above 60 dB. However, at sound levels between 51 and 60 dB, the proportion annoyed by railway noise was higher than that annoyed by noise from road traffic. These findings are in conflict with the majority of other European studies, which show that railway noise is less annoying than road traffic noise. Situational factors may have contributed to the extensive annoyance effects from railway noise, for example, changes and plans for new railway infrastructure. New comparative studies should be conducted to further evaluate annoyance due to road and railway noise and the relevance of a railway bonus in residential areas exposed to railway lines with a very large number of trains.

ACKNOWLEDGMENTS

This study was funded by the Swedish National Board of Health and Welfare, the Swedish Environmental Protection Agency, the County Administrative Board of Västra Götaland, RUS (Regional Follow-up System), the Municipality of Lerum, and the Sahlgrenska University Hospital.

Champelovier, P., Cremezi, C., Charlet, C., and Lambert, J. (**2004**). "Evaluation de la gêne due à l'exposition combinée aux bruits routier et ferroviaire!," ("Evaluation of annoyance due to combined exposure to noise from road and railway traffic!"), Lyon, France, INRETS, Report No. INRETS. 242.

Cremezi, C., Gautier, P. E., Lambert, J., and Champelovier, P. (**2001**). "Annoyance due to combined noise sources: advanced results," *Proceedings ICA International Commission for Acoustics: 17th ICA - International Congress on Acoustics*, Rome, Italy, edited by A. Alippi, September 2–7 (CD ROM).

EU Commission (**2002**). "A study of European priorities and strategies for railway noise abatement," *Final Report No.* 01.980, EU Commission, Directorate-General for Energy and Transport, Belgium, ODS Reference No.: 00.2335.

Huybregts, C. (**2003**). "Community response to changes in railway noise exposure—a review," *Proceedings of the 8th Western Pacific Acoustics Conference (Wespac VIII)*, Melbourne, Australia, 7–9 April.

ISO (E). (**2003**). "Technical specification, acoustics – Assessment of noise annoyance by means of social and socio-acoustic surveys," ISO/TS15666: (E).

Jonasson, H., and Nielsen, H. L. (**1996**). *Road traffic noise-Nordic prediction method*, Nordic Council of Ministers, Tema Nord, Copenhagen, Denmark, Vol. **525**, ISBN 92 9120 836 1.

Joncour, S., Cailhau, D., Gautier, P. E., Champelovier, P., and Lambert, J. (**2000**). "Annoyance due to combined noise sources," *Proceedings of the 29th International Congress and Exhibition on Noise Control Engineering*, Nice, France, 27–30 August.

Kaku, J., and Yamada, I. (**1996**). "The possibility of a bonus for evaluating railway noise in Japan," J. Sound Vib. **193**(1), 445–450.

Kurra, S., Morimoto, M., and Maekawa, Z. I. (**1999**). "Transportation annoyance—A simulated-environment study for road, railway and aircraft noises, Part I: Overall annoyance," J. Sound Vib. **220**(2), 251–278.

Lambert, J., Champelovier, P., and Vernet, I. (**1998**). "Assessing the railway bonus: The need to examine the "new infrastructure" effect," *Proceedings of Inter Noise 1998*, 16–18 November, Christchurch, New Zealand.

Lercher, P., Brauchle, G., and Widmann, U. (**1999**). "The interaction of landscape and soundscape in the Alpine area of the Tyrol: an annoyance perspective," *Proceedings of Inter Noise 1999*, Fort Lauderdale, FL, 2–4 December.

Miedema, H. (**2004**). "Relationship between exposure to multiple noise sources and annoyance," J. Acoust. Soc. Am. **116**(2), 949–957.

Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.

Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.

Moehler, U., Liepert, M., Schuemer, R., and Griefahn, B. (**2000**). "Differences between railway and road traffic noise," J. Sound Vib. **231**(3), 853–864.

Morihara, T., Yano, T., and Sato, T. (**2002**). "Comparison of dose-response relationships between railway and road traffic noises in Kyushu and Hokkaido, Japan," *Proceedings of Inter Noise 2002*, Dearborn, MI, 19–20 August.

Nilsson, M. E. (**2001**). "Perception of traffic sounds in combination," Ph.D. dissertation, Archives of the Center for Sensory Research, Stockholm University and Karolinska Institute, Stockholm, Sweden, ISBN 91-887-8410-X.

Öhrström, E. (**1989**). "Sleep disturbance, psychosocial and medical symptoms—A pilot survey among persons exposed to high levels of road traffic noise," J. Sound Vib. **133**, 117–128.

Öhrström, E., and Skånberg, A.-B. (**1996**). "A field survey on effects of exposure to noise and vibrations from railway traffic. Part I: Annoyance and activity disturbance effects," J. Sound Vib. **193**(1), 39–47.

Öhrström, E. (**1997**). "Community reactions to railway traffic—Effects of countermeasures against noise and vibration," *Proceedings of Inter Noise*

1997, Budapest, Hungary, Vol. **2**, pp. 1065–1070.
Öhrström, E. (**2000**). "Sleep disturbances caused by road traffic noise—Studies in laboratories and field," Noise & Health **8**, 71–78.
Öhrström, E., and Skånberg, A. (**2004**). "Sleep disturbances from road traffic and ventilation noise—Laboratory and field experiments," J. Sound Vib. **271**, 279–296.
Öhrström, E. (**2004**). "Longitudinal surveys on effects of changes in road traffic noise—Annoyance, activity disturbances and psycho-social well being," J. Acoust. Soc. Am. **115**(2), 719–729.
Öhrström, E., Barregård, L., Skånberg, A., Svensson, H., Ångerheim, P., Holmes, M., and Bonde, E. (**2005a**). "Undersökning av hälsoeffekter av buller från vägtrafik, tåg och flyg i Lerums kommun," ("Study of health effects from road traffic, railway and aircraft noise in Lerum municipality"), Report No. ISSN 1400-5808, ISRN GU-MMED-R-2005/1-SE, Department of Occupational and Environmental Medicine, The Sahlgrenska Academy at Göteborg University, Göteborg, Sweden.
Öhrström, E., Skånberg, A., Barregard, L., Svensson, H., and Angerheim, P. (**2005b**). "Effects of simultaneous exposure to noise from road- and railway traffic," *Proceedings of Inter Noise 2005*, Rio de Janeiro, 6–10 August, Paper no 1570 (CD ROM).
Öhrström, E., and Skånberg, A. (**2006**). "Litteraturstudie–Effekter avseende buller och vibrationer från tåg- och vägtrafik." ("Literature study - Effects of noise and vibrations from railway and road traffic"), Report No. 112, ISSN 1650-4321, ISBN 91-7876-111-5, Ocupational and Environmental Medicine, The Sahlgrenska Academy at Göteborg University, Göteborg, Sweden.
Öhrström, E., Skånberg, A., Svensson, H., and Gidlöf Gunnarsson, A. (**2006**). "Effects of road traffic noise and the benefit of access to quietness," J. Sound Vib. **295**, 40–59.
Ota, A., Yokoshima, S., Kamitani, K., and Tamura, T. (**2006**). "A study on evaluation methods of combined traffic noises, Part 2: Community response to road traffic and conventional railway noises," *Proceedings of Inter-Noise 2006*, Honolulu, Hawaii, 3–6 December, Paper No. 412 (CD ROM).
Ronnenbaum, T., Schulte-Fortkamp, B., and Weber, R. (**1996**). "Synergetic effects of noise from different sources: A literature study," *Proceedings of Inter-Noise 96*, Liverpool, U.K., edited by F. A. Hill and R. Lawrence July, Vol. **5**, pp. 2241–2246.
National Institute for Communication Analysis (SIKA) (**2000**). "Utveckling av godstransporter 2010," ("Prognoses of freight transportation 2010"), Report No. 2000:7.
National Institute for Communication Analysis (SIKA) (**2002**). "Utveckling av persontransporter, prognoser for 2010, 2020," ("Development of passenger transport, forecast for 2010, 2020"), Report No. 2002:1.
van Dongen, J. E. F., and van den Berg, R. (**1983**). "Gewenning aan het geluid van een nieuwe spoorlijn," (Adapting to noise from a new railway line), IMG-TNO Report No. RL-HR-03-02, Delft, The Netherlands.
Vos, J. (**2004**). "Annoyance caused by the sounds of a magnetic levitation train," J. Acoust. Soc. Am. **115**(4), 1597–1608.
World Health Organization (WHO). (**2000**). "Guidelines for community noise" edited by B. Berglund, T. Lindvall, D. H. Schwela, and K.-T. Goh, Guideline Document, Geneva, Switzerland.
Yano, T., Murakami, Y., Kawai, K., and Sato, T. (**1998**). "Comparison of responses to road traffic and railway noises," *Proceedings Noise Effects '98—7th International Congress on Noise as a Public Health Problem*, Sydney, Australia, 22–26 November, Noise Effects '98 PTY Ltd.

Elastic characterization of closed cell foams from impedance tube absorption tests

F. Chevillotte and R. Panneton[a)]
GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke Quebec, Canada, J1K2R1

(Received 18 June 2007; revised 1 August 2007; accepted 16 August 2007)

A method is presented to determine the bulk elastic properties of isotropic elastic closed-cell foams from impedance tube sound absorption tests. For such foams, a resonant sound absorption is generally observed, where acoustic energy is transformed into mechanical vibration, which in turn is dissipated into heat due to structural damping. This article shows how the bulk Young's modulus, Poisson's ratio, and damping loss factor can be deduced from the resonant absorption. Also, an optimal damping loss factor yielding 100% of absorption at the first resonance is defined from the developed theory. The method is introduced for a sliding edge condition which is an ideal condition. Then, the method is extended to a bonded edge condition which is more easily achievable and additionally enables the identification of the Poisson's ratio. The method is experimentally tested on expanding closed-cell foams to find their elastic properties in both cases. Using the found properties, sound absorption predictions using an equivalent solid model with and without surface absorption are compared to measurements. Good correlations are obtained when considering surface absorption. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783126]

PACS number(s): 43.55.Ev, 43.20.Ye, 43.40.At [LLT] Pages: 2653–2660

I. INTRODUCTION

Many authors worked on the propagation of acoustical and elastic waves in elastic open-cell porous materials. The most used model is the one proposed by Biot[1] in which the porous medium is modelled in two superimposed phases (fluid and solid). The fluid phase (usually air) forms an interconnected cell network which opens to the surrounding medium. The two phases are elastically coupled and a relative motion between the two phases exists. Three waves propagate in the porous material (two compressional waves and one shear wave). The energy carried by the waves is dissipated through structural damping loss, viscous loss (due to relative motion between the two phases), and thermal loss.[2] In the case of elastic closed-cell foams, there is no relative motion between the fluid and solid phases. Consequently, the only energy dissipation mechanisms are the structural damping and thermal losses.[3] From the Biot's theory, only the elastic compression and shear waves now propagate in the closed-cell foams. In this case, few specific models were proposed to study the acoustic dissipation within this type of foams. The most common way of modelling these foams is to use a solid model with equivalent elastic properties. Ingard[3] proposed a method to derive these equivalent properties from the knowledge of the elastic properties of the material from which the skeleton is made, and the properties of the trapped air in the closed cells. The equivalent properties independently account for the thermal loss (which is frequency dependent) and the structural damping loss of the skeleton material. However, in many situations the properties of the skeleton material are not known after the foaming process. Also, even if the core of these foams is made of closed cells (i.e., no propagation of acoustical waves within the core), their surface may show some irregularities and open cells (e.g., exploded cells). In this case, the surface may be seen as a thin resistive layer showing some surface sound absorption.

To account for the surface sound absorption of closed-cell foams, a model was worked out by Wojtowicki and Panneton.[4] In this work, it was shown that closed cell foams show resonant sound absorption (i.e., sound absorption at elastic resonances) with residual surface absorption (apart from resonances). To model this type of acoustic behavior, the closed-cell foams are modeled as a two-layer material: A resistive layer covering a core made of an equivalent solid with bulk elastic properties. The bulk elastic properties account for the structural damping and thermal losses. The resistive surface layer is characterized by its static airflow resistivity. Although the resistivity can be easily identified, the key element in the proper use of this so-called "surface absorption solid model" is the fine characterization of the bulk elastic properties.

The main objective of this work is to develop a method, based on sound absorption measurements, for the characterization of the bulk elastic properties of closed-cell foams—notably those from heat expanding foam process.[4] In this particular case, the produced samples have generally irregular surface and thickness that prevent the use of existing methods for the elastic characterization of column foam samples.[5–10] Brief descriptions of these methods are given elsewhere.[5]

This article is organized as follows. In Sec. II, the theory behind the method is first introduced for the sliding edge condition for which the Young's modulus and damping loss factor can be deduced. Then, the method is extended to the

[a)]Author to whom correspondence should be addressed. Electronic mail: raymond.panneton@usherbrooke.ca

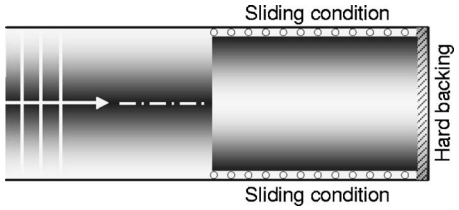


FIG. 1. Configuration of the equivalent solid foam column tested in a rigid wall impedance tube.

bonded edge condition which additionally allows for the identification of the Poisson's ratio. In Sec. III, the method is experimentally tested for characterizing the bulk elastic properties of two foams: one with the sliding edge condition and one with the bonded edge condition. Then, sound absorption predictions using the identified bulk elastic properties are compared to impedance tube results. Finally, Sec. IV concludes this work.

II. THEORY

A. Sliding edge condition

1. First natural frequency

In the following, the elastic closed-cell foam under consideration is considered as being isotropic and homogeneous so it can be modeled *a priori* as an equivalent solid. Also, it is assumed that its elastic properties are constant, or only slightly vary, with the frequency—at least around the first resonant absorption frequency. A column of this foam is now excited by a normal incidence acoustic plane wave as shown in Fig. 1. The column is backed by a rigid wall and a sliding condition is applied on its periphery. Although it is difficult to obtain in practice, the sliding condition may be approximated by lubricating the sample periphery and slightly undercutting sample diameter. Under this configuration, the normal sound absorption is of a resonant type[3,4] with the first peak absorption occurring at the first compression natural frequency of the column[11]

$$f_1 = \frac{\omega_1}{2\pi} = \frac{1}{4L}\sqrt{\frac{K_b}{\rho_1}}, \tag{1}$$

where L is the height of the column, ρ_1 the bulk density of the foam, and K_b its bulk compression modulus.

2. Finding Young's modulus

Assuming that the height and the bulk density are known, the first resonant frequency is determined from the normal sound absorption coefficient curve and the bulk modulus of the foam is deduced from Eq. (1):

$$K_b = \rho_1(4f_1L)^2. \tag{2}$$

As the equivalent solid is isotropic, one can relate the bulk modulus to the Young's modulus E and Poisson's ratio ν by

$$K_b = E\frac{(1-\nu)}{(1+\nu)(1-2\nu)}. \tag{3}$$

Equation (3) shows that an infinite number of couples (E,ν) yields the same bulk modulus K_b. Hence, the Poisson's ratio is required in order to find the Young's modulus from the bulk modulus. Assuming Poisson's ratio known, the Young's modulus is given by

$$E = \frac{(1+\nu)(1-2\nu)}{(1-\nu)}K_b. \tag{4}$$

As the Poisson's ratio has little influence for a sliding edge condition, and based on the works by Gibson *et al.*,[12] a mean Poisson's ratio of 0.33 can be used for typical closed-cell foams. Also, as Poisson's ratio is only slightly frequency dependent, one could use a static method to measure the Poisson's ratio.

3. Finding bulk damping loss factor

The normal sound absorption coefficient of the column shown in Fig. 1 is expressed in terms of its surface acoustic impedance Z_s and the characteristic impedance of air Z_0

$$\alpha(\omega) = 1 - \left|\frac{Z_s(\omega) - Z_0}{Z_s(\omega) + Z_0}\right|^2, \tag{5}$$

where ω is the angular frequency. As the cells of the foam are closed, there is no relative motion between the air in the cells and the structure, therefore no friction losses as in open-cell foams.[3] The only losses are due to structural damping and heat conduction (relative to the harmonic compression of the air in the cells). Making use of a bulk damping loss factor η combining both dissipation mechanisms, the losses can now be introduced in the bulk modulus as follows:

$$\tilde{K}_b = K_b\sqrt{1+j\eta}. \tag{6}$$

Considering only elastic compressional waves in the material, the normal surface impedance of the sliding edge sample on hard backing may be written as

$$Z_s = \sqrt{\rho_1\tilde{K}_b}\coth\left(j\omega L\sqrt{\frac{\rho_1}{\tilde{K}_b}}\right), \tag{7}$$

where $(\rho_1\tilde{K}_b)^{1/2}$ is the characteristic impedance of the material, and $(\rho_1/\tilde{K}_b)^{1/2}$ is the inverse of the propagation speed of the compressional waves in the material. Using Eqs. (1) and (6), the surface impedance (7) now rewrites

$$Z_s = \frac{2}{\pi}\rho_1\omega_1L\sqrt{1+j\eta}\coth\left(j\frac{\pi}{2}\frac{\omega}{\omega_1}\frac{1}{\sqrt{1+j\eta}}\right). \tag{8}$$

At the first resonance ($\omega=\omega_1$), the surface impedance becomes

$$Z_s(\omega_1) = \rho_1\omega_1L\left(\coth\left(j\frac{\pi}{2}\frac{1}{\sqrt{1+j\eta}}\right)\Big/\frac{\pi}{2}\frac{1}{\sqrt{1+j\eta}}\right). \tag{9}$$

Assuming the damping loss factor is low enough (less than 1), Eq. (9) simplifies to

$$Z_s(\omega_1) \cong \tfrac{1}{2}\rho_1\omega_1L\eta. \tag{10}$$

Consequently, at the first elastic resonance, substituting Eq. (10) into Eq. (5) yields the approximated sound absorption coefficient

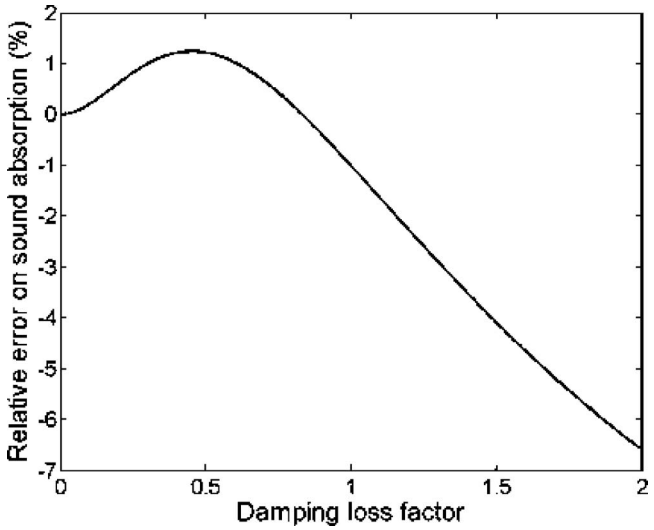


FIG. 2. Relative error on sound absorption due to the approximated surface impedance [Eq. (10)].

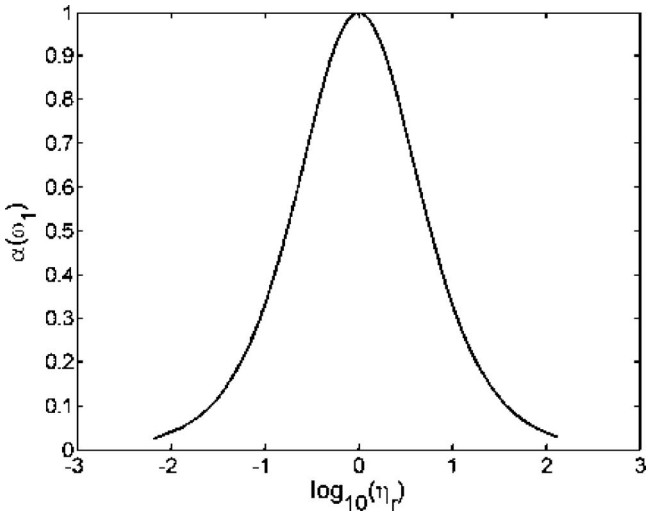


FIG. 3. Sound absorption coefficient at the first resonance in function of the reduced structural damping loss factor for closed-cell foams.

$$\alpha(\omega_1) = 1 - \left| \frac{\rho_1 \omega_1 L \eta - 2Z_0}{\rho_1 \omega_1 L \eta + 2Z_0} \right|^2, \tag{11}$$

or, using Eq. (1),

$$\alpha(\omega_1) = 1 - \left| \frac{\pi \sqrt{\rho_1 K_b}\, \eta - 4Z_0}{\pi \sqrt{\rho_1 K_b}\, \eta + 4Z_0} \right|^2. \tag{12}$$

Equation (11) implies that the bulk damping loss factor of the material can be deduced from the sound absorption coefficient, knowing the bulk density of the material, the first resonance frequency, and the characteristic impedance of air. This is true only if the normal sound absorption coefficient shows the first resonant peak absorption in the measured frequency range.

Equation (12) is presented in order to show that the amplitude of the resonant absorption does not depend on the thickness of the sample. Consequently, changing the thickness will shift the peak absorptions only in frequencies and not in amplitude.

As the absorption coefficient given by Eq. (11) is valid for low damping loss factors, Fig. 2 shows how this approximation diverges from the exact solution in function of the loss factor. One can note that the absolute error is less than 2% for damping loss factors smaller than 1. Consequently, Eq. (11) can be used with confidence for a large number of elastic closed-cell foams as most of them have a loss factor less than 1.

4. Optimal and reduced damping loss factors

Figure 3 shows how the sound absorption coefficient at first resonance varies with the damping loss factor. The sound absorption coefficient reaches almost 100% absorption for an optimal damping loss factor. This optimal loss factor is deduced from Eq. (11) by setting $\alpha(\omega_1)=1$. This yields:

$$\eta_{\text{opt}} = \frac{2Z_0}{\omega_1 \rho_1 L}. \tag{13}$$

Now, one can define the reduced damping loss factor as the ratio of the actual damping loss factor to the optimal value

$$\eta_r = \frac{\eta}{\eta_{\text{opt}}}. \tag{14}$$

Substituting Eq. (14) into Eq. (11), the sound absorption at first resonance reduces to

$$\alpha(\omega_1) = 1 - \left| \frac{\eta_r - 1}{\eta_r + 1} \right|^2. \tag{15}$$

Apart from the optimal damping loss factor, Eq. (15) and Fig. 3 show that there exist two reduced damping loss factors for a given value of $\alpha(\omega_1)$. The first solution corresponds to a low damping solution, and the second one corresponds to a high damping solution. These solutions are, respectively,

$$\eta_r^{\text{sol1}} = \frac{4 - 2\alpha(\omega_1) - 4\sqrt{1-\alpha(\omega_1)}}{2\alpha(\omega_1)},$$
$$\eta_r^{\text{sol2}} = \frac{4 - 2\alpha(\omega_1) + 4\sqrt{1-\alpha(\omega_1)}}{2\alpha(\omega_1)}, \tag{16}$$

with $\eta_r^{\text{sol1}} \leqslant \eta_{\text{opt}} \leqslant \eta_r^{\text{sol2}}$.

Figure 4 shows the normal incidence sound absorption coefficient of an equivalent solid for the optimal damping loss (13), and the damping loss solutions obtained from Eqs. (14) and (16) with $\alpha(\omega_1)=0.7$. The absorption coefficient is computed using Eqs. (5) and (7) with $L=30$ mm, $\rho_1=90$ kg m^{-3}, $K_b=536$ kPa, and $Z_0=412$ Pa s m^{-1}. As expected, the main difference between sol1 and sol2 lies in the width of the peak and not in its amplitude.

Here, the optimal damping introduced through Eq. (13) must be used with caution since it is only valid for a sliding edge condition on small samples, or for samples of laterally infinite extent (or large extent using a free field measuring method[13]). If other boundary conditions are used on small samples, another set of equations needs to be developed.

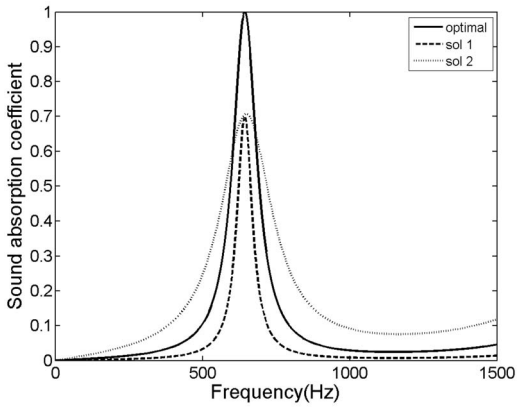


FIG. 4. High and low solutions of the reduced damping loss factor.

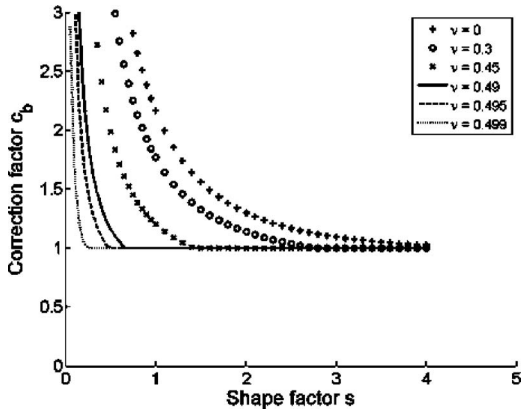


FIG. 5. Computation of the bonded correction factor c_b.

B. Bonded edge condition

The previous section has shown the theory of a method to find the bulk Young's modulus and damping loss factor for a sliding sample of solid or closed-cell porous materials on hard backing. Unfortunately, the sliding condition in an impedance tube cannot be perfectly ensured in practice. Therefore, this section extends the proposed method to another type of boundary condition: the bonded edge condition. In this case, the same assumptions as previously defined can be first made (i.e., the Poisson's ratio is assumed to be known and the sound absorption at the resonance is only governed by the bulk damping loss factor).

1. Finding Young's modulus

Considering a bonded edge condition and a hard backing, the first compression resonance depends on the shape factor s and the Poisson's ratio ν. This frequency can be expressed by

$$f_1^b = \frac{\omega_1^b}{2\pi} = \frac{c_b(\nu,s)}{4L}\sqrt{\frac{K_b}{\rho_1}}, \tag{17}$$

where the shape factor s is the ratio of the diameter of the sample to its thickness: i.e., $s=D/L$.

Comparing to the first resonance frequency of a sliding sample given by Eq. (1), Eq. (17) becomes:

$$f_1^b = c_b(\nu,s) f_1. \tag{18}$$

The factor c_b is called the bonded edge correction factor. This coefficient only depends on the shape factor s and Poisson's ratio ν. A two dimensional axisymmetric finite element method (FEM) model has been used in order to compute this coefficient. A CPU Intel Pentium D, 3 GHz, 4 GB RAM has been used for the entire study. An eigenfrequency analysis is employed for the bonded sample. The correction factor c_b is simply the ratio of the first resonance frequency calculated with the FEM model to the frequency given by Eq. (1). Note that the FEM simulations for a sliding sample correlates with Eq. (1). From the FEM results, an abacus of the correction factor is generated—this eliminates the use of FEM simulations in the future. This table contains the bonded correction factor c_b for $s \in [0.05\ 5]$ and $\nu \in [0\ 0.499]$. Then, a function is implemented for interpolating the bonded correction factor from a given couple (s,ν). Figure 5 shows how this correction factor evolves with the shape factor for some Poisson's coefficients. The low limit of the correction factor is one. A factor below 1 would imply that the first resonance is no longer a compression resonance. The theory would not be valid anymore. Consequently, with a shape factor of 1.5, the method would work for characterizing materials with Poisson's ratio up to 0.45. This is quite acceptable for typical closed-cell foams.

Still assuming that the Poisson's ratio is known, the bulk modulus is obtained from Eqs. (2) and (18), knowing the bulk density ρ_1, the thickness L and the shape factor s of the sample:

$$K_b = \rho_1\left(4\frac{f_1^b}{c_b(\nu,s)}L\right)^2. \tag{19}$$

Finally, the Young's modulus is deduced from Eq. (4).

Contrary to the sliding edge condition, when the bonded edge condition is considered, the bonded sample undergoes shear stresses and the Poisson's ratio plays an important role. Taking advantage of this observation, a method for finding the Poisson's ratio from impedance tube absorption tests is given in a next paragraph.

2. Finding bulk damping loss factor

Considering a bonded sample on a hard backing, the sound absorption coefficient at the resonance still reaches a maximum for an optimal damping loss factor η_{opt}^b. This behavior is similar to the sliding case illustrated in Fig. 3. Thus, the sound absorption coefficient could still be expressed by

$$\alpha(\omega_1) = 1 - \left|\frac{\eta_r - 1}{\eta_r + 1}\right|^2, \quad \eta_r = \frac{\eta}{\eta_{\text{opt}}^b}. \tag{20}$$

Once the reduced damping loss factor has been found from the sound absorption at the resonance, the optimal damping loss factor is required. Unfortunately, Eq. (13) cannot be used anymore. Indeed $\eta_{\text{opt}}^b \neq \eta_{\text{opt}}$ and the greatest difficulty lies in obtaining this optimal factor. One calls this factor the bonded edge optimal loss factor in opposition to the sliding edge optimal loss factor defined by Eq. (13). Now, one needs to describe how η_{opt}^b evolves in function η_{opt}. To do so, the bonded edge optimal loss factor is com-

TABLE I. Properties used for the linear relation between the two optimal loss factors

s	ν	c_b	b	ms	K_b (Pa)	η^b_{opt}
3.35	0.2	1.0077	0.083	0.0977	4,038,500	0.095
					448,720	0.105
					200,000	0.12
					134,620	0.13
3.95	0.1	1.008	0.054	0.1075	4,038,500	0.065
					448,720	0.08
					200,000	0.095
					134,620	0.105
1.25	0.3	1.4997	0.001	0.2181	4,038,500	0.015
					448,720	0.04
					200,000	0.055
					134,620	0.07
1.4	0.2	1.5103	0.001	0.2196	4,038,500	0.015
					448,720	0.04
					200,000	0.055
					134,620	0.07
0.5	0.3	3.2744	0.003	0.2535	4,038,500	0.01
					448,720	0.02
					200,000	0.035
					134,620	0.04
0.6	0.1	3.3316	0.000	0.3123	4,038,500	0.01
					448,720	0.025
					200,000	0.035
					134,620	0.045

puted for a number of couples (s,ν). The bulk density of the sample is fixed to 90 kg m^{-3}, its diameter to 30 mm, and the characteristic impedance of the air is fixed to 412 Pa s m^{-1}. The couples (s,ν), the bulk modulus K_b and the bonded edge correction factor c_b are given in Table I. Figure 6 shows the linear relation linking the two optimal loss factors for 3 correction factors c_b. The three graphics correspond to the three sections of Table I, respectively. Each correction factor is obtained from two different couples (s,ν) and a section corresponds to an approximately equivalent correction factor (e.g., 1.0, 1.5, or 3.3). All these cases lead to a linear relation of the form:

$$\eta^b_{\text{opt}} = \frac{\pi s}{c_b} m(c_b)\,\eta_{\text{opt}} + b(c_b), \tag{21}$$

where m and b are two factors depending on the bonded edge correction factor c_b.

These factors are computed for a range of correction factors $c_b \in [0,5]$. Figure 7 represents ms and b as functions of the correction factor. For each coefficient, a polynomial approximation is carried out. These polynomial approximations enable a significant simplification for computing the bonded optimal loss factor from the sliding optimal loss factor given by Eq. (13). These polynomial relations are also plotted in Fig. 7 and are expressed as follows:

$$\begin{aligned} c_b < 2 \Rightarrow m(c_b)s &= 0.0028x^5 - 0.0484x^4 + 0.3206x^3 \\ &\quad - 1.0238x^2 + 1.6004x - 0.7462, \\ c_b \geqslant 2 \Rightarrow m(c_b)s &= 0.0256x + 0.1876, \end{aligned} \tag{22}$$

and

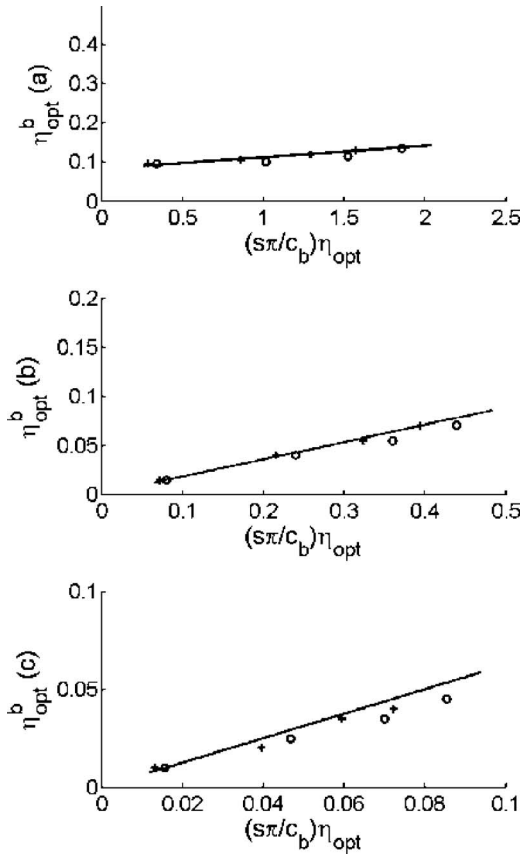


FIG. 6. Linear relation between the bonded and the sliding optimal damping loss factors. (1) c_b=1.0, (2) c_b=1.5, and (3) c_b=3.3.

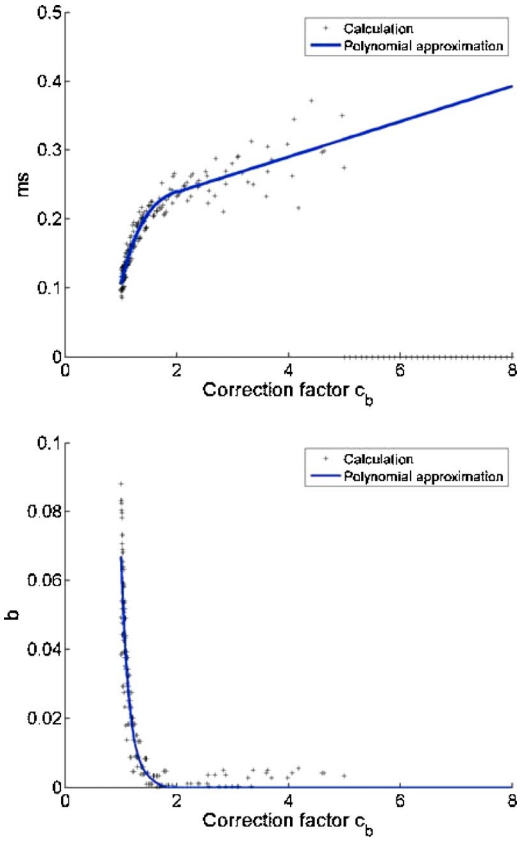


FIG. 7. (Color online) Computations of m and b in function of the bonded correction factor c_b.

$$c_b < 2 \Rightarrow b(c_b) = 0.179x^6 - 2.0054x^5 + 9.2993x^4 - 22.859x^3 + 31.447x^2 - 22.996x + 7.0021,$$
$$c_b \geqslant 2 \Rightarrow b(c_b) = 0. \tag{23}$$

Finally, assuming that the Poisson's ratio is known, the bonded edge correction factor c_b is deduced from the abacus previously computed (see Fig. 5). The coefficients m and b are calculated with relations (22) and (23). Then, the bonded edge optimal loss factor η^b_{opt} is deduced from Eq. (21). The actual damping loss factor is, at last, calculated from the sound absorption coefficient at the resonance using Eqs. (16) and (20).

3. Finding Poisson's ratio

By looking at Eq. (1), one can note that the product $f_1 L$ is kept for the same material. Thus, considering two samples defined by $[\rho_1, K_b, s_1, L_1, f_1^b(\nu, s_1)]$ and $[\rho_1, K_b, s_2, L_2, f_2^b(\nu, s_2)]$, residue R is defined as follows:

$$R = \left| \frac{f_1^b(\nu, s_1)}{c_b(\nu, s_1)} L_1 - \frac{f_1^b(\nu, s_2)}{c_b(\nu, s_2)} L_2 \right|. \tag{24}$$

This residue is equal to zero for an admissible solution of the Poisson's ratio. Due to the nature of c_b, R is similar to a second order polynomial having two roots. Sometimes, two solutions are admissible for the Poisson's ratio. To get around this problem, a third sample is used. This sample is defined by $[\rho_1, K_b, s_3, L_3, f_3^b(\nu, s_3)]$. The residue R becomes

$$R = \left| \frac{f_1^b(\nu, s_1)}{c_b(\nu, s_1)} L_1 - \frac{f_1^b(\nu, s_2)}{c_b(\nu, s_2)} L_2 \right| + \left| \frac{f_1^b(\nu, s_1)}{c_b(\nu, s_1)} L_1 - \frac{f_1^b(\nu, s_3)}{c_b(\nu, s_3)} L_3 \right|. \tag{25}$$

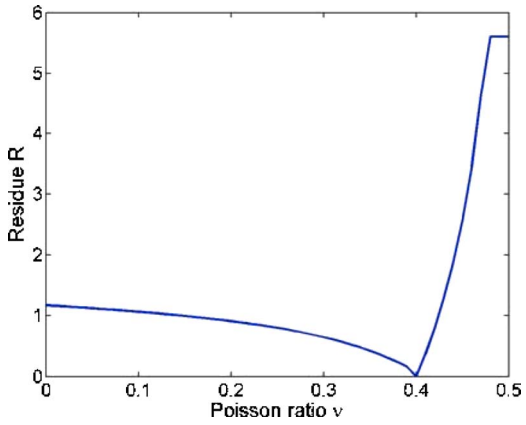


FIG. 8. (Color online) Residue in function of the Poisson's ratio.

Independently of the Young's modulus, the residue is equal to zero for the actual value of the Poisson's ratio. Thus, knowing the three frequencies f_1^b, f_2^b, and f_3^b and the geometries of the samples $(L_1, s_1, L_2, s_2, L_3, s_3)$, the Poisson's ratio can be deduced from the computation of residue R. This computation is illustrated in Fig. 8.

III. EXPERIMENTAL RESULTS

In this section, the previous theory is experimentally tested on two expanding closed foams: one with a sliding edge condition and one with a bonded edge condition. The normal sound absorption coefficients are measured with a 29-mm Bruel & Kjaer 4206 impedance tube following standard ASTM E1050-98 or ISO 10534-2:1998.

A. Sliding edge condition

In this first test, the sample is cut using pressurized water jet to ensure nice circularity and a diameter of $29.0^{+0.0}_{-0.2}$ mm. With this diameter, once mounted in the 29-mm impedance tube, it is assumed the sample can slide freely along the tube axis, and leakage around the sample is negligible. In this case, the first compression resonance frequency and related maximum absorption are easily determined from the measured absorption shown in Fig. 9. These results together with the known material properties and dimensions are listed in Table II

Assuming that the Poisson's ratio is equal to 0.33, the bulk modulus is deduced from Eq. (2) and the Young's modulus using Eq. (4). This yields, respectively, $K_b = 546\,362$ Pa and $E = 368\,753$ Pa. On the other hand, the optimal damping loss factor, reduced damping loss factor, and bulk damping loss factor are obtained from Eqs. (13), (16), and (14), respectively. This yields, $\eta_{\text{opt}} = 0.041$, $\eta_r^{\text{sol2}} = 11.160$, and $\eta = 0.458$.

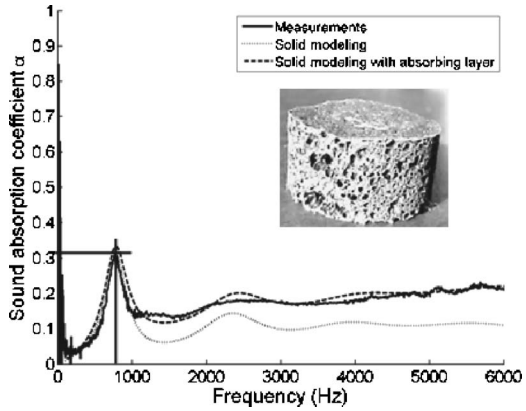


FIG. 9. Normal sound absorption coefficient of the tested expanding closed-cell foam with the sliding edge mounting condition. Comparison between impedance tube measurement, solid simulation, and solid simulation with surface absorption.

Using the found parameters, the sound absorption coefficient predicted using the equivalent solid model [Eqs. (5) and (7)] is compared to impedance tube measurements in Fig. 9. The simulation correlates well with the measurement at the resonance peak. However, for frequencies higher than 1000 Hz, the predicted absorption is underestimated. This is due to the fact that the surface sound absorption is not considered with the equivalent solid model (i.e., surface is assumed impervious). In order to correct for this, the solid model with surface absorption proposed in Ref. 4 can be used. In this case, one needs to define the thickness of the surface absorption layer by

$$d = \frac{1}{2}\sqrt{\frac{2P_0}{\omega_{\min}\sigma\phi}}, \tag{26}$$

where $\omega_{\min}$ is a low frequency (here 628 rad s^{-1}, 100 Hz, is used), P_0 the atmospheric pressure (101 325 Pa), and $\sigma\phi$ the product of the static airflow resistivity and open porosity of the surface absorption layer. In this case, the value of $\sigma\phi$ was iteratively found to be $\sigma\phi=10^7$ N s/m^4. This yields, from Eq. (26), $d=2.8$ mm, i.e., the thickness of the surface absorption layer (thin resistive layer accounting for the fact the surface is not purely impervious).

Consequently, with a 2.8-mm surface absorption layer covering an 11.3-mm equivalent solid core (total thickness is 14.1 mm), the prediction of the surface absorption model is compared to measurements in Fig. 9. This time the correlation with measurements is excellent for the whole frequency range.

For this material, one can note that the sound absorption at the first resonance is relatively low. This is logical since the actual damping loss factor of the material (0.458) is very far from the optimal value (0.041).

TABLE II. Properties of the 29-mm diameter sample of expanding closed-cell foam used for the sliding Edge Condition test.

Bulk density ρ_1 (kg m^{-3})	299
Thickness L (mm)	14.1
Resonance frequency f_1 (Hz)	776
Sound absorption $\alpha(\omega_1)$	0.308

TABLE III. Properties of the three 29-mm diameter samples of the expanding closed-cell foam used for the bonded edge condition test.

	Sample 1	Sample 2	Sample 3
Bulk density ρ_1 (kg m^{-3})	83.0	80.4	81.8
Thickness L (mm)	18.8	27.1	35.8
Shape factor s	1.60	1.11	0.84
Resonance frequency f_1 (Hz)	1176	992	896
Sound absorption $\alpha(\omega_1)$	0.56875	0.53925	0.51417

B. Bonded edge condition

For this second test, the method is now experimentally tested on samples with the bonded edge mounting condition. Three samples of a different length are required for computing the Poisson's ratio. Here, each foam sample was carefully and directly heat expanded in a hollow cylinder having a 29-mm inner diameter. The part of the foam running over the ends of the cylinder was cut. Then, the cylinder was mounted on the tube using a special end termination ensuring a rigid backing.

The data related to each sample are summarized in Table III, and the absorption curves are given in Fig. 10. Using these data, the Poisson's ratio is calculated with the residue R defined by Eq. (25). This yields $\nu=0.43\pm0.005$.

The correction factor can now be read from the abacus in Fig. 5 and the bulk modulus can be computed for each sample from Eq. (19). In order to obtain a good approximation of the Poisson's ratio, the three shape factors must be different. Moreover, from this abacus, one can note that the correction factors converge to 1 when either the shape factor is increased or the Poisson's ratio is increased. It is thus easier to find low Poisson's ratio ($\nu<0.45$). The ability to find the Poisson's ratio is increased as the shape factors are decreased.

The Young's modulus can now be deduced from the found Poisson's ratio and the average bulk modulus using Eq. (4). This yields $E=214\,897\pm2853$ Pa. Finally, the optimal damping loss factor is obtained from Eqs. (21) and (13). Similarly to the sliding condition, the actual bulk damping loss factor is deduced from Eqs. (16) and (20). This yields $\eta=0.22\pm0.04$.

The sound absorption coefficient is now computed with the equivalent solid model and compared to the impedance tube measurements in Fig. 10. The correlation with measurements is good especially at the resonance peak. Again, the surface absorption model [4] is used in order to correct for the underestimation of the predicted absorption (here, $\sigma\phi=5\times10^6$ N s/m^4 and $d=4.01$ mm were used). The residual variation between these latter predictions and the measurements may be attributed to the fact that the real elastic parameters of the expanded foam are not necessarily constant with the frequency.

IV. CONCLUSION

In this work, a method was proposed to determine the bulk elastic properties of soft equivalent solids or closed-cell foams from simple impedance tube absorption tests. The method was tested experimentally with success in laboratory.

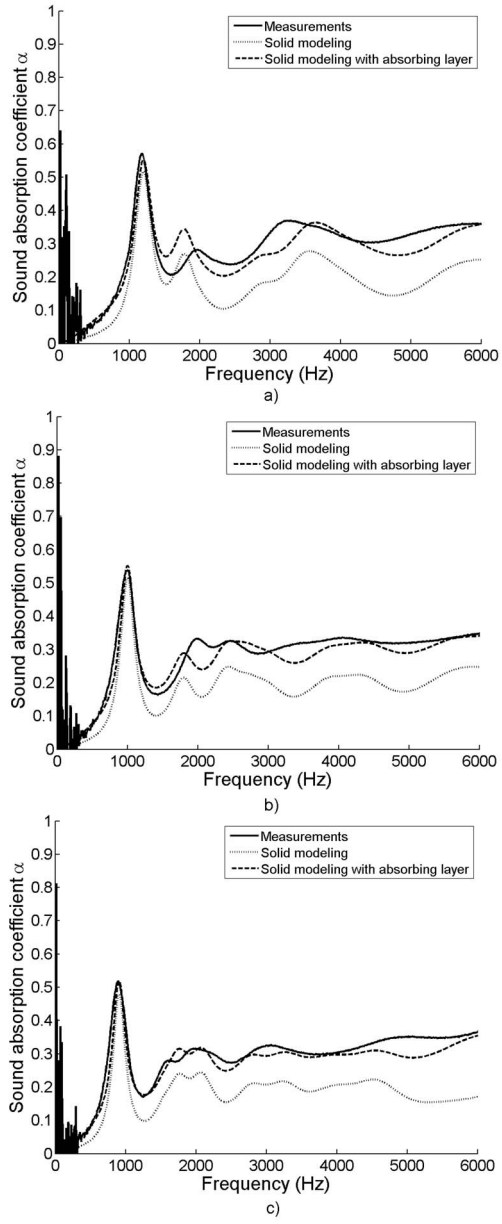


FIG. 10. Normal sound absorption coefficients of the three samples of the tested expanding closed-cell foam with the bonded edge mounting condition. Comparison between impedance tube measurement, solid simulations, and solid simulation with surface absorption. (a) Sample 1; (b) Sample 2; and (c) Sample 3.

It has revealed that the properties found with the proposed method can be used in the surface absorption solid model worked out in Ref. 4 to yield excellent correlations with measurements.

However, the accuracy of the method relies mostly in the proper control of the mounting conditions in the impedance tube. This is actually the most important limitation of the method, especially on small sample diameters for which the absorption measurement is very sensitive to boundary conditions. Further tests are required to validate the robustness of the proposed method.

ACKNOWLEDGMENTS

This work was supported by Henkel Technologies, NSERC Canada, FQRNT Quebec, and NCE Auto21.

[1]M. Biot, "The theory of propagation waves in a fluid-saturated porous solid. I. Low frequency range. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[2]J. F. Allard, *Propagation of Sound in Porous Media. Modeling Sound Absorbing Materials* (Elsevier Applied Science, New York, 1993).
[3]U. Ingard, *Notes on Sound Absorption Technologies* (Noise Control Fundation, New York, 1994).
[4]J.-L. Wojtowicki and R. Panneton, "Improving the efficiency of sealing parts for hollow body network," SAE Tech. Paper, Document No. 2005-01-2279 (2005).
[5]C. Langlois, R. Panneton, and N. Atalla, "Polynomial relations for quasi-static mechanical characterization of poroelastic materials," J. Acoust. Soc. Am. **110**, 3032–3040 (2001).
[6]M. Melon, E. Mariez, C. Ayrault, and S. Sahraoui, "Acoustical and mechanical characterization of anisotropic open-cell foams," J. Acoust. Soc. Am. **104**, 2622–2627 (1998).
[7]Y. K. Kim and H. B. Kingsbury, "Dynamic characterization of poroelastic materials," Exp. Mech. **19**, 252–258 (1979).
[8]A. Wijesinghe and H. B. Kingsbury, "Complex modulus of a poroelastic column," J. Acoust. Soc. Am. **65**, 90–95 (1979).
[9]T. Pritz, "Transfer function method for investigating the complex modulus of acoustic materials: rod-like specimen," J. Sound Vib. **81**, 359–376 (1982).
[10]S. Sim and K.-J. Kim, "A method to determine the complex modulus and Poisson's ratio of viscoelastic materials from FEM applications," J. Sound Vib. **141**, 71–82 (1990).
[11]R. D. Blevins, *Formulas for Natural Frequency and Mode Shape* (Krieger Publishing Co., Malabar, FL,2001).
[12]L. J. Gibson and M. F. Ashby, *Cellular Solids: Structure and Properties*, 2nd ed. (Cambridge University Press, Cambridge, 1997).
[13]J.-F. Allard and P. Delage, "Free field measurements of absorption coefficients on square panels of absorbing materials," J. Sound Vib. **101**, 161–170 (1985).

High-resolution plane-wave decomposition in an auditorium using a dual-radius scanning spherical microphone array

Boaz Rafaely,[a] Ilya Balmages,[b] and Limor Eger[c]
Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

(Received 17 January 2007; revised 19 August 2007; accepted 20 August 2007)

The spatial and temporal distribution of early reflections in an auditorium is considered important for sound perception. 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 analysis based on the windowed Fourier transform is employed to detect direction of arrival of individual reflections at selected frequencies. Experimental results of sound-field analysis in a real auditorium are also presented.
 [DOI: 10.1121/1.2783204]

PACS number(s): 43.55.Mc, 43.55.Br, 43.60.Fg [NX] Pages: 2661–2668

I. INTRODUCTION

The way in which sound propagates in an auditorium from the stage area to the seating area defines the acoustic performance of the auditorium. The required spatial and temporal behavior of the sound field depends on the purpose of the auditorium, e.g., concert hall, opera house, or lecture hall. Recent advances in the ability to measure and analyze the spatial behavior of sound fields in auditoria were facilitated by microphone arrays. In particular, the use of spherical microphone arrays[1–3] was employed in the directional analysis of enclosed sound fields. Gover *et al.* presented a 32-element spherical array which was used for directional sound-field analysis in rooms.[4,5] Rigelsford and Tennant designed a 64-element volumetric array[6] and employed beamforming for acoustic imaging in a room.[7] Park and Rafaely[8] used a 98-element spherical scanning array around a rigid sphere to analyze the sound field in an auditorium using plane-wave decomposition.[9] Although systems with a much smaller number of microphones have been presented for directional sound-field analysis in rooms,[10] these have limited spatial resolution and therefore rely on temporal separation.

Recently, Balmages and Rafaely[11] presented the design of spherical microphone arrays with open-sphere configurations. These are useful for spatial analysis at relatively low frequencies where an array with a large radius is required to achieve high spatial resolution. In this case an array configured around a large rigid sphere might not be practical. Therefore, open-sphere configurations were studied which overcome the poor numerical conditioning associated with open-sphere arrays that use omni-directional microphones.[11,12] In particular, a dual-sphere configuration was presented with an optimal ratio between the two radii, which depends on the array order and the frequency range. This paper presents several contributions. First, the paper presents an experimental verification of the dual-sphere microphone array and its application to sound-field analysis in an auditorium. Second, as the measurement system is based on a rotating microphone sampling a spherical surface at 882 positions and achieving a Rayleigh resolution of 9° in separating incoming room reflections, this paper provides an experimental validation to the practicality of high spatial-resolution room acoustics measurement systems. Finally, the paper proposes the use of the windowed Fourier transform for the time-frequency-space analysis of the sound field.

II. SPHERICAL ARRAY PROCESSING

The spherical array processing employed in this work is outlined in this section. The reader is referred to previous work[2,11] for further details. The sound pressure on the surface of an open or an imaginary sphere as measured by omnidirectional or pressure microphones is denoted by $p(k,\Omega)$, where k is the wave number. Here $\Omega \equiv (\theta, \phi)$ and r is the sphere radius, both combined to define the spherical coordinates (r, θ, ϕ). The first step in this spherical-harmonics based array processing is to compute the spherical Fourier transform[13] of the pressure on the sphere given by

$$p_{nm}(k) = \int_{\Omega \in S^2} p(k,\Omega) Y_n^{m^*}(\Omega) d\Omega \quad (1)$$

where $Y_n^m(\Omega)$ are the spherical harmonics[14] of order n and degree m. In practice, the pressure on the sphere surface is spatially sampled by microphones (see, for example, Rafaely[2] for a review of sampling methods). In this work we use the Gaussian sampling method, and so the spherical Fourier transform of the pressure is approximated by the summation

[a]Electronic mail: br@ee.bgu.ac.il
[b]Electronic mail: balmages@bgu.ac.il
[c]Electronic mail: egerl@bgu.ac.il

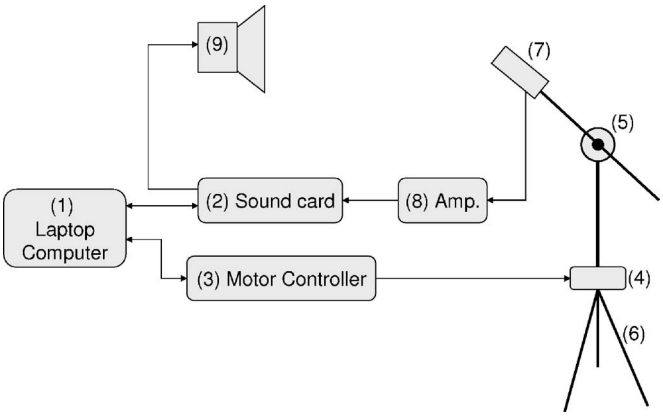


FIG. 1. Diagram of the measurement system: (1) IBM Thinkpad X41 Laptop computer; (2) Ego-Sys Waveterminal U2A stereo sound card with USB connection; (3) MicroLynx motor controller; (4) Velmex B5990TS rotating stage; (5) Newport M-RS40 manual rotating stage; (6) Manfrotto 475 tripod; (7) Bruel & Kjaer 4133 microphone with a 2669 preamplifier; (8) Bruel & Kjaer Nexus microphone conditioning amplifier; and (9) KRK RP-6 loudspeaker with build-in power amplifier.

$$p_{nm}(k) = \sum_{j=1}^{M} a_j p(k,\Omega_j) Y_n^{m^*}(\Omega_j), \tag{2}$$

where $\Omega_j = (\theta_j, \phi_j)$ are the samples on the sphere representing the microphones positions, and a_j are the sampling weights which depend on the sampling configuration. M represents the total number of samples. As shown previously[9] plane-wave decomposition is calculated as the spherical Fourier transform of p_{nm}/b_n, where for an open sphere

$$b_n(k) = 4\pi i^n j_n(kr), \tag{3}$$

where j_n is the spherical Bessel function. In this work we also introduce a weighting vector c_n as suggested by Meyer and Elko[1] and also presented in recent work.[2,15] The array output can now be written as

FIG. 2. Picture of the rotating microphone mechanical system.

FIG. 3. (Color online) A picture of the Sonnenfeldt auditorium.

$$y(k,\Omega_0) = \sum_{n=0}^{N} \sum_{m=-n}^{n} c_n(k) \frac{p_{nm}(k)}{b_n(k)} Y_n^{m^*}(\Omega_0), \tag{4}$$

where N is the array order, and c_n can be frequency dependent. Array output is a function of Ω_0, the array look direction.

In this paper we use the dual-sphere microphone array, as recently presented by Balmages and Rafaely,[11] in order to overcome the numerical ill-conditioning at nodal frequencies of the spherical Bessel function when a single sphere is used with pressure microphones. The formulation is similar to the one presented here, only the functions p_{nm} and b_n are the result of combining data from the two radii. Let p_{1nm} and b_{1n} denote data from the first sphere of radius $r=r_1$, and p_{2nm} and b_{2n} denote data from the second sphere of radius $r=r_2$. We would like to avoid using b_n with small values and so we select at each wave number k and order n the data set with the largest value of $|b_n|$. We first define a radius selection parameter

$$\beta_n(k) = \begin{cases} 0, & |b_{1n}(k)| \geqslant |b_{2n}(k)| \\ 1, & |b_{1n}(k)| < |b_{2n}(k)| \end{cases} \tag{5}$$

and so the combination of data from the two radii is given by

$$p_{nm}(k) = [1 - \beta_n(k)] p_{1nm}(k) + \beta_n(k) p_{2nm}(k), \tag{6}$$

$$b_n(k) = [1 - \beta_n(k)] b_{1n}(k) + \beta_n(k) b_{2n}(k). \tag{7}$$

Now, Eq. (4) can be used with the combined data.

III. SOUND FIELD ANALYSIS

Sound field analysis can be performed by analyzing array output $y(k,\Omega_0)$ in various ways as presented in this section.

A. Narrow-band sound field directivity

Here $y(k,\Omega_0)$ is evaluated for the entire range of $\Omega_0 = (\theta_0, \phi_0) \in ([0,\pi],[0,2\pi])$, and a narrow bandwidth around $k=k_0$. A two-dimensional (2D) plot of $|y|$ as a function of (θ_0, ϕ_0) can therefore be presented,[9] where significant peaks in magnitude represent arrival directions of (plane) waves with significant amplitude.

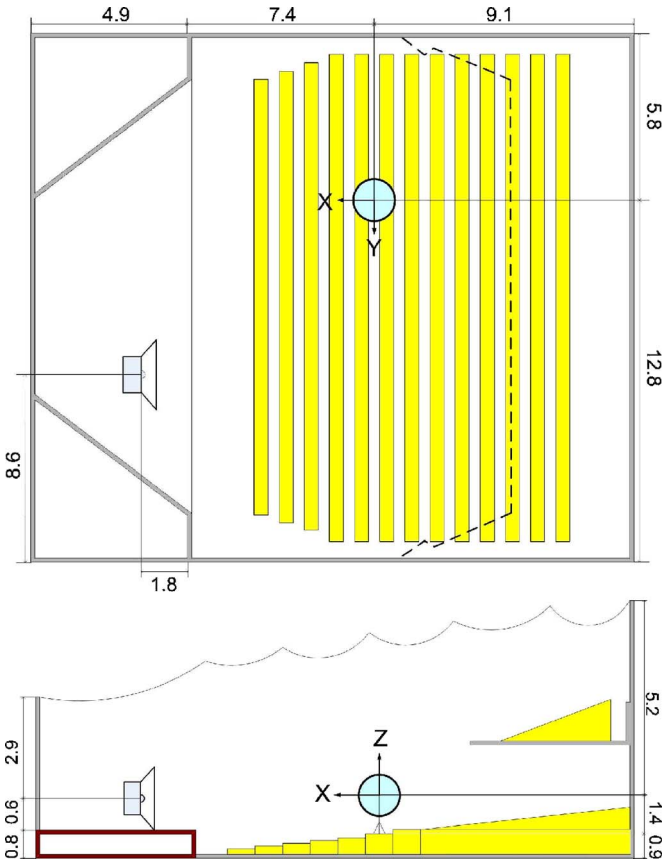


FIG. 4. (Color online) Schematic diagram of the Sonnenfeldt auditorium and the experimental setup: (upper) top view; (lower) side view. Loudspeaker and array (circle) are marked. Distances are in meters.

B. Directional impulse response

By taking the inverse Fourier transform (or the inverse discrete Fourier transform for discrete frequency values) of $y(k,\Omega_0)$, where $k=2\pi f/c$, with f the frequency and c the speed of sound, we get $y(t,\Omega_0)$, which represents the directional impulse response at look direction Ω_0. We expect enhanced contribution of waves arriving from directions close to Ω_0 in the response y.

C. Sound-field directivity with time-frequency selectivity

To achieve plane-wave decomposition we aim to separate the contribution of individual waves to the array output. Beamforming as presented here allows spatial separation, but due to the impulsive nature of the signal in a room impulse response, we can use time windowing to select a narrow section of the impulse response with only a small number of significant waves.[8] Time windowing can be combined with frequency selectivity by using a modulated time window.[16] Window delay, width, and modulation frequency can control the position of the window in time, and its width in both time and frequency. Denoting by $g(t-\tau)$ the time window delayed by τ, and by f_0 the modulation frequency, the windowed Fourier transform of a directional impulse response is given by[16]

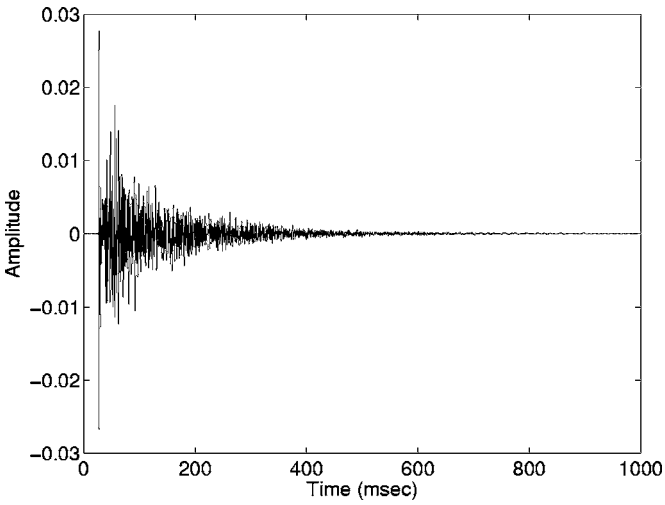


FIG. 5. Impulse response from the loudspeaker to the microphone at position $(r,\theta,\phi)=(0.4,0,0)$.

$$z(\tau,f_0,\Omega_0)=\int_{-\infty}^{\infty} y(t,\Omega_0)g(t-\tau)e^{i2\pi f_0 t}\,dt. \tag{8}$$

Further, the time-averaged power of z aimed at removing fluctuations in z typically at f_0, can be calculated for example using similar windowed Fourier transform with a modulation frequency set to zero

$$\xi(\tau,f_0,\Omega_0)=\int_{-\infty}^{\infty} |z(t,f_0,\Omega_0)|^2 g(t-\tau)dt, \tag{9}$$

with ξ representing the envelope of the power of z, and can therefore be used in the analysis of the power spectral density over time, frequency, and space. Also note that $\xi(\tau,f_0,\Omega_0)$ and $y(k,\Omega_0)$ are both functions of frequency and array look direction. However, ξ provides additional time selectivity and therefore improved overall spatial selectivity, but at the expense of potentially reduced frequency selectivity. The Heisenberg uncertainty theorem[16] suggests that the product of the time and frequency resolutions has a lower bound, and so greater time resolution comes at the expense of frequency resolution and vice versa.

TABLE I. Summary of data for the plane-wave decomposition analysis. For each of the six waves the table shows the arrival time, the distance traveled from the source to the array, the peak magnitude with an average-power analysis around 2800 Hz, the arrival direction, and the physical path followed in the auditorium from the source to the array.

No.	Time (ms)	Distance (m)	Peak level (dB)	θ	ϕ	Path
1	28.3	9.72	−5.26	96.4	25.7	Direct
2	42.5	14.58	−40.40	57.9	25.7	Ceiling reflector I
3	46.0	15.78	−51.54	36.4	25.7	Ceiling reflector II
4	48.3	16.58	−37.33	92.1	42.9	Left stage wall
5	57.2	19.61	−25.92	92.1	300.0	Right auditorium wall
6	63.7	21.84	−37.10	92.1	68.6	Left auditorium wall

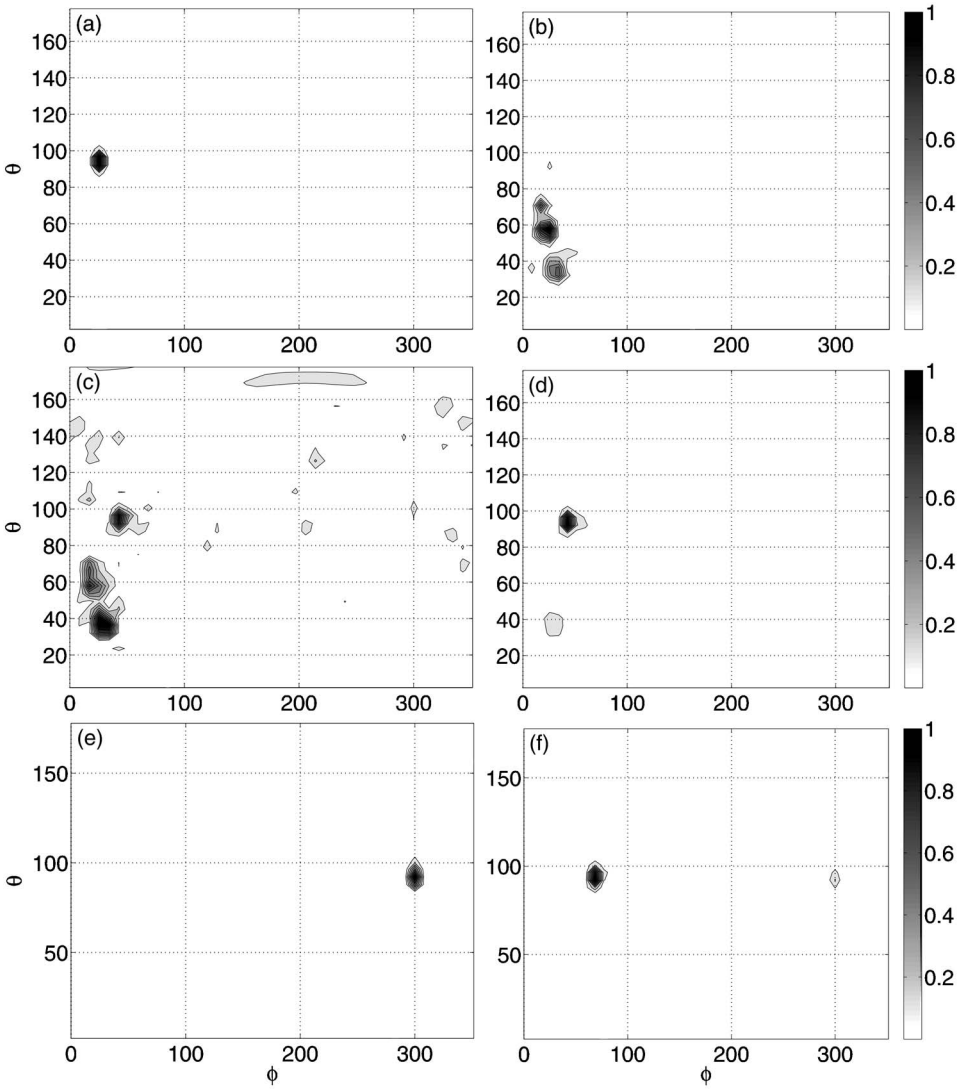


FIG. 6. Average-power function $\xi(\tau, f_0, \Omega_0)$, evaluated for $f_0=2800$ Hz, and τ values for (a)–(f) equal the arrival time of waves 1–6 in Table I. The figures therefore represent the directional density over $\Omega_0 \in ([0,\pi],[0,2\pi])$, (represented in degrees in the figures) of the averaged power of the array output localized around 2800 Hz and $\tau_i, i=1,\ldots,6$. Levels have been normalized to peak values shown in Table I.

IV. THE SPHERICAL MICROPHONE ARRAY SYSTEM

In this section the microphone array system employed in this paper is presented. A microphone placed on a boom attached to a motorized rotating system is used to spatially sample the sound field around the surface of a sphere. The sphere radius is set by the position of the microphone on the boom. The azimuth of the microphone position is controlled by a computer via a motor controller and a rotating stage, while the elevation of the microphone position is controlled via a manual rotating stage. Once the microphone is positioned, the impulse response between a loudspeaker placed on the auditorium stage and the microphone placed in the seating area is measured using a linearly swept-sine signal of 6 s duration in the range 0–3000 Hz. A sound card connected to a computer running MATLAB is used to play and record the signals. An entire measurement session includes measuring and saving the impulse response data for microphone positions covering the surface of two spheres with different radii. A schematic diagram of the microphone array measurement system is presented in Fig. 1, including details of the measurement equipment employed. Figure 2 shows a picture of the rotating microphone mechanical system.

V. SPHERICAL ARRAY DESIGN

The microphone array measurement system was designed to achieve high resolution with minimal error in the operating frequency range, as summarized below. The reader is referred to Rafaely[2] for further details on array design.

- A measurement area covering the surfaces of two spheres with radii of 40 and 43 cm was employed in this dual-sphere array configuration, to avoid errors due to nulls in b_n. The ratio of the two radii was close to the optimal value[11] given by $[1+(\pi/2N)]$.
- A spherical array order of $N=20$ was employed to achieve a Rayleigh resolution limit[17] of $\pi/N=9°$,[9] which represents the ability to separate two proximate plane waves. The order was limited only by practical constraints relating to the overall duration of the experiment which increases with N and the total number of spatial samples.

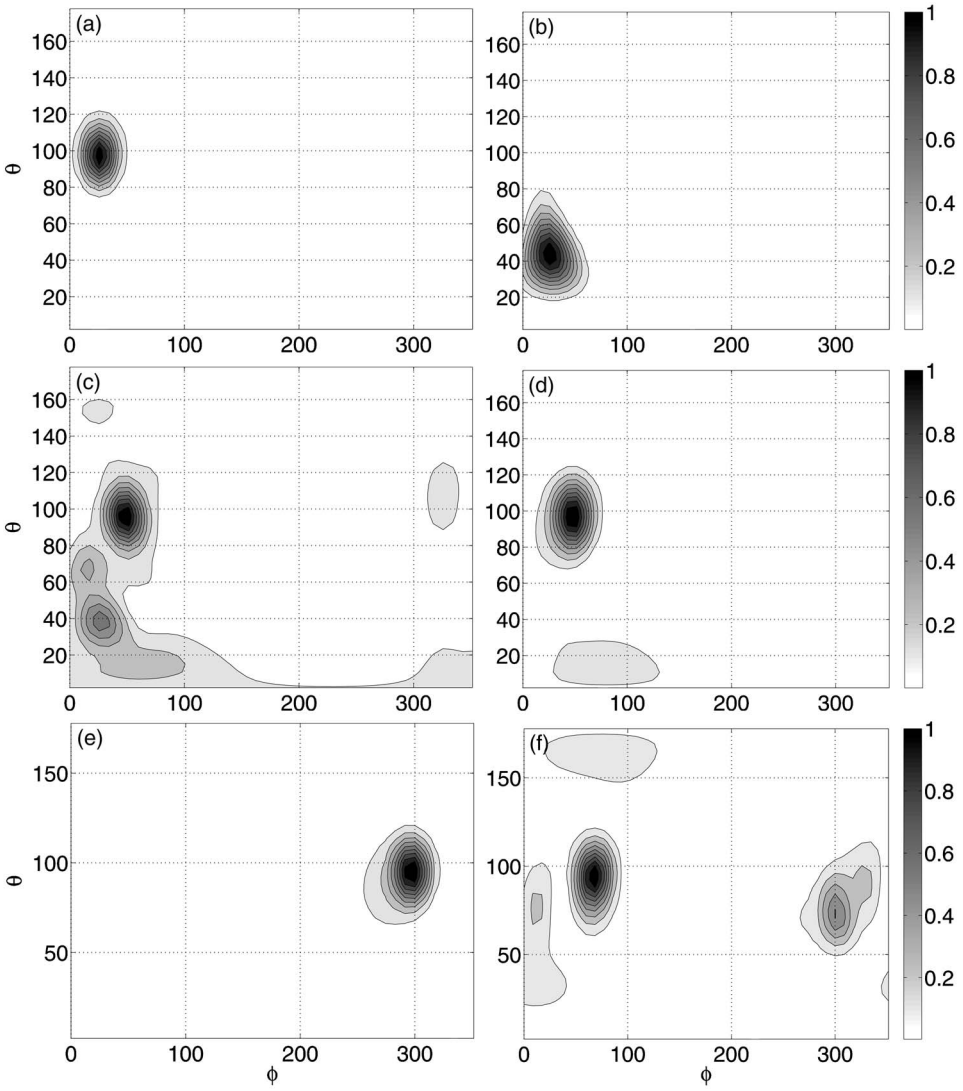


FIG. 7. Average-power function $\xi(\tau, f_0, \Omega_0)$, evaluated for $f_0=800$ Hz, and τ values for (a)–(f) equal to the arrival time of waves 1–6 in Table I. The figures therefore represent the directional density over $\Omega_0 \in ([0,\pi],[0,2\pi])$, (represented in degrees in the figures) of the averaged power of the array output localized around 800 Hz and $\tau_i, i=1,\ldots,6$. Levels have been normalized to peak values.

- Microphone positions followed the Gaussian sampling method,[2] with a total of $2(N+1)^2=882$ samples per sphere.
- An operating frequency range of $f<3000$ Hz was defined, with the upper limit set to reduce spatial aliasing, by avoiding frequencies above the limit of $kr \approx N$.[2] A corresponding sampling rate of 6000 Hz was employed.
- An array weighting function of $c_n=|b_n|^2$ in Eq. (4) was used, which achieved constant robustness due to the constant white-noise gain.[15]

VI. EXPERIMENTAL PROCEDURE

The measurement system was positioned in the Sonnenfeldt auditorium at Ben-Gurion University. The loudspeaker was placed on the stage, and the rotating array at the seating area in the sixth row. Figure 3 shows a picture of the Sonnenfeldt auditorium. The auditorium has 444 seats with approximate dimensions $21\times18\times6=2268$ m^3. A reverberation time of 0.95 s was measured by the system as detailed in the next section. The experiment was performed in the following stages:

(1) The loudspeaker and microphone system were placed in position, resetting the microphone position to $(\theta,\phi)=(0,0)$, with $\theta=0$ facing the ceiling, and $\phi=0$ facing the back stage wall.
(2) The loudspeaker played a 6 s long linearly swept-sine signal. The microphone signal along with the loudspeaker input signal were recorded by the computer.
(3) The microphone was then repositioned to the next spatial sampling position, and Stage 2 was repeated.
(4) Stage 3 was repeated to cover all positions for the two spheres.

At the end of the measurement session, the input and output signals for the loudspeaker-microphone response were recorded for the two spheres, and were used in the analysis

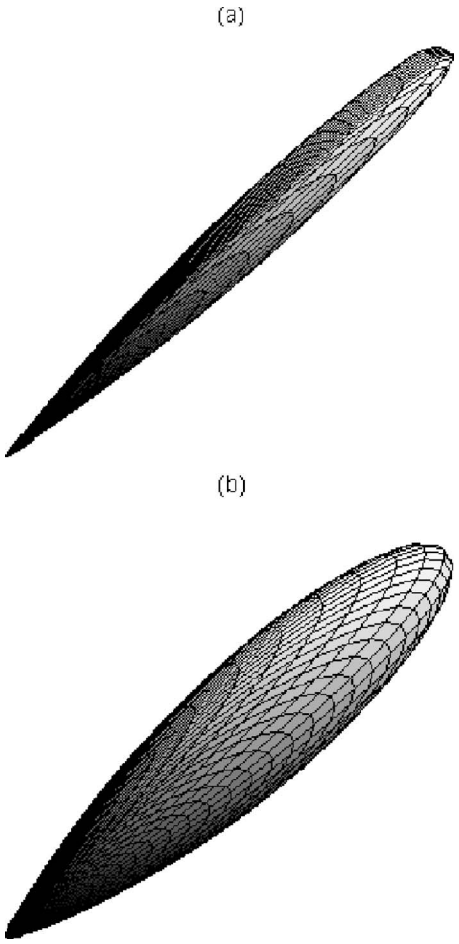


FIG. 8. (a) Presentation of Fig. 6(a) in a 3D directivity-type plot; and (b) presentation of Fig. 7(a) in a 3D directivity-type plot.

stage. Figure 4 presents a schematic diagram of the auditorium and the positions of the loudspeaker and microphone measurement array.

VII. ANALYSIS OF RESULTS

Figure 5 shows an example impulse response, measured from the loudspeaker to the microphone at position $(r,\theta,\phi)=(0.4,0,0)$. Impulse responses were computed as the inverse fast Fourier transforms (FFTs) of the ratio of the zero-padded FFTs of the microphone output signal and the loudspeaker input signal. The reverberation time was calculated from this impulse response[18] and was found to be 0.95 s. Sound field directivity $\xi(\tau,f_0,\Omega_0)$ was calculated using Eqs. (8) and (9) for two frequencies $f_0=2800$ and 800 Hz and plotted for time values with peak magnitude. A Hanning window of duration 3.3 ms (21 samples long at 6 kHz sampling rate) was used for function $g(t)$. The resulting function $\xi(\Omega_0)$ represented directional power distribution of the plane waves composing the sound field, localized in frequency and arrival time. Table I details the six arrival times investigated corresponding to the direct sound and first reflections in the auditorium. Figures 6 and 7 show $\xi(\theta_0,\phi_0)$ for $f_0=2800$ and 800 Hz, respectively, evaluated at the six arrival times given in Table I. Note that Figs. 6 and 7 show magnitude levels normalized to the peak level as detailed in Table I.

Figure 6(a) shows a clear peak around $(\theta,\phi)=(96.4°,25.7°)$, corresponding to the direct sound from the loudspeaker. Figures 6(b) and 6(c) show two main peaks and one smaller peak, arriving from directions corresponding to the ceiling. These are reflections from the ceiling curvatures, arriving with proximity in both space and time, as detailed in Table I (waves 2 and 3). Although the ceiling reflections are separated by less than 4 ms in time and 22° in space, they can be clearly distinguished due to the high spatial resolution facilitated by the array order of $N=20$. Similarly, Figs. 6(d)–6(f) show reflections from various auditorium walls. Note that the physical path of each reflection from the loudspeaker to the microphone as detailed in Table I was deduced by correlating travel distances and arrival direction of reflections as measured by the array with the auditorium geometry and dimensions.

Figure 7 shows $\xi(\Omega_0)$ similar to Fig. 6, but with $f_0=800$ Hz. Due to the array processing employed with $c_n=|b_n|^2$ to maintain constant robustness at the entire frequency range,[15] the effective array order decreased at low frequencies, and at 800 Hz it is about $N\approx 6$. It is clear that a reduction in effective array order results in a reduction in spatial resolution, and although most of the reflections can still be identified, Figs. 7(b) and 7(c) show that the two ceiling reflections (2 and 3 in Table I) cannot be clearly distinguished, but rather merge into a single peak. This result clearly emphasizes the need for a measurement and analysis system with a high spatial resolution for room acoustics applications. Table I also confirms that reflections with peak magnitude smaller by more than 45 dB compared to the direct sound, e.g., reflection 3, can be clearly detected as illustrated in Fig. 6(c).

Figure 8 presents the data in Figs. 6(a) and 7(a) in a 3D directivity-type plot. This provides an illustration of the 3D directional characteristics of the array and associated processing in detecting the direct sound.

Directional impulse responses $y(t,\Omega_0)$ were computed as described in Section III, for various look directions Ω_0 corresponding to the reflections arrival directions detailed in Table I. Figure 9 shows all six directional impulse responses, with the inverted triangle at the top of each figure representing the reflection arrival time detailed in Table I. Figure 9(a) shows that the direct sound has been emphasized compared to the remaining part of the impulse response. Figures 9(b) and 9(c) confirm that the ceiling reflections are of relatively low level, even with the array looking at their arrival direction. In Fig. 9(d) the reflection from the side stage wall is emphasized, although the direct sound is also clear. This is since the direct sound and the reflection from the side stage wall are relatively proximate in their arrival direction (see Table I), and so the spatial separation by the array is limited. Figures 9(e) and 9(f) show clear peaks corresponding to the auditorium walls reflections. Note that a high-pass filter at 500 Hz has been employed on the directional impulse responses to reduce low-frequency magnitude which was enhanced due to the relatively low array directivity at this frequency range.

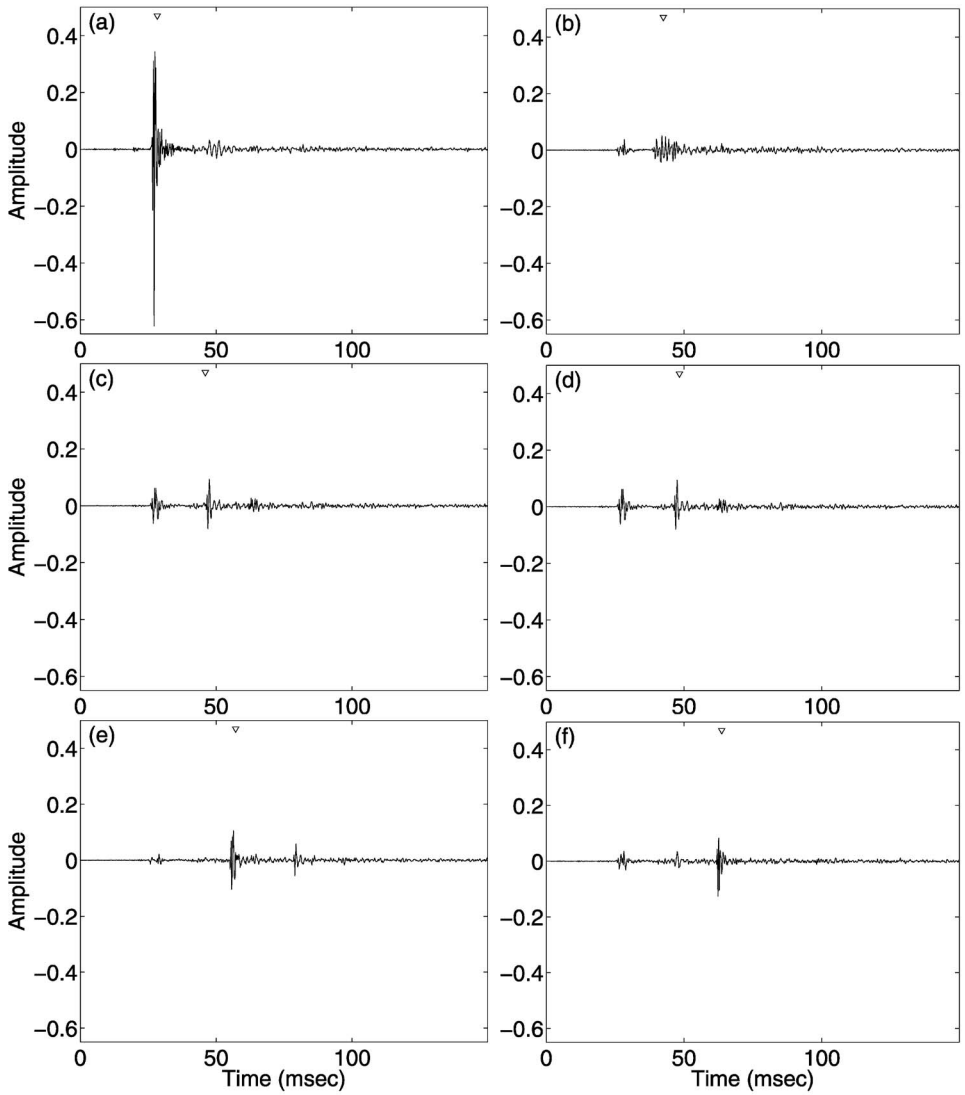


FIG. 9. Directional impulse responses, $y(t,\Omega_0)$, for look directions corresponding to the arrival directions of the direct sound and reflections as detailed in Table I, with (a)–(f) representing waves 1–6 in Table I. The inverted triangle at the top denotes the approximate arrival time of the wave incident from the array look direction.

VIII. CONCLUSION

This paper presented a sound field measurement and analysis system suitable for the study of auditorium acoustics. The use of a dual-sphere array and corresponding array processing to improve array robustness at the null frequencies of the spherical Bessel function was verified experimentally. An array processing method based on the windowed Fourier transform has been proposed and investigated for the detection of the arrival times and directions of room reflections around selected frequencies. The sound field in a real auditorium was measured and analyzed, verifying the ability of the proposed system to detect room reflections, and even separate ceiling reflections which are proximate in both time and space. The use of optimal array processing methods for sound field analysis in an auditorium is proposed for future work.

[1]J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," Proceedings ICASSP 2002, **II**, pp. 1781–1784.

[2]B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**, 135–143 (2005).

[3]S. Moreau, J. Daniel, and S. Berlet, "3D sound field recording with higher order ambisonics—objective measurements and validation of a 4th order spherical microphone," Proceedings of 120th AES Convention, 6857, Paris, May 2006.

[4]B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," J. Acoust. Soc. Am. **112**, 1980–1991 (2002).

[5]B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurement of directional properties of reverberant sound fields in rooms using a spherical microphone array," J. Acoust. Soc. Am. **116**, 2138–2148 (2004).

[6]J. M. Rigelsford and A. Tennant, "A 64 element acoustic volumetric array," Appl. Acoust. **61**, 569–475 (2000).

[7]J. M. Rigelsford and A. Tennant, "Acoustic imaging using a volumetric array," Appl. Acoust. **67**, 680–688 (2006).

[8]M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," J. Acoust. Soc. Am. **118**, 3094–3103 (2005).

[9]B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," J. Acoust. Soc. Am. **116**, 2149–2157 (2004).

[10]H. Okubo, M. Otani, R. Ikezawa, S. Komiyama, and K. Nakabayashi, "A system for measuring the directional room acoustical parameters," Appl. Acoust. **62**, 203–215 (2000).

[11]I. Balmages and B. Rafaely, "Open-sphere designs for spherical micro-

phone arrays," IEEE Trans. Audio, Speech and Lang. Proc. **15**, 727–732 (2007).
[12]A. Laborie, R. Bruno, and S. Montoya, "A new comprehensive approach of surround sound recording," Proceedings of 114th AES Convention, 5715, Amsterdam, March 2003.
[13]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).
[14]G. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 5th ed. (Academic, San Diego, 2001).
[15]B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," IEEE Signal Process. Lett. **12**, 713–716 (2005).
[16]S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, San Diego, 1999).
[17]H. L. V. Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory*, Part IV, 1st ed. (Wiley-Interscience, New York, 2002).
[18]M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).

A spread spectrum technique for the study of outdoor noise propagation[a)]

David C. Waddington[b)] and Jamie A. S. Angus
Acoustics Research Centre, University of Salford, Salford M5 4WT, United Kingdom

(Received 17 May 2007; revised 10 August 2007; accepted 10 August 2007)

This paper describes field measurements to assess innovative correlation techniques for the study of meteorological and topographical effects on sound propagation. To take advantage of the properties of coded signals in a time-varying system, the correlation signal is produced by the modulation of a code sequence onto an acoustic carrier. An established method of increasing signal-to-noise ratio is to use correlation techniques with maximum length sequences. However, this standard method is restricted in its use outdoors because of the time-variant nature of the atmosphere. On the other hand, the correlation properties of a directly carrier-modulated code sequence modulation signal may be exploited in a time-varying environment. An experiment is described in which the correlation properties of the spread spectrum signal are demonstrated and are used to calculate accurate times of flight that compare well with sonic anemometer measurements of speed of sound. The results illustrate that an acoustical spread spectrum system can provide significantly improved ways of measuring sound propagation outdoors. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2783120]

PACS number(s): 43.60.Ek, 43.28.Tc, 43.28.Vd, 43.60.Rw [NX] Pages: 2669–2678

I. INTRODUCTION

Maximum length sequences (MLS) and related coded signals are widely used in acoustical applications. Nearly all sequences in common use possess two-valued autocorrelation functions and it is this correlation property that is exploited. An analysis of transfer function measurements based on maximum length sequences was presented by Rife[1], while Xiang[2] described simultaneous acoustic multiple-source measurements using orthogonal correlation sequences. Correlation methods have also been used for the investigation of outdoor noise propagation. There are, however, limitations due to the nonstationarity of the outdoor noise propagation system. The refraction of sound is an important feature of outdoor propagation. Refraction occurs due to changes in sound speed, which depends upon vector wind speed and temperature. However wind, temperature, and humidity vary spatially, both vertically and horizontally, and with time. Furthermore, turbulent fluctuations of these physical qualities result in the scattering of sound, another important feature of outdoor propagation. The combination of refractive and scattering effects results in changes in the direction and length of the sound propagation and deviation from the straight-line path. Moreover, these changes in the direction and length of the propagation path vary with time.

One assumption for the use of the MLS measurement technique is the time invariance of the system during the travel time of the sequence. This assumption often does not hold outdoors due to changes in the propagation path caused by wind and turbulence. One approach that exploits the advantages of the correlation method for the study of outdoor propagation is to use the pseudorandom sequence to modulate the amplitude of random noise. This is known as the *m*-sequence correlation method as presented by Hallberg.[3] Unfortunately, this results in a loss of the temporal resolution of the impulse response. Another way to avoid the problem of time variance is to reduce the length of the sequence until the system can be regarded as stationary. The maximal usable length of the correlation sequence then depends on the measurement distance, and Heutschi[4] obtained results at propagation distances up to 200 m for flat grassy terrain. However, the shorter the sequence duration, the more sensitive measurements become to fluctuations of meteorological quantities. What is required is some means of exploiting the advantages of these short sequences with the ability to perform long-term averaging in time varying environments. Such techniques would be useful in propagation model evaluation, in particular the effects of meteorology and topography, and in acoustic remote sensing, especially for the study of time-varying atmospheric properties. This paper describes a novel method of measuring outdoor sound paths via the use of correlation techniques using spread spectrum methods, in which the correlation signal is produced using a coded signal to modulate an acoustic carrier signal.

II. SPREAD SPECTRUM SIGNAL DESIGN

A. Spread spectrum modulation

When sound propagates in an inhomogeneous, turbulent, moving medium, it experiences time-varying delay, Doppler frequency shifts, and Doppler frequency spreading. In the fully general case, these can be quite complex, as discussed by Ostashev.[5] Spread spectrum modulation is one technique that can be employed to overcome these effects. In wireless

[a)]Aspects of this work have been presented at the 12th Long Range Sound Propagation Symposium, New Orleans, LA.
[b)]Author to whom correspondence should be addressed. Electronic mail: d.c.waddington@salford.ac.uk

communications, for example, the modeling and estimation of time-varying channels is an important design challenge, and the effects of time delays and Doppler frequency shifts caused by the mobility of objects must be considered.[6] Several spread spectrum techniques are available, including chirping, time hopping, and various hybrid combinations of modulation forms.[7] The two most prevalent forms of spread spectrum modulation, however, are direct sequence modulation and frequency hopping. Direct sequence systems are the best-known and most widely used spread spectrum systems, in part due to their relative simplicity. Furthermore, direct sequence spread spectrum modulation is well known for its high performance in multipath fading environments, where signal components arriving by different propagation path add destructively. In applications other than communications, direct sequence spread spectrum signals are used to obtain accurate range and velocity measurements in radar and navigation. The Jet Propulsion Laboratory (JPL) ranging technique described by Golomb[8] and used in various space applications is a direct sequence approach, and was an inspiration for this project.

Although ideally suited to ranging, implementation of the JPL ranging technique in the acoustic domain reveals that these signals are not suitable for environmental noise investigations due to their tonal nature and, consequently, exceedingly annoying sound. Nevertheless, it has been recognized that success of the ranging codes lay not just in the properties of the number sequences, but also in the application of spread spectrum techniques. Spread spectrum techniques have of course been used for many years in military applications, but have recently found widespread commercial applications in mobile phones, and in particular, GPS global positioning systems.

Spread spectrum systems, however, are less widely used in acoustics. Spread spectrum methods are employed in underwater acoustic communications systems since they offer an improved performance in time-varying multipath channels. A comprehensive overview of the field is given by Kilfoyle,[9] and an example of an application to oceanic acoustic tomography was presented by Zheng.[10] Indeed, much work has been presented addressing modulation and demodulation techniques motivated in part by the large time delay and Doppler spreads often encountered in underwater acoustic channels. In airborne sound, on the other hand, an experiment technique for the measurement of multipath propagation time delay has been presented by Swanson,[11] and a wide band characterization of the acoustic spreading function in the surface layer of the atmosphere was investigated by Mahon,[12] whilst more recently Mizutani[13] described implementing differential biphase shift keying for digital communication with robots.

Of the many application fields for spread spectrum techniques, the main properties of interest are high-resolution timing and ranging, multipath and interference rejection, and operation under time-varying propagation conditions. Driven by a revolution in the information society, considerable research effort has been invested in developing techniques to optimize these properties for radio channels, and application of just a few to the study of transmission over an acoustic medium could potentially yield great reward. In the investigation of outdoor noise propagation, time variance and multipath propagation are significant challenges. On the other hand, signal to noise is often a problem when using electroacoustic noise sources. Consequently, application of spread spectrum techniques implementing correlation sequences to the investigation of outdoor sound propagation is therefore an exciting opportunity.

This work describes the design and development of an acoustic signal for the study of meteorological and topographical effects on outdoor sound propagation. This approach to the investigation of outdoor sound propagation has significant advantages over traditional methods. Just as sound reflected from surfaces carries information about the properties of the material, similarly sound refracted during propagation carries information about the properties of the atmosphere. This raises the possibility of deriving detailed propagation data from a traditional measurement setup, significantly improving environmental noise investigations. This paper introduces the approach taken, and describes some results from a field measurement performed to test the performance of such a direct sequence spread spectrum signal.

B. Acoustic spread spectrum signal design

The correlation signal is produced by the modulation of a coded sequence onto an acoustic carrier, to allow exploitation of the properties of coded signals in a time-varying system. A pseudonoise (PN) sequence is a code sequence whose autocorrelation has properties similar to those of white noise. The most widely known binary PN code sequences are the maximum length shift register sequences, known as MLS or *m*-sequences. An important characteristic of the periodic PN sequence is its autocorrelation function, and since the sequence is periodic, the autocorrelation sequence is also periodic with the same period. For long *m*-sequences, the size of the off-peak values of the autocorrelation function relative to the peak value is small. Consequently, *m*-sequences are very close to ideal PN sequences in terms of the autocorrelation function. In the acoustic spread spectrum signal design presented here, the correlation properties of an *m*-sequence are utilized by the modulation of an acoustic carrier, resulting in a direct sequence spread spectrum (DSSS) signal. By combining correlation techniques such as MLS with spread spectrum techniques such as DSSS, an improved performance in time-varying airborne multipath channels may be achieved through coherent detection of high-resolution phase-modulated spreading codes. First, the binary sequence with elements $\{0,1\}$ is mapped onto a corresponding bipolar sequence with elements $\{+1,-1\}$. The acoustic carrier is then directly modulated by the bipolar sequence, as illustrated by Fig. 1. This process, expressed mathematically as

$$f(t) = A\cos\left(\omega t \pm \hat{a}_n \frac{\pi}{2}\right) \tag{1}$$

with amplitude A, frequency ω, and phase $\hat{a}_n . \pi/2$, where $\hat{a}_n$ is the bipolar sequence element $\{\pm 1\}$, is sometimes referred to as directly carrier-modulated code sequence modulation.

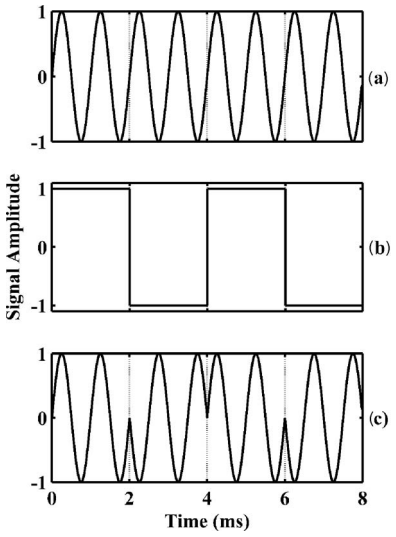


FIG. 1. Schematic showing direct sequence modulation of a 1 kHz acoustic carrier. (a) Acoustic carrier. (b) Spreading sequence. (c) Modulated acoustic signal.

A spectrum typical of this signal format is illustrated in Fig. 2 and has a main lobe bandwidth of $2R_c$, twice the clock rate of the code sequence used as a modulating signal. The code generator clock rate R_c is known as the chip rate, and so the chip time $1/R_c$ signifies the time occupied by the carrier for each code sequence element. The chip is the basic element in a DSSS signal. The center frequency of the main lobe is the acoustic carrier frequency f_c. Each of the sidelobes has a bandwidth from null-to-null that is equal to the clock rate R_c. Although the spectral power envelope of a direct sequence signal is of the form $[\sin(x)/x]^2$, with careful design the signal can be tailored to emulate the spectrum of octave band filtered white noise. This allows general application and analysis of the noise source without reference to the correlation properties. To illustrate this design feature, Fig. 2 shows the power spectrum density of acoustic spread spectrum signals created using the same spreading sequence but with differing numbers of cycles per chip. It can be seen that four cycles per chip of a 1 kHz phase modulated linear sequence produces a main lobe with a bandwidth that approximates the 1/3 octave filter.

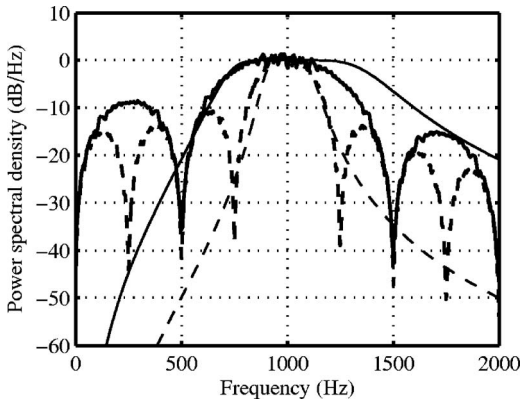


FIG. 2. Power spectral density of acoustic direct sequence spread spectrum signals implemented using 2 (bold solid) and 4 (bold dashed) cycles per chip. Compared with 1/1 (solid) and 1/3 (dashed) octave filter curves.

Although a DSSS signal can be passed through a bandlimiting filter with a minimal effect, important timing information is contained in the sidelobes. This is because the sidelobes contain much of the harmonic power of the modulation, and so limiting the bandwidth is the equivalent of restricting the rise and fall times of the modulating code. When the received signal is correlated against an ideal model of itself, the ratio of the value of correlation when the signals align versus the maximum value of the correlation for all other time lags is a measure of the clarity with which the signal can be measured in a noisy environment. Consequently, for extending application of the technique, the cross-correlation properties of the PN sequences are as important as the autocorrelation properties. Other methods for generating PN sequences with good periodic cross-correlation properties that are similar to *m*-sequences have been developed, for example by Gold and by Kasami.[14] These sequences have controlled cross-correlation properties and are simple to generate. An example of the use of these mutually orthogonal PN sequences as spreading codes would be multiple source applications, so that the level of interference from different sources is zero.

Compared with MLS this approach would appear to be at a disadvantage given the additional computational requirements for signal processing. However, there are also a number of advantages to this approach.

(1) Some immunity to time variance of the medium, from the correlation properties of the spreading code, and from recovery and tracking of the carrier, since by recovering and tracking the carrier some compensation for time variance can be applied.
(2) Accuracy of the times of flight allowed by the technique, from synchronization of the signal between source and receiver by correlation at the sample level
(3) Measurements can be performed with a large range ambiguity, using a long spreading code.
(4) Concentration of the energy into the relevant bandwidth.

The challenge for the study of outdoor noise propagation, however, lies in designing the spreading code to achieve both fine resolution for precise measurement of the time of flight, and sufficient length of code for a large range ambiguity. The design of one such spreading code, referred to here as the "inner and outer" (I&O) sequence, is detailed in the following.

C. Resolution, range ambiguity, and the I&O sequence

Ranging has been the most prominent and best-known use of spread spectrum systems. Any signal used will be subject to the same distance/time relations; the signal reaching a microphone at a given instant left the speaker some time before. The difference in a signal wave form measured at a microphone from that present at the speaker can be related directly to the distance between them and the time of flight. A spread spectrum signal has an advantage, however, in that its phase is easily resolvable. Resolution in ranging is therefore determined in accordance with the sample rate and

code rate used, while the sequence length determines maximum unambiguous range. The basic resolution is one code chip, and consequently the higher the chip rate the better the measurement capability. In practice, measurements are commonly made to within a fraction of a chip, with range determined to ~1/1000th of a chip period by high resolution spread spectrum radar systems.[7]

The ranging code used by a system must be long enough not to repeat over the maximum distance measured. Otherwise, the sequence will repeat and the receiver would recognize synchronization at more than one range. Much work has been performed in the area of developing special codes which have the required length for long-range measurement, but which also have synchronization properties that permit a range to be searched out without analyzing the entire code length. The JPL ranging codes have been particularly successful. These sequences are made up of three component codes assembled in such a way that the overall ranging sequence has a length that is the product of the three component code lengths.

Our approach to resolve the problem of range ambiguity measurement in outdoor sound propagation at long distances is here refered to as the I&O sequence. This composite code is generated by the combination of an inner linear maximal sequence in an outer linear maximal sequence pattern. This means that for a bipolar code in the outer sequence the code symbol 1 is taken as the inner sequence, while the code symbol −1 in the outer sequence is taken as an inverted version of the inner sequence.

For example, let the inner code be a and the outer code be b. If

$$a = (+1\ -1\ -1) \tag{2}$$

and

$$b = (+1\ +1\ -1\ +1\ -1\ -1\ -1) \tag{3}$$

then the I&O code c will be

$$c = (\bar{a}\ \bar{a}\ a\ \bar{a}\ a\ a\ a) \tag{4}$$

where $\bar{a}$ is NOT(a), so

$$c = (+1\ -1\ -1\ \ +1\ -1\ -1\ \ -1\ +1\ +1\ \ +1\ -1\ -1\ \ -1\ +1\ +1\ \ -1\ +1\ +1\ \ -1\ +1\ +1). \tag{5}$$

This can be implemented efficiently in software using the Kronecker tensor product of a and b,

$$a' \otimes b = \begin{bmatrix} +1 \\ -1 \\ -1 \end{bmatrix} \otimes [+1\ \ +1\ \ -1\ \ +1\ \ -1\ \ -1\ \ -1] = \begin{bmatrix} +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ -1 & -1 & +1 & -1 & +1 & +1 & +1 \\ -1 & -1 & +1 & -1 & +1 & +1 & +1 \end{bmatrix}. \tag{6}$$

Taking elements column-wise

TABLE I. Characteristics of an I&O signal utilizing 8 bit MLS for both inner and outer coding, a 1 kHz carrier, and two cycles per bit.

Characteristic	Calculation	Result
Basic resolution of one chip	$2\times\frac{1}{1000}$	2 ms
Resolution of time of flight (assuming synchronization $<\frac{1}{4}$ chip)	$\frac{1}{4}\times 2\times\frac{1}{1000}$	0.5 ms
Duration of one inner signal	$2\times\frac{1}{1000}\times(2^8-1)$	0.51 s
Duration of one outer signal	$0.51\times(2^8-1)$	130.05 s
Range ambiguity	340×130	44 km

$$c = [+1\ -1\ -1\ +1\ -1\ -1\ -1\ +1\ +1\ +1\ -1\ -1\ -1\ +1\ +1\ -1\ +1\ +1\ -1\ +1\ +1]. \tag{7}$$

The inner MLS code a does not determine the maximum range and therefore its length can be short, reducing synchronization time and increasing the basic resolution of one chip. The outer MLS code b, on the other hand, does not determine the basic resolution or the rate of synchronization, and all that is required is that it should be long enough to exceed the measurement range under investigation. Advantages of this I&O sequence include simple generation and a greatly increased range ambiguity, whilst retaining the excellent ranging properties of the MLS.

Table I considers an example with an 8 bit MLS used for both inner and outer coding, and modulation of a 1 kHz carrier using two cycles per bit. These characteristics are determined by the I&O code design and are carrier frequency dependent, with the signal duration and range ambiguity being inversely proportional to frequency. Since the resolution of the time of flight ultimately depends on the sample rate, this means that there can be a significant advantage to operating with a higher sample rate. The disadvantage is of course an increase in computational load, although online application using modern DSP processors might realistically be expected to operate with sample rates ~192 kHz. The repetition rate of ~0.5 s per inner sequence with a resolution of ~0.5 ms on the time of flight provides a realistic and useful measure of the variation in the acoustical propagation path. On the other hand, the range ambiguity determined by the outer sequence is sufficiently large for most conceivable applications.

III. EXPERIMENTAL INVESTIGATION

The field measurements involved synchronous recording at multiple positions of noise from an omnidirectional loudspeaker. Experiments were performed at the Shoeburyness firing range on the East Coast of England on the early afternoon of 16 May 2006. The correlation signals used implemented the I&O sequence modulated on octave band frequencies from 125 to 4 kHz. A signal composed of the combined 125–4 kHz octave band modulated sequences was also used. However, for brevity, results will only be presented here for the 1 kHz carrier investigation, although

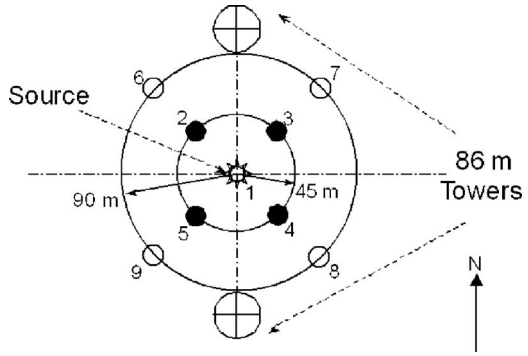


FIG. 3. Schematic illustrating plan view of experimental setup. Numbers identify microphone positions. (●) Microphone and sonic anemometer. (○) Microphone only.

similar results were obtained for the other carrier frequencies. These measurements were performed over a duration of 17 min 10 s starting at 1326UTC. Simultaneously, sonic anemometers logged variations in wind speed and direction. The weather was bright with no rain. The mean wind direction measured at the source during these measurements was 90° relative to North and the mean wind speed was 3.0 m/s. Microphones were positioned simultaneously in horizontal concentric circles, vertically on the North 86 m tower. The equipment configurations are illustrated in Figs. 3 and 4, and shown in photographs in Fig. 5.

Four microphones numbered 2–5 were positioned at a radial distance of 45 m, and four microphones numbered 6–9 at a radial distance of 90 m. Distances were determined using a tape measure and later measured using differential GPS. As a result of these investigations, it was discovered that the distance measurement should be as precise as possible. The orientation of the concentric circle of microphones was rotated by 45° from the North to minimize the flow effects of the towers. A single microphone, numbered 1, monitored the source at a radial distance of 1 m. The above-mentioned microphones were set at a nominal height of 1.5 m from the ground. The ground surface was a flat grassland fairway, with the elevation determined by a differential GPS varying from 1.33 m at microphone 1 (source) to −1.08 m at microphone 9 (SW90m). The grassland was moist and had the light compaction typical of a firing range. Short-range excess attenuation measurements were performed within the experimental circle, and the ground impedance calculated from an average obtained from measurements in four locations. Best-fitting Attenborough's 2PA model[15] to the mean ground impedance produced effective values of flow resistivity 75 kN s m^{-4} and rate to change of porosity with depth 0 m^{-1}.

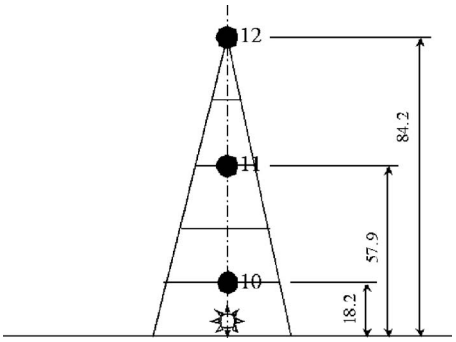


FIG. 4. Schematic illustrating a side view of the experimental setup on the tower. Numbers identify microphone positions. (●) Microphone and sonic anemometer. (ýý) Source.

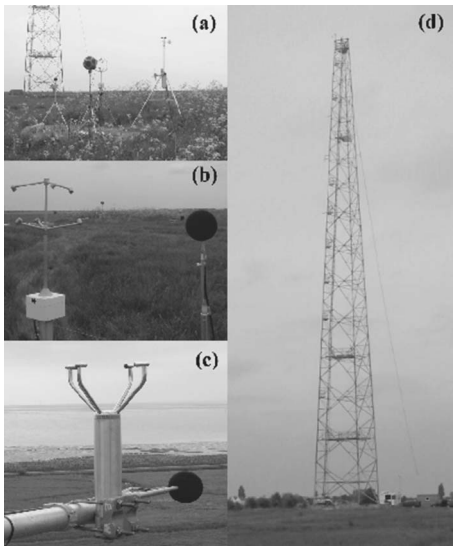


FIG. 5. Photographs showing the experimental setup. (a) Sound source, microphone monitoring the source at a distance of 1 m, and sonic anemometer. South tower visible behind. (b) Microphone and sonic anemometer at position 5, SW of the source at a distance of 45 m. View looking NE toward the source. (c) Detail of microphone and sonic anemometer on North tower at height 84.2 m. (d) View NW of North tower.

Three microphones numbered 10–12 were also positioned on the 86 m North tower at convenient heights of 18.2, 57.9 and 84.2 m, with additional microphones numbered 13–16 located at the base of the North tower. Three sonic anemometers were positioned close to microphones 10–12 on the tower. Field data from the sonic anemometers were also logged at the four 45 m microphone positions 5–9 and at 1 m radial distance from the source position 1. The nominal measurement height of the sonic anemometers was 1.6 m. Each sonic anemometer had a dedicated laptop for data logging, the clocks of which were synchronized by radio time signal receivers. Each laptop recorded raw data from the sonic anemometer at 10 Hz and turbulence parameters once every 10 min.

A dodecahedron speaker was located at a height of 2 m midway between the two 86 m towers in the center of the concentric circle of microphones. Correlation signal wav files were replayed at 12 kHz sample rate from a laptop computer. Data from all the microphones were logged at a sample rate of 50 kHz. Calibration tones of 1 kHz at 93.8 dB were recorded before the commencement of the correlation sequence play back. The logger provides synchronization across all channels within ~1 ns, with allowance made for cable lengths. Meteorological measurements were made using an AWS, the Salford LIDAR, the Salford radiometer, and radiosondes. Although not essential for the correlation experiment, these meteorological measurements contribute to a comprehensive data set to be used elsewhere.

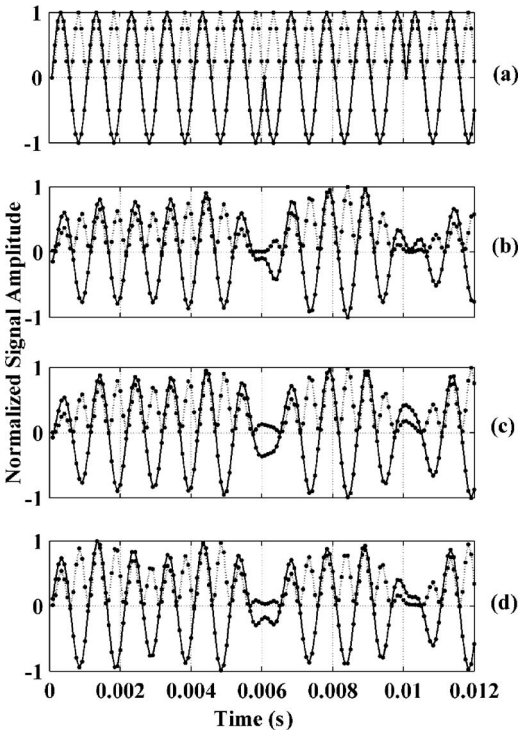


FIG. 6. Signals (solid) and squares of signals (dashed) for first six chips. (a) The signal sent. (b), (c), and (d) Three different receiver distances. (a) Original signal. (b) Signal at 1m, (microphone 1). (c) Signal at SW45m (microphone 5). (d) Signal at SW90m (microphone 9).

IV. RESULTS

A. Carrier recovery

The carrier is recovered by squaring the direct sequence biphase signal. The result is a continuous wave signal at twice the original modulated carrier frequency. It may be recalled from Fig. 2, however, that there is no modulated signal at twice the carrier frequency. Whilst in the context of rf spread spectrum communications this fact reduces the covert applications of direct sequence signaling, for acoustic remote sensing this makes direct sequence biphase signals the preferred option since it allows for direct detection and tracking of the signal carrier.

Figure 6 shows the measured signals at 1 m from the source (microphone 1), SW45m (microphone 5), and SW90m (microphone 9) receivers for the first six chips of the first inner sequence. The positions at SW45m (microphone 5) and at SW90m (microphone 9) are downwind of the source. The signals have been passed through a high pass filter and aligned for display purposes. Also shown are the squares of the aligned filtered signals, and the recovered carrier at twice the original carrier frequency is clearly evident. Figure 7 shows the carrier frequency calculated using the squaring technique for the first 255 inner sequences at the same receivers. The carrier frequency is calculated for each inner sequence using a correlation-based comparison of sinusoids. The correlation-based comparison of sinusoids is a standard technique for the estimation of recovered carrier frequency and is related to the sliding correlator.[6] The carrier estimation procedure does not explicitly require precise signal timing. This technique does not exploit the correlation properties of the spreading sequence.

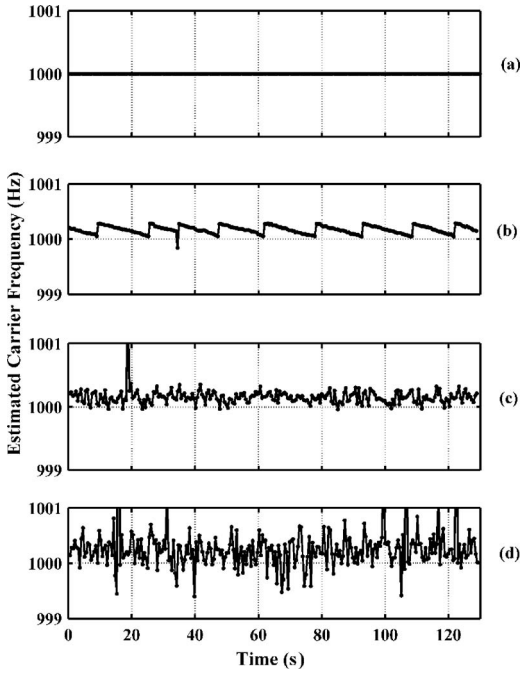


FIG. 7. Carrier frequencies estimated from squares of signals for the first 255 inner sequences. (a) The signal sent. (b), (c), and (d) Three different receivers. (a) Original signal. (b) Signal at 1m, (microphone 1). (c) Signal at SW45m (microphone 5). (d) Signal at SW90m (microphone 9).

These investigations show that the correlation-based comparison of sinusoids can provide accurate estimates of carrier frequency of the acoustic spread spectrum signal following propagation through a real time-varying outdoor acoustic medium. On this occasion, the precision of the car-

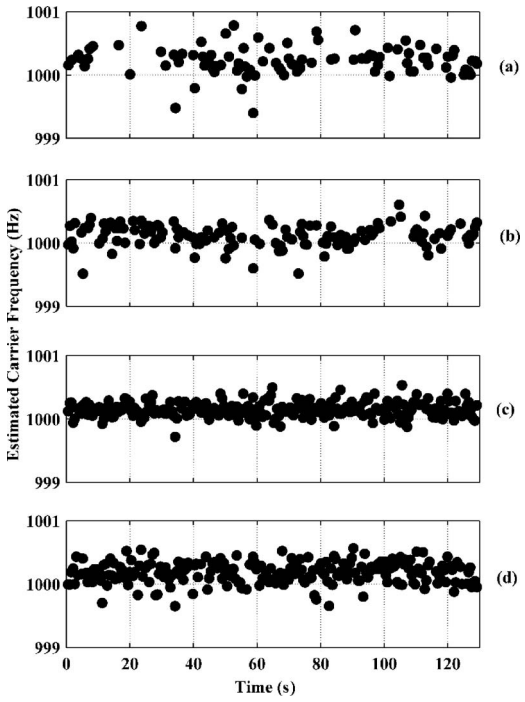


FIG. 8. Carrier frequencies estimated from squares of signals for microphones on the North tower for the first 255 inner sequences. Plot shows variations within 1 Hz of the nominal 1 kHz carrier frequency for the signal at four receiver heights. (a) Signal at 84.2m, (microphone 12). (b) 57.9m, (microphone 11). (c) 18.2m, (microphone 10). (d) 1.5m, (microphone 16).

TABLE II. Summary of numbers of significant Doppler measured by ground-based microphones at 1.5 m.

	Original signal	1 m microphone 1	SW45m microphone 5	SW90m microphone 9
Std (Hz)	0.00	0.07	0.12	0.74
Number of shifts in frequency >10 Hz	0	0	0	1

rier frequency determined by the correlation-based comparison of sinusoids was within ±0.01 Hz, although there is no reason a finer match could not be achieved. However, it is evident from the results of Fig. 7 that the variations in carrier frequency are significantly greater than ±0.01 Hz. Even at 1 m, the standard deviation is 0.07 Hz, although the sawtooth form of the estimated carrier frequency suggests that this may be an artifact of the playback sound card rather than an atmospheric phenomenon. The mean estimated carrier at 1 m of 1000.17±0.01 Hz indicates that the playback from a laptop was at a slightly higher rate than 12 kHz, and the sawtooth variation in the carrier estimate may be due to a systematic variation in sample rate. Alternatively, possible though less likely is a systematic error in the sampling rate on the Racal-Heim 380 recording unit. The latter is, however, a high-specification dedicated synchronous sampling multichannel data recorder. Playback, on the other hand, was through a low-cost sound card, and the capability of the carrier estimation routine to track this variation is reassuring. Subsequently, these playback variations can be calculated and accommodated by the carrier estimation method.

Figure 8 shows the calculated frequencies for each inner sequence over the first outer sequence for microphones 10–12 and 16 on the North tower. It is noteworthy that these receivers are crosswind of the source, and this is reflected in the variability of the carrier on the ground, i.e., at 1.5 m height, as compared with those downwind, i.e., at SW45m (microphone 5), and at SW90m (microphone 9). The standard deviations in the recovered carrier frequency are compared in Tables II and III.

Causes of shifts in carrier frequency include Doppler, refraction, and turbulence. Doppler shifts result from the effects of the moving medium. Refraction results from variability in the speed of sound in both the horizontal and vertical directions, and causes variations in the path length that lead indirectly to shifts in the frequency of the received signal. Turbulent fluctuations of wind vector, temperature, and humidity cause the scattering of sound, but combined with small-scale refractive effects again lead indirectly to shifts in the frequency of the received signal.

In Table II, an increasing standard deviation of estimated carrier frequency is seen with distance for the microphones at 1.5 m. The variations seen at 1 m are hardly evident within the SW45m (microphone 5) estimates, while the SW90m (microphone 9) estimates show much larger variations again. At SW90m (microphone 9) the shift in carrier frequency is seen to be within ~1 Hz for 10/255, i.e., ~4% of the time. This compares with 1/255 at SW45m (microphone 5), a tenfold increase with a doubling of the distance. The relative velocity of the source and receivers as calculated from these estimates of carrier frequency reflect the above-described properties. The relative velocities are seen to be small, <2 m/s in most cases at SW90m (microphone 9). A useful representation might be to consider these variations as changes in the main propagation path length. However, as discussed earlier, changes in propagation path length depend not only on refraction but also on the variations in wind speed.

Comparing the estimations of carrier frequency crosswind on the tower in Table III with those obtained downwind at 1.5 m from the ground, a larger number of significant shifts in carrier frequency are observed. The vast majority of shift in carrier frequency are less than 1 Hz, and these were displayed in Fig. 8. Considering a shift in carrier frequency of 10 Hz, however, this corresponds to relative velocity of 3.4 m/s. The number of shifts of the order of 10 Hz increases rapidly for microphones higher up the tower, and away from the ground between microphone 10 at 18.2 m and microphone 12 at 84.2 m, the variation of shifts in carrier frequency is seen to increase significantly with height. However, while no such significant shifts in carrier frequency are seen from microphone 10 at 18.2 m, some are seen from microphone 16 at the height 1.5 m. One suggestion is that this may be due to turbulence generated by wind shear near the ground resulting in large variations in sound speed gradients along the path from source to receiver. The larger and more variable shifts in carrier frequency for microphones at greater heights are at least in part due to the increased path length, but also perhaps due to increased wind speeds with height. The variation of shift in carrier frequency close to the

TABLE III. Summary of numbers of significant Doppler measured by tower microphones.

	Mast 1.5 m Microphone 16	Mast 18.2 m Microphone 10	Mast 57.9 m Microphone 11	Mast 84.2 m Microphone 12
Std (Hz)	2.63	0.41	5.24	6.51
Number of shifts in frequency >10 Hz	13	0	22	87

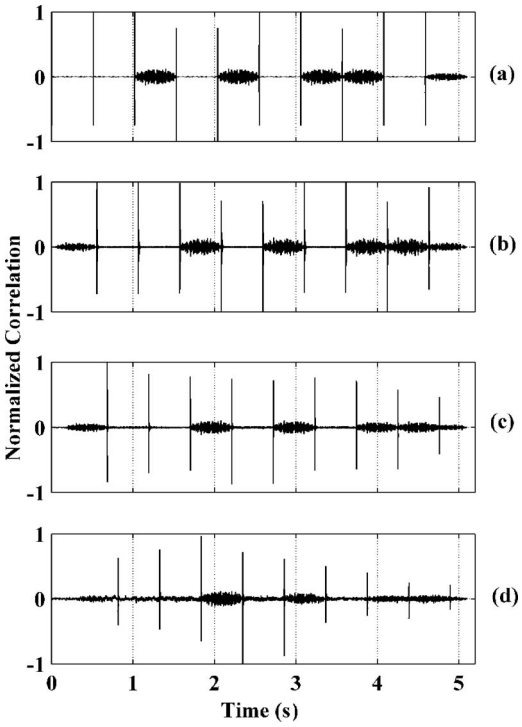


FIG. 9. Normalized cross correlations of modulated inner sequence with original and recorded signals for the first 10 inner sequences. (a) Original signal. (b) Signal at 1m, (microphone 1). (c) Signal at SW45m (microphone 5). (d) Signal at SW90m (microphone 9).

ground is significantly greater crosswind than downwind. This variation of shift in carrier frequency close to the ground is likely to be due to variations in the wind and temperature gradients in the surface layer.

B. Correlation properties and time of flight

Correlation functions were produced by the cross correlation of the modulated inner signal with the received signal of outer length duration. Figure 9 shows the correlation functions for the received signals of microphone 1 at the source, microphone 5 at SW45m, and microphone 9 at SW90m. Also shown is the cross correlation of the modulated inner signal with the playback signal. It is evident from the correlation functions that the amplitude of the peaks varies with range and with time indicating that transfer function information is present.

Precise location of the discrete correlation peaks allows an accurate calculation of time of flight of the sound between microphones. Since the correlations are performed in the passband, that is on the received rather than demodulated acoustic spread spectrum signal, the time of flight can be resolved to the order of the sample rate. Furthermore, since the correlation peaks are determined for each inner code signal, a mean time of flight is calculated with periodicity ~0.5 s. It is convenient to display the time of flight normalized by the geometrical separation of the microphones as speed of sound. Figure 10(a) shows the mean speed of sound per frame of inner code signal between the source (microphone 1) and SW45m (microphone 5), and SW90m (microphone 9). Also shown is the speed of sound as measured by the sonic anemometer positioned at SW45m.

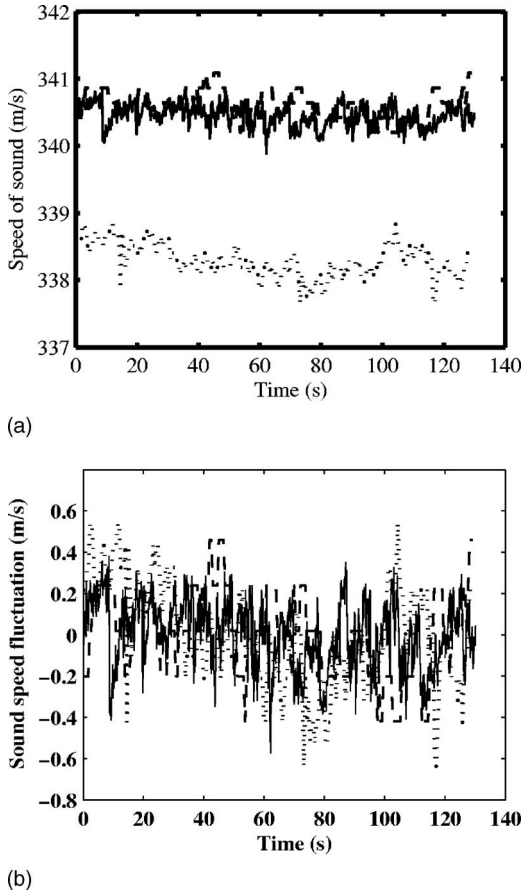


FIG. 10. (a) Speed of sound as measured by sonic anemometer, and as calculated from time of flight to each microphone for each inner sequence component over one outer sequence. (Dashed) Sonic anemometer at SW45m. (Solid) Microphone 5 at SW45m. (Dotted) Microphone 9 at SW90m. (b) Sound speed fluctuations as measured by sonic anemometer, and as calculated from time of flight to each microphone for each inner sequence component over one outer sequence. (Dashed) Sonic anemometer at SW45m. (Solid) Microphone 5 at SW45m. (Dotted) Microphone 9 at SW90m.

The significance of accurate geometries is seen in the displacement of the speed of sound data. For the 90 m range microphone, a ±0.25 m uncertainty in range corresponds to a ±1 m/s uncertainty in the speed of sound data. Consequently, while it is encouraging that the speed of sound measured by the sonic anemometer at the SW45m position shows excellent agreement with the mean speed of sound calculated between the source (microphone 1) and the SW45m receiver (microphone 5), the uncertainties in geometry would suggest a ±1 m/s displacement of the latter could result. Furthermore, the sonic anemometer provides a point measurement averaged over an air volume of about 150 × 150 × 150 mm, contrasting with the linear spatial average from the correlation measurement. Nevertheless, the estimates of speed of sound for the correlation measurements and the sonic anemometer measurements show excellent agreement within the limits of uncertainty.

Figure 10(b) shows these data presented as sound speed fluctuations rather than absolute sound speed values, in order to be independent of uncertainties in distance measurements. The comparison of sound speed fluctuations between source and receivers with those measured at the sonic anemometer

show good overall agreement. The standard deviation of sound speed fluctuations for the sonic anemometer was 0.16 m/s, while between the source (microphone 1) and SW45m (microphone 5) was 0.21 m/s, and between the source (microphone 1) and SW90m (microphone 9) the standard deviation was 0.25 m/s. This result confirms that the sound speed fluctuated more over the larger propagation path.

V. DISCUSSION AND APPLICATIONS

The above-mentioned results indicate that a directly carrier-modulated code sequence modulation system can provide significantly improved ways of measuring sound propagation outdoors. Nevertheless, in the field the technique will ultimately be limited by signal to noise ratio and a shift in frequency, and clearly further work is required to investigate these bounds. These factors are likely to be of most significance to applications in outdoor noise propagation. On the other hand, the results also show that the uncertainty in the measurement of the speed sound is determined by the precision of the geometrical determination of receiver positions. This factor is likely to be of most significance to applications in acoustic tomography.

One objective of this research is to provide researchers with an improved method for investigating the propagation of noise outdoors. The problem identified is that of obtaining both adequate signal-to-noise ratio and phase information in medium to long-range measurements. This is a major limitation in the study of meteorological and topographical effects on sound propagation. However, the sound power level of the source is also often required to be limited due to the need to use levels that do not cause a noise disturbance during testing. These initial investigations, with a maximum range of ~125m to the microphone at the top of the tower, indicate that the technique might therefore be useful for the investigation of environmental noise propagation, for example in urban conditions. However, whilst these investigations are encouraging for measurements at greater distances, for the study of outdoor noise propagation, especially the effects of meteorology, evidence of the applicability of the technique for long-range measurements would be required.

An interesting development of the technique would be extraction of the impulse response between two receivers, with the intention of deriving both phase and amplitude of the frequency response function. The sound attenuation during propagation as a function of frequency could then be found by filtering and integrating over the corresponding time interval. In addition to the time of travel of the sound between the source and receiver and the attenuation of the channel, the strength of disturbing noise during the period of signal transmission might also then be estimated, allowing a simultaneous measurement of the background noise.

Application in urban environments is in part suggested because of the broadband, low intrusive nature of the tailored sound source. On the other hand, spread spectrum systems are well known for operation in environments with high background noise levels and low signal-to-noise ratios. As mentioned, this property is likely to allow operation in the urban environment without disturbance to residents. However, the capacity to operate over low signal-to-noise propagation paths makes the technique useful for the investigation of other acoustical problems. Combined with the time-varying capabilities of the technique, the acoustic spread spectrum signal would seem well suited for the investigation of the properties of barriers in real-world turbulence conditions. Similarly, propagation experiments could also be performed in uneven and hilly terrain where accurate measurements are currently difficult to obtain by traditional methods. Propagation could also be determined in the presence of high background noise such as from motorways.

Meteorological classifications for the prediction of long-term average sound levels in Europe were recently presented by Heimann.[16] However, within the urban environment sound speed profiles vary significantly in the horizontal direction, and vertical gradients of temperature and wind speed are significantly stronger at the very small heights above the ground. Acoustic time travel measurements are valuable in the determination of horizontal gradients and the investigation of turbulence structures in the atmospheric surface layer,[17] and so the continuous fine resolution for time of flight in potentially orthogonal directions obtained by the spread spectrum technique suggests an application in acoustic tomography.

The acoustic spread spectrum technique described provides a direct measurement of the sound speed, rather than meteorology from which the required profile must be inferred. This presents the possibility of the direct determination of sound speed profiles within the urban environment or a forest canopy, both environments where it has previously been difficult to obtain reliable measurements. Realistic sound speed profiles are essential for accurate predictions of outdoor sound propagation. Examples include noise mapping exercises, detection and ranging, and evaluation of the environmental noise impact from a proposed development. Measurements of sound speed profiles using the spread spectrum signal approach could take the form of a suspended string of inexpensive microphones at precisely determined separations, to receive the coded signal and thus provide accurate and precise sound speed profiles. Application of the spread spectrum method could allow tailoring of the signal for the environment. For example, in the urban environment a signal source resembling traffic noise would allow daytime experiments to take place with minimum disturbance to the population, even in the presence of significant background sources. On the other hand, combinations of orthogonal coding and multiple carrier frequencies would permit continuous sound speed measurements in numerous vertical and horizontal directions, an approach comparable with acoustic tomographic applications.[18]

VI. CONCLUSIONS

An acoustic spread spectrum technique for the study of outdoor noise propagation has been described. The correlation signal was produced by modulation of an acoustic carrier signal by a code sequence. Optimization of the correlation signal through careful design of the code sequence was

explained. A limited case study was presented introducing the approach taken and some of the properties of the technique for the investigation of propagation in a real turbulent atmosphere. Careful design of the spread spectrum signal allowed recovery and tracking of the carrier frequency, and time variance of the medium was measured by examining variation in the Doppler of the carrier. These results indicate that the time variance encountered in outdoor sound propagation measurements can be determined and compensated by this method.

In a separate analysis, the correlation properties of the acoustic spread spectrum signal were demonstrated in the analysis of recorded signals, and were used to calculate accurate times of flight. The times of flight were used to estimate mean speed of sound over the propagation distance, and these estimates compared well with sonic anemometer point measurements of speed of sound. The results indicate that the acoustic spread spectrum technique is likely to prove valuable in time-varying propagation environments, under potentially low signal-to-noise ratio conditions. Moreover, the technique provides a direct measurement of the sound speed, as opposed to the meteorology from which the required profile must be inferred. This presents the possibility of the direct determination of detailed sound speed profiles, for example within the urban environment or a forest canopy, an environment where it has previously been difficult to obtain reliable measurements.

ACKNOWLEDGMENTS

We would like to thank Dr. Rodger Munt and Dr. Ray Browne from QinetiQ (Farnborough) Ltd. for their support with the field trial. The EPSRC (UK) funded this work with Grant No. GR/S80431/01. The authors would like to thank the referees for their helpful comments.

[1]D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–443 (1989).

[2]N. Xiang and M. R. Schroeder, "Reciprocal maximum-length sequence pairs for acoustical dual source measurements," J. Acoust. Soc. Am. **113**, 2754–2761 (2003).

[3]B. Hallberg, C. Larsson, and S. Isrealsson, "Measurements of meteorological effects on long-range sound propagation using m-sequence correlation," J. Acoust. Soc. Am. **78**, 1038–1044 (1985).

[4]K. Heutschi and A. Rosenheck, "Outdoor sound propagation measurements using an MLS technique," Appl. Acoust. **51**, 13–32 (1997).

[5]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (Spon, London, 1997).

[6]J. G. Praokis, *Digital Communications*, 4th ed. (McGraw-Hill, New York, 2001).

[7]R. C. Dixon, *Spread Spectrum Systems With Commercial Applications*, 3rd ed. (Wiley, New York, 1994).

[8]S. W. Golomb *et al.*, *Digital Communication with Space Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1964).

[9]D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).

[10]H. Zheng, A. Kaneko, and N. Gohda, "A multistation coastal acoustic tomography system with coherent operation," J. Acoust. Soc. Am. **120**, 3063 (2006).

[11]D. C. Swanson, T. D. Mast, M. P. Mahon, and D. Norris, "Atmospheric multipath resolution using spread spectrum acoustic signals," J. Acoust. Soc. Am. **101**, 3102 (1997).

[12]M. P. Mahon, "A wide band characterization of the acoustic spreading function in the surface layer of the atmosphere," Ph.D. thesis, Pennsylvania State University, 2001.

[13]K. Mizutani, I. Odanaka, and N. Wakatsuki, "Differential biphase shift keying (DBPSK) using airborne sound for data communication," presented at 13th International Congress on Sound and Vibration (ICSV13), Vienna, Austria, 2006.

[14]P. Fan and M. Darnell, *Sequence Design for Communications Applications* (Wiley, New York, 1996).

[15]K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).

[16]D. Heimann and E. M. Salomons, "Testing meteorological classifications for the prediction of long-term average sound levels," Appl. Acoust. **65**, 925–950 (2004).

[17]D. K. Wilson, A. Ziemann, V. E. Ostashev, and A. G. Voronovich, "An overview of acoustic travel-time tomography in the atmosphere and its potential applications," Acust. Acta Acust. **87**, 721–730 (2001).

[18]S. N. Vecherin, V. E. Ostashev, G. H. Goedecke, D. K. Wilson, and A. G. Voronovich, "Time-dependent stochastic inversion in acoustic travel-time tomography of the atmosphere," J. Acoust. Soc. Am. **119**, 2579–2588 (2006).

Array gain for a cylindrical array with baffle scatter effects

Derek C. Bertilone,[a] Damien S. Killeen,[b] and Chaoying Bao[c]
Defence Science and Technology Organisation, Building A51, HMAS Stirling, P.O. Box 2188, Rockingham DC, Western Australia 6958 Australia

(Received 22 February 2007; revised 16 August 2007; accepted 22 August 2007)

Cylindrical arrays used in sonar for passive underwater surveillance often have sensors surrounding a cylindrical metal baffle. In some operational sonars, the phones in each stave (i.e., each line of phones aligned with the cylinder axis) are hardwired together so that the array is equivalent to a baffled circular array of directional elements, where each element corresponds to a line array of omnidirectional phones steered to broadside. In this paper a model is introduced for computing the array gain of such an array at high frequencies, which incorporates baffle scatter using infinite, rigid cylinder scattering theory, and with ambient noise described by an angular spectral density function. In practice the phones are often offset from the baffle surface, and the acoustic field sampled by the staves is distorted at high frequencies due to interference between the incident and scattered fields. Examples are given to illustrate the resulting array gain degradation, using three noise distributions that are frequently used in sonar performance modeling: three-dimensional isotropic, two-dimensional isotropic, and surface dipole noise. [DOI: 10.1121/1.2785033]

PACS number(s): 43.60.Fg [EJS] Pages: 2679–2685

I. INTRODUCTION

Circular and cylindrical arrays are widely used in radar, sonar, and communications.[1–3] The cylindrical receiving arrays used in sonar for passive underwater surveillance are often comprised of several rings of hydrophones surrounding a cylindrical metal baffle. The phones in different rings are vertically aligned (i.e., parallel to the cylinder axis) in groupings known as "staves." For some operational sonars, the outputs of the phones in each stave are not available for individual processing. Instead, the phones in each stave are hardwired so that each stave has a single output that corresponds to summing the associated phone outputs with zero time delay. This is done to reduce the computational load, and in this case the array is equivalent to a baffled circular array of directional elements, where each element corresponds to a vertical line array steered to broadside. In typical detection applications, the acoustic signals from distant vessels arrive at almost zero elevation, i.e., perpendicular to the cylinder axis, while ambient noise can arrive from any elevation. Thus the stave hardwiring has the effect of suppressing ambient noise components that arrive from nonzero elevations, without suppressing the signals. Beamforming is carried out by processing the outputs of the staves that lie on an arc centered on the steering direction.

The purpose of this paper is to introduce a model for computing the array gain for this type of array, which incorporates scattering of both the signal and ambient noise fields by the baffle. Array gain quantifies the improvement in signal-to-noise ratio (SNR) obtained by beamforming the array, compared to a single omnidirectional phone, and is a widely used performance metric.[1,4] The baffle is modeled as an infinite, rigid cylinder,[5,6] the incident signal is taken to be a plane wave, and the ambient noise field is modeled as a superposition of independent plane waves arriving from all directions in space, with a power distribution described by an angular spectral density function.[3] Although it is reasonable to apply this model to a metal baffle at high frequencies, it is inapplicable at very low frequencies or near mechanical resonances, and it also neglects effects associated with the finite length of the baffle.[5,6] To illustrate the application of the model, we examine the impact of baffle scatter on array gain at high frequencies, when the staves are offset from the baffle surface. A stave offset is often introduced in practice for the purpose of improving phone coupling to the water, and to reduce the coupling of baffle vibrations to the phones. When an offset is present, interference between the incident acoustic wave and the wave field scattered by the baffle can distort the signal at the spatial locations occupied by the phones, and this can degrade high frequency performance. Illustrative examples are given for three idealized noise fields that are often used in sonar performance modeling: three-dimensional (3D) isotropic, two-dimensional (2D) isotropic, and surface dipole.[3,4] Also, two beamformers are considered: The optimum spatial filter that maximizes SNR for the given acoustic field,[1,3] and a spatial filter that is proportional to the array output vector when signal-only is present. The latter maximizes SNR when the noise at the output of different staves is independent and identically distributed (IID).

The work presented here builds on earlier research by Meyer,[7] who incorporated scattering theory into the analysis of beamforming of a circular array of omnidirectional and dipole sensors mounted on a rigid sphere. He analyzed the directivity index and white noise gain, and also discussed some effects that arise when the sensors are offset from the surface of the sphere. The directivity index corresponds to array gain when the noise field is nondirectional (i.e., 3D isotropic), while the white noise gain assumes that the noise

[a]Electronic mail: derek.bertilone@dsto.defence.gov.au
[b]Electronic mail: damien.killeen@dsto.defence.gov.au
[c]Electronic mail: chaoying.bao@dsto.defence.gov.au

at the output of different sensors is IID. The current paper extends this work by showing how scattering theory can be included in array gain computations for a type of cylindrical array that is often used operationally, using directional noise fields that are often used in sonar performance modeling. In related work, Daigle *et al.*[8] have compared directivity indices for a circular array of omnidirectional phones mounted on the surface of spheres and infinite cylinders that are rigid or impedance scatterers, and discuss how the characteristics of the baffle can enhance array performance. Other studies of cylindrical and circular array directivity do not include a baffle, but include effects associated with sensor directionality; see, for example, Queen[9] and Rahim and Davies.[10] Recently, Teutsch and Kellermann[11] have utilized scattering theory to develop algorithms for wideband detection and localization using circular arrays of omnidirectional sensors mounted on the surface of rigid cylinders.

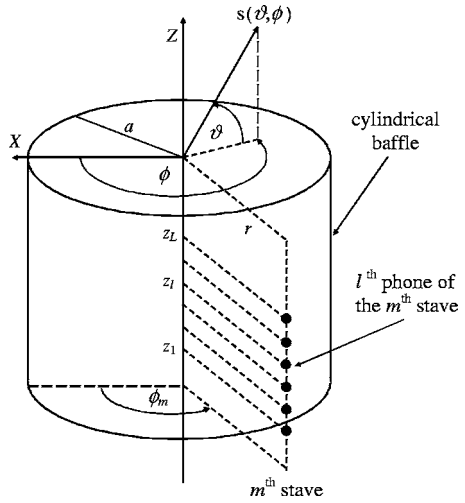


FIG. 1. Schematic illustration of a single baffled stave, showing the coordinate system and a unit direction vector $\mathbf{s}(\vartheta,\phi)$.

II. GEOMETRY OF THE ARRAY

Let Q sensors be equispaced on a circle of radius r, surrounding a cylindrical baffle of radius $a \leqslant r$. If the sensors are offset from the baffle surface (i.e., $a<r$), then the supports are assumed to be acoustically transparent. Each sensor is a stave of omnidirectional phones, i.e., an identical linear array of L phones, aligned with the cylinder axis. The phones in each stave are hardwired, so that each stave has a single output that corresponds to the sum of the associated phone outputs with zero time delay. A beamformer processes $M \leqslant Q$ staves that lie on an arc centered on the beamsteer direction. Thus as the array is steered through 360° to provide full coverage of azimuths at zero elevation, the arc swings around with it and selects different staves for processing. The full circle of Q staves is often not used for beamforming, because staves that face in the opposite direction to beamsteer are acoustically shielded by the baffle and make little contribution to detection and localization by the array at high frequencies.

Let $m=1,\ldots,M$ label the staves that lie on the processing arc, and let $l=1,\ldots,L$ label the phones in each stave. Figure 1 is a schematic illustration of a single baffled stave, and shows the cylindrical coordinates used to locate the phones; the lth phone of the mth stave is located at (r,ϕ_m,z_l), where ϕ_m is the azimuth angle of the stave, and z_l is the axial location of the phone. Often the phones can be taken to be equally spaced on the stave, $z_l=(l-1)\Delta z$, but this need not be the case. Also shown in Fig. 1 is unit vector $\mathbf{s}(\vartheta,\phi)$, with elevation angle ϑ and azimuth angle ϕ, that we use to indicate the direction of an incident plane wave. While acoustic signals are assumed to arrive from zero elevation, ambient acoustic noise components can arrive from any direction in space.

III. PLANE-WAVE RESPONSE OF A BAFFLED STAVE

A. Formulation

Suppose an acoustic plane wave arrives from direction $\mathbf{s}(\vartheta,\phi)$. The complex narrow-band output of the lth phone of the mth stave in bandwidth Δf at frequency f, as might be obtained by Fourier transforming a segment of data of duration $T_0 \sim 1/\Delta f$, is given by

$$x_{m,l}(f,\vartheta,\phi) = p(f)\{e^{-jk[z_l \sin\vartheta + r\cos\vartheta\cos(\phi-\phi_m)]} + \xi^{\text{scatt}}_{m,l}(f,\vartheta,\phi)\}. \quad (1)$$

Here it is assumed that Δf is sufficiently small that the narrow-band approximation[1] is applicable. The first term in curly braces gives the incident plane-wave contribution, the second term gives the baffle scatter contribution, $k=2\pi f/c$ is the wave number (c is the speed of sound in water), and $p(f)$ is a complex amplitude. For convenience, we assume that the phone outputs have been calibrated so that $\langle|p(f)|^2\rangle$ is the power spectral density of the incident acoustic wave, where $\langle\cdot\rangle$ denotes statistical expectation. The output of the stave is obtained by summing the outputs of all phones that comprise the stave,

$$x_m(f,\vartheta,\phi) = p(f)\sum_{l=1}^{L}\{e^{-jk[z_l \sin\vartheta + r\cos\vartheta\cos(\phi-\phi_m)]} + \xi^{\text{scatt}}_{m,l}(f,\vartheta,\phi)\}. \quad (2)$$

In this paper, the baffle is modeled as a rigid cylinder of infinite length. Then the acoustic field scattered by the baffle can be represented by a cylindrical harmonic expansion as follows:[5]

$$\xi^{\text{scatt}}_{m,l}(f,\vartheta,\phi) = e^{-jkz_l \sin\vartheta}\gamma_m(f,\vartheta,\phi), \quad (3)$$

$$\gamma_m(f,\vartheta,\phi) = \sum_{n=0}^{\infty}\varepsilon_n(-j)^n b_n(f,\vartheta)H_n^{(1)}(kr\cos\vartheta)\times\cos[n(\phi-\phi_m)], \quad (4)$$

$$\varepsilon_0 = 1, \quad \varepsilon_n = 2, \quad n>0, \quad (5)$$

$$b_n(f,\vartheta) = -\frac{J_n'(ka\cos\vartheta)}{H_n^{(1)\prime}(ka\cos\vartheta)}. \tag{6}$$

Here b_n is the scattering coefficient, J_n and $H_n^{(1)}$ are Bessel and Hankel functions, respectively, of the first kind and order n,[12] and primes denote derivatives. Substituting Eq. (3) into Eq. (2), we obtain the following representation of the output of the mth stave:

$$x_m(f,\vartheta,\phi) = p(f)h(f,\vartheta)\psi_m(f,\vartheta,\phi), \tag{7}$$

$$h(f,\vartheta) = \sum_{l=1}^{L} e^{-jkz_l \sin\vartheta}, \tag{8}$$

$$\psi_m(f,\vartheta,\phi) = e^{-jkr\cos\vartheta\cos(\phi-\phi_m)} + \gamma_m(f,\vartheta,\phi). \tag{9}$$

The outputs of the M staves that lie on the processing arc are placed in an array output vector,

$$\mathbf{x}(f,\vartheta,\phi) = [x_1(f,\vartheta,\phi), x_2(f,\vartheta,\phi), \ldots, x_M(f,\vartheta,\phi)]^T, \tag{10}$$

where the superscript T is matrix transpose. Using Eq. (7) we can represent the array output vector as follows:

$$\mathbf{x}(f,\vartheta,\phi) = p(f)h(f,\vartheta)\boldsymbol{\psi}(f,\vartheta,\phi), \tag{11}$$

$$\boldsymbol{\psi}(f,\vartheta,\phi) = [\psi_1(f,\vartheta,\phi), \psi_2(f,\vartheta,\phi), \ldots, \psi_M(f,\vartheta,\phi)]^T. \tag{12}$$

This result for the response of the array to a plane wave lies at the foundation of our model.

The directional response of a baffled stave is given by the product of two functions [see Eq. (7)]; $\psi_m(f,\vartheta,\phi)$ gives the contribution due to the baffle, and $h(f,\vartheta)$ gives the contribution due to the unbaffled stave. The former accounts for effects associated with the interaction of the acoustic field with the baffle, which includes acoustic shadowing by the baffle, while the latter corresponds to the directional response of a linear array steered to broadside in an unbounded medium.[1,3] The azimuthal response of a baffled stave is fully determined by the baffle contribution, while the elevational response depends on both contributions. The stave contribution to the elevational response plays a key role in the suppression of ambient noise components that arrive from nonzero elevations.

B. Distortions due to stave offset

Suppose an acoustic plane wave is incident on the array from zero elevation and zero azimuth. Infinite, rigid cylinder scattering theory shows that if the staves are flush against the surface of the baffle, then those staves that do not lie in the acoustic shadow (i.e., $|\phi_m| < 90°$) will sample a wave field with almost the same phase as the incident wave, and an amplitude that is close to double that of the incident wave (although dropping away as the shadow boundary is approached). If the staves are offset from the baffle surface, however, then the staves will sample a wave field that is distorted due to interference between incident and scattered fields. Figure 2 illustrates this point, by showing the amplitude and phase of $\Gamma(f,\phi_m) = \psi_m(f,0,0)/\exp(-jkr\cos\phi_m)$ plotted as sensor angle ϕ_m varies from −120° to +120°, for normalized frequencies $kr = 20, 40, 60$. Here Γ corresponds to the ratio of the narrow-band output of the mth stave of the baffled array, to the output that the stave would have if it sampled the incident acoustic field only (i.e., with no scattered contribution). The dash-dot curves show results for staves with zero offset; $r/a = 1$. In this case, if the stave is not shadowed by the baffle, then the phase of Γ is almost flat as a function of sensor angle, while the amplitude is flat near the center but starts to drop away as the shadow boundary is approached. The solid curves show the corresponding results for a nonzero stave offset with $r/a = 1.04$. In this case both the amplitude and phase of the stave output are increasingly distorted as frequency increases. For example, the amplitude drops almost to zero at $\phi_m = 0°$ when $kr = 40$; at this combination of stave location and frequency, the path difference between incident and reflected acoustic rays is almost half a wavelength.

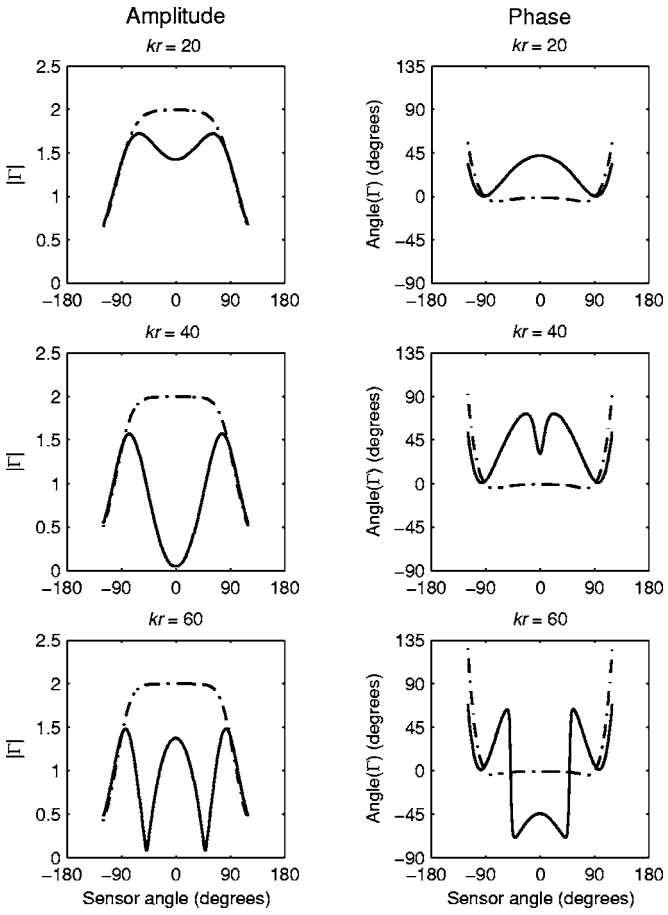


FIG. 2. Illustration of distortions in the stave output, when the stave is offset from the baffle surface. Shown are the amplitude (left) and phase (right) of Γ, the ratio of the narrow-band output of the baffled stave to the output that the stave would have if it sampled the incident acoustic field only, when a plane wave is incident from the zero-elevation, zero-azimuth direction. Results are plotted as a function of stave angle, for normalized frequencies $kr = 20$ (top), 40 (middle), and 60 (bottom). Solid curves: Nonzero stave offset with $r/a = 1.04$. Dash-dot curves: Zero stave offset.

IV. CROSS-SPECTRAL MATRIX

A. Single plane wave

The cross-spectral matrix[1] gives the spatial correlation between the narrow-band outputs of all pairs of staves on the processing arc. If we assume that the array is excited by a single plane wave arriving from a direction $\mathbf{s}(\vartheta,\phi)$, so that the array output vector is given by $\mathbf{x}(f,\vartheta,\phi)$, then the $M \times M$ cross-spectral matrix is given as follows:

$$\mathbf{R} = \langle \mathbf{x}(f,\vartheta,\phi)\mathbf{x}^H(f,\vartheta,\phi)\rangle, \tag{13}$$

where the superscript H denotes matrix conjugate transpose. Substituting Eq. (11) into Eq. (13) we obtain

$$\mathbf{R} = \langle |p(f)|^2\rangle |h(f,\vartheta)|^2 \boldsymbol{\psi}(f,\vartheta,\phi)\boldsymbol{\psi}^H(f,\vartheta,\phi). \tag{14}$$

For the array gain computations, we assume that the array is placed in an acoustic field comprised of a plane-wave signal together with ambient noise. The signal-only and noise-only cross-spectral matrices are derived from Eq. (14), as outlined in the following.

B. Signal-only matrix

The incident acoustic signal is assumed to be a plane wave at zero elevation. For convenience, we also assume that the coordinate system is aligned so that the signal arrives at zero azimuth. Using Eq. (14), the cross-spectral matrix for signal-only will be

$$\mathbf{R}^{\text{signal}} = L^2\langle |p^{\text{signal}}(f)|^2\rangle \boldsymbol{\psi}(f,0,0)\boldsymbol{\psi}^H(f,0,0), \tag{15}$$

where $\langle |p^{\text{signal}}(f)|^2\rangle$ is the power spectral density of the incident plane-wave signal. The factor L^2 arises due to the coherent summation of the signal in the phones of each stave.

C. Noise-only matrix

The ambient noise field is modeled as a superposition of independent plane waves arriving from all directions in space, and is characterized by an angular spectral density function, $N(f,\vartheta,\phi)$.[3] This is defined such that $N(f,\vartheta,\phi)\,d\Omega$ gives the contribution to the power spectral density at the output of an unbaffled point omnidirectional phone (i.e., one sufficiently small so as to leave the acoustic field undisturbed) from plane waves arriving within a solid angle $d\Omega = \cos\vartheta d\vartheta d\phi$ in direction $\mathbf{s}(\vartheta,\phi)$. Using Eq. (14), it can be shown that the noise-only cross-spectral matrix can be represented as follows:

$$\mathbf{R}^{\text{noise}} = \int\int_{4\pi} N(f,\vartheta,\phi) \times |h(f,\vartheta)|^2 \boldsymbol{\psi}(f,\vartheta,\phi)\boldsymbol{\psi}^H(f,\vartheta,\phi)d\Omega \tag{16}$$

with elements

$$r^{\text{noise}}_{m,m'} = \int\int_{4\pi} N(f,\vartheta,\phi) \times |h(f,\vartheta)|^2 \psi_m(f,\vartheta,\phi)\psi^*_{m'}(f,\vartheta,\phi)d\Omega. \tag{17}$$

This follows by noting that the component of the wave field that arrives from directions within a differential solid angle $d\Omega$ can be treated as a plane wave at the array. Replacing $\langle |p(f)|^2\rangle$ with $N(f,\vartheta,\phi)\,d\Omega$ in Eq. (14), and integrating over solid angle to account for incoherent summation of contributions from all directions, we obtain Eqs. (16) and (17).

Underwater noise fields are often modeled using azimuth-independent functions; $N(f,\vartheta,\phi)=N_1(f,\vartheta)$. In this case, computation of the noise cross-spectral elements can be simplified by taking advantage of the symmetry. One such representation is

$$r^{\text{noise}}_{m,m'} = \sum_{n=0}^{\infty} \varepsilon_n c_n(f)\cos[n(\phi_m - \phi_{m'})], \tag{18}$$

$$c_n(f) = 2\pi\int_{-\pi/2}^{+\pi/2} N_1(f,\vartheta)|h(f,\vartheta)|^2|g_n(f,\vartheta)|^2\cos\vartheta d\vartheta, \tag{19}$$

$$g_n(f,\vartheta) = J_n(kr\cos\vartheta) + b_n(f,\vartheta)H_n^{(1)}(kr\cos\vartheta). \tag{20}$$

Here ε_n and $b_n(f,\vartheta)$ are defined in Eqs. (5) and (6), respectively. The above-presented result is derived in the appendix, and was used to compute $\mathbf{R}^{\text{noise}}$ for the examples given in Sec. IV involving 3D isotropic, 2D isotropic, and surface dipole noise fields.[3,4]

V. ARRAY GAIN

Array gain compares the SNR obtained by beamforming an array to the SNR obtained at the output of a single omnidirectional phone that leaves the acoustic field undisturbed.[1,4] We assume that the array is placed in an acoustic field comprised of a single signal together with an additive and independent ambient noise field, and that the beamformer is steered to the direction of the signal. The cross-spectral matrices for signal and noise at the stave outputs are given by Eqs. (15) and (16), respectively. Then the SNR at the output of a frequency-domain beamformer[1] used to process the stave outputs will be given by

$$\text{SNR}_{\text{BF}} = \frac{\mathbf{v}^H\mathbf{R}^{\text{signal}}\mathbf{v}}{\mathbf{v}^H\mathbf{R}^{\text{noise}}\mathbf{v}}, \tag{21}$$

where $\mathbf{v}$ is the $M\times 1$ beamformer weight vector. Also, the SNR at the output of an omnidirectional phone that does not disturb the acoustic field will be given by

$$\text{SNR}_{\text{Omni}} = \frac{\langle |p^{\text{signal}}(f)|^2\rangle}{\int\int_{4\pi} N(f,\vartheta,\phi)d\Omega}. \tag{22}$$

Array gain is defined as the ratio of SNR_{BF} to SNR_{Omni}, and is usually expressed in decibels,

$$\text{AG} = 10\log_{10}\frac{\text{SNR}_{\text{BF}}}{\text{SNR}_{\text{Omni}}}. \tag{23}$$

Note that array gain is a function of frequency, and depends on the angular spectral density of the noise.

In Sec. VI we give illustrative array gain computations for two beamformers. The first beamformer (BF1) is the op-

timum spatial filter that maximizes SNR in the given acoustic field.[1,3] In this case the weight vector is proportional to

$$\mathbf{v} = \frac{(\mathbf{R}^{\text{noise}})^{-1}\boldsymbol{\psi}(f,0,0)}{\boldsymbol{\psi}^H(f,0,0)(\mathbf{R}^{\text{noise}})^{-1}\boldsymbol{\psi}(f,0,0)}. \tag{24}$$

This beamformer is used to establish the upper limit on array gain, as no other processor can give a larger array gain for the given array when it is immersed in the given acoustic field. The second beamformer (BF2) is derived from Eq. (24) by assuming that the noise at the output of different staves is IID (even though we will apply it when this is not the case). Here the weight vector is given by

$$\mathbf{v} = \frac{\boldsymbol{\psi}(f,0,0)}{\boldsymbol{\psi}^H(f,0,0)\boldsymbol{\psi}(f,0,0)} \tag{25}$$

and is proportional to the array output vector when signal-only is present. Note that if there was no baffle scatter, so that the staves sampled the incident acoustic field only, then Eq. (25) would correspond to the weight vector for an unshaded conventional beamformer (CBF). Thus we can view Eq. (25) as a generalization of CBF that accounts for effects associated with baffle scatter.

VI. ILLUSTRATIVE EXAMPLES

A. Model parameters

Consider an array with $Q=32$ staves in total, $M=8$ staves on the processing arc, and $L=6$ equispaced phones per stave with normalized spacing $\Delta z/r=\pi/32$. Array gain was computed for three idealized, azimuth-independent noise fields: 3D isotropic, 2D isotropic, and surface dipole.[3,4] For 3D isotropic noise, the angular spectral density is independent of direction. For 2D isotropic noise, the power arrives from a single elevation, which we take to be zero. Thus the angular spectral density has an elevation dependence given by $N_1=N_0\delta(\vartheta)$, where δ is the Dirac delta function. This is sometimes used to model the directionality of ocean noise at low frequencies, where noise is dominated by distant shipping. For surface dipole noise, the elevation dependence is given by $N_1=N_0 \sin\vartheta$ for $0° \leqslant \vartheta \leqslant 90°$ and $N_1=0$ otherwise. This is used to model the directionality of ocean noise at intermediate frequencies, where noise is dominated by wave action at the ocean surface (which we assume to lie above the array, perpendicular to its axis). Computations were carried out for frequencies up to $3f_o$, where $f_o=cQ/(4\pi r)$ is the frequency at which the stave spacing on the arc equals half a wavelength.

B. Zero stave offset

Figure 3 plots array gain for each noise field, assuming staves are flush against the baffle surface, $r/a=1$. Solid curves show results for beamformer BF2. These curves are indistinguishable from those for unshaded CBF, because there is little difference between the two weight vectors in this case. Dashed curves show corresponding results for beamformer BF1, and show the maximum array gain achievable by any processing scheme. Note that array gain is highest for dipole noise, and lowest for 2D isotropic noise. This

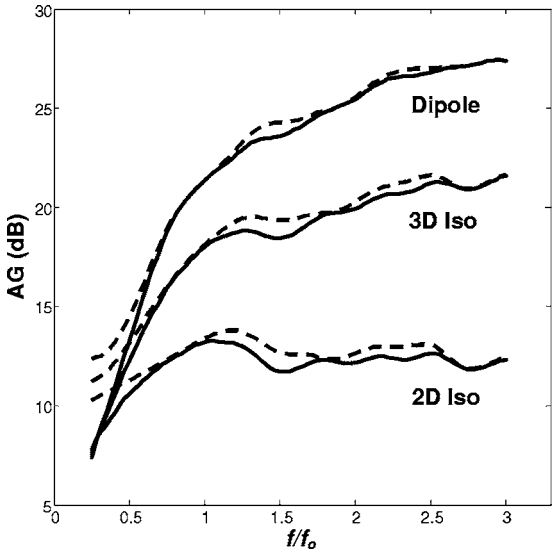


FIG. 3. Array gain for an array with zero stave offset. Solid curves: Beamformer BF2. Dashed curves: Beamformer BF1. The dashed curves indicate the maximum array gain for the given array in the given signal and noise fields.

is due, in large part, to the elevational response of the individual staves, which tends to suppress plane waves arriving from nonzero elevations. Dipole noise arrives with greatest power from high elevations, so the noise at the stave outputs is reduced by the elevational stave response, while 2D isotropic noise arrives from zero-elevation only. Note that the curves for BF1 and BF2 almost touch at several frequencies; at these frequencies the noise output of different staves is almost IID, so that the weight vectors for BF1 and BF2 are almost the same. At other frequencies, beamformer BF2 does not achieve the maximum array gain possible, so there is potential for more sophisticated beamformers to improve performance. The improvement is most significant at low frequencies, where the noise at the output of different staves is more highly correlated.

C. Nonzero stave offset

Now suppose the staves are offset from the baffle by a distance of one-tenth of the wavelength at frequency f_o, or $r/a=1.04$. Figure 4 plots array gain for beamformer BF2 in each noise field. Solid curves show the results for nonzero stave offset, while for comparison, dashed curves show the corresponding results for zero stave offset. The presence of the offset causes a dramatic drop in array gain, so that a minimum is reached at a frequency near $2.7f_o$. This degradation is a consequence of interference between the incident and scattered acoustic fields. The lowest frequency at which an acoustic ray can pass through a point phone and be canceled by its normal reflection from the baffle surface is $2.5f_o$. However, clearly an incident plane wave will not be perfectly canceled by its own reflection in all staves at the same frequency, due to the curved array geometry. The array gain degradation is observed to extend down to quite low frequencies compared to $2.7f_o$. In fact, for dipole and 3D isotropic noise fields, array gain is noticeably reduced even at frequency f_o. Also note that array gain dips below zero for the

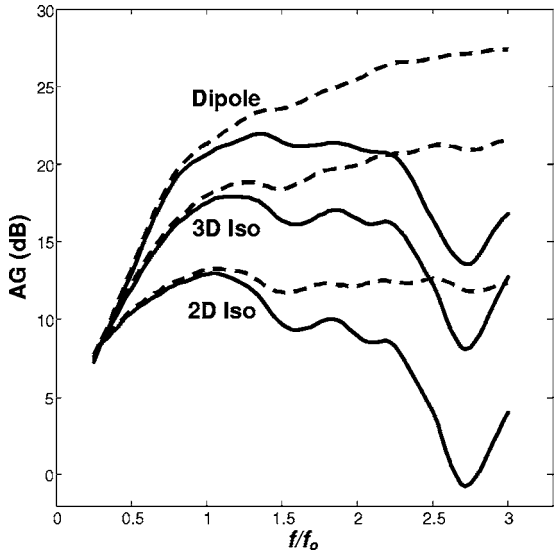


FIG. 4. Array gain for beamformer BF2. Solid curves: Nonzero stave offset. Dashed curves: Zero stave offset. The nonzero stave offset causes a degradation in array gain at high frequencies.

2D isotropic noise case; here the array performs worse than the single omnidirectional phone used as a reference. This is possible because the reference phone is unbaffled, so there is no scattering to destructively interfere with the incident field. Finally, Fig. 5 shows results obtained using beamformer BF1, and demonstrates that the array gain lost due to the presence of the stave offset cannot be fully recovered by more sophisticated processing.

VII. CONCLUSION

This paper has introduced a model for computing array gain at high frequencies for a particular type of baffled cylindrical array that is often used in passive underwater surveillance. The array is equivalent to a baffled circular array of directional elements, where each element corresponds to a linear array of phones that is aligned with the cylinder axis and steered to broadside. The model incorporates baffle scatter using infinite, rigid cylinder scattering theory, and directional noise fields that are characterized by angular spectral density functions. The model can be used to examine the high frequency performance degradation that arises when the sensors are offset from the baffle surface. An offset is often included in practice, for the purpose of improving phone coupling to the water, and to reduce the coupling of baffle vibrations to the phones. Some illustrative examples were given, using noise fields that are often used in sonar performance modeling.

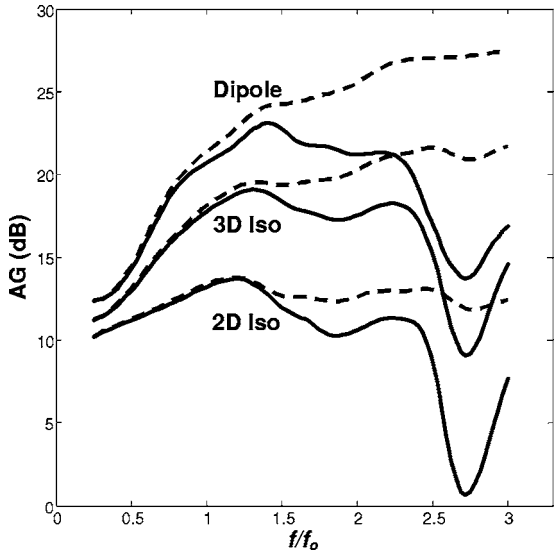


FIG. 5. Array gain for beamformer BF1. Solid curves: Nonzero stave offset. Dashed curves: Zero stave offset. The degradation due to nonzero stave offset is not eliminated through optimum processing that maximizes array gain.

APPENDIX: NOISE CROSS-SPECTRAL ELEMENTS FOR AZIMUTH-INDEPENDENT NOISE

Here we derive Eqs. (18) and (19) for the noise cross-spectral elements in an azimuth-independent noise field. We begin with the following cylindrical harmonic expansion:[5]

$$e^{-jkr\cos\vartheta\cos(\phi-\phi_m)} = \sum_{n=0}^{\infty} \varepsilon_n(-j)^n J_n(kr\cos\vartheta) \times\cos[n(\phi-\phi_m)]. \tag{A1}$$

Using Eq. (A1) together with Eqs. (4), (9), and (20), we find that $\psi_m(f,\vartheta,\phi)$ can be represented by

$$\psi_m(f,\vartheta,\phi) = \sum_{n=0}^{\infty} \varepsilon_n(-j)^n g_n(f,\vartheta)\cos[n(\phi-\phi_m)]. \tag{A2}$$

Thus

$$\psi_m(f,\vartheta,\phi)\psi^*_{m'}(f,\vartheta,\phi) = \sum_{n=0}^{\infty}\sum_{n'=0}^{\infty} (-j)^n(+j)^{n'}\varepsilon_n\varepsilon_{n'}g_n(f,\vartheta)g^*_{n'}(f,\vartheta) \times\cos[n(\phi-\phi_m)]\cos[n'(\phi-\phi_{m'})]. \tag{A3}$$

We introduce the following orthogonality relation, which is valid for $n, n'=0,1,2\ldots$, and where $\delta_{n,n'}$ is the Kronecker delta function; $\delta_{n,n'}=1$ if $n=n'$ and $\delta_{n,n'}=0$ otherwise,

$$\int_0^{2\pi} \cos[n(\phi-\phi_m)]\cos[n'(\phi-\phi_{m'})]d\phi = 2\pi\frac{\delta_{n,n'}}{\varepsilon_n}\cos[n(\phi_m-\phi_{m'})]. \tag{A4}$$

Integrating both sides of Eq. (A3) over ϕ from 0 to 2π, and utilizing Eq. (A4), we obtain

$$\int_0^{2\pi} \psi_m(f,\vartheta,\phi)\psi^*_{m'}(f,\vartheta,\phi)d\phi = 2\pi\sum_{n=0}^{\infty}\varepsilon_n|g_n(f,\vartheta)|^2\cos[n(\phi_m-\phi_{m'})]. \tag{A5}$$

Substituting $N(f,\vartheta,\phi)=N_1(f,\vartheta)$ and $d\Omega=\cos\vartheta d\vartheta d\phi$ into Eq. (17), integrating over ϕ using Eq. (A5), and then revers-

ing the order of integration and summation, we obtain Eqs. (18) and (19).

[1]H. L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation and Modulation Theory* (Wiley-Interscience, New York, 2002).
[2]D. E. N. Davies, "Circular arrays," in *The Handbook of Antenna Design*, edited by A. W. Rudge, K. Milne, A. D. Olver, and P. Knight (Peter Peregrinus, London, 1983), Vol. **II**.
[3]W. S. Burdic, *Underwater Acoustic System Analysis*, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ, 1991).
[4]J. E. Barger, "Sonar systems," in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley-Interscience, New York, 1997), Vol. **I**.
[5]E. A. Skelton and J. H. James, *Theoretical Acoustics of Underwater Structures* (Imperial College Press, London, 1997).
[6]M. C. Junger and D. Feit, *Sound, Structures, and their Interaction*, 2nd ed. (MIT, Cambridge, MA, 1986).
[7]J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**, 185–193 (2001).
[8]G. A. Daigle, M. R. Stinson, and J. G. Ryan, "Beamforming with air-coupled surface waves around a sphere and circular cylinder," J. Acoust. Soc. Am. **117**, 3373–3376 (2005).
[9]W. C. Queen, "Directional characteristics of cylindrical receiving arrays with nonuniform hydrophone response," J. Acoust. Soc. Am. **62**, 400–403 (1977).
[10]T. Rahim and D. E. N. Davies, "Effect of directional elements on the directional response of circular antenna arrays," IEE Proc., Part H: Microwaves, Opt. Antennas **129**, 18–22 (1982).
[11]H. Teutsch and W. Kellermann, "Acoustic source detection and localisation based on wavefield decomposition using circular microphone arrays," J. Acoust. Soc. Am. **120**, 2724–2736 (2006).
[12]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).

Optimal array pattern synthesis for broadband arrays

Shefeng Yan[a)]
Department of Electronics and Telecommunications, Norwegian University of Science and Technology, 7491, Trondheim, Norway

Yuanliang Ma
Institute of Acoustic Engineering, Northwestern Polytechnical University, 710072, Xi'an, People's Republic of China

Chaohuan Hou
Institute of Acoustics, Chinese Academy of Sciences, 100080, Beijing, People's Republic of China

(Received 30 March 2007; revised 21 August 2007; accepted 23 August 2007)

Broadband beamformers with constant mainlobe response over the frequency of interest are desirable in many applications including underwater acoustics, ultrasonics, acoustic imaging and communications, and so on. Solutions to this problem have been presented for specific array geometry often requiring a larger number of sensors. And the array pattern synthesis error minimization is employed for the whole field of view, which leads to suboptimal designs. In this paper, a broadband array pattern synthesis approach to designing time-domain constant mainlobe response beamformer is proposed. By imposing constraints both on the mainlobe spatial response variation over frequency and on the sidelobes of the beamformer, several optimization criteria are presented and the corresponding convex second-order cone programming implementations are given. In this approach, no preliminary desired beampattern is required and the beam responses variation minimization is employed only in the mainlobe region and not in the sidelobe regions, which improves the beamformer mainlobe synthesis accuracy. Equally, one can obtain lower sidelobes at the same mainlobe synthesis accuracy. This approach is applicable to arrays with arbitrary geometry. Simulation and experimental results are presented to illustrate the effectiveness of this approach. Performance comparisons of the proposed beamformers and the existing beamformer are also provided. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2785037]

PACS number(s): 43.60.Fg [EJS] Pages: 2686–2696

I. INTRODUCTION

Array signal processing has wide applications in sonar, radar, wireless communications, astronomy, seismology, medical imaging, etc.[1–3] Broadband beamforming is one of the most important tasks in array signal processing.[4] Broadband beamforming includes discrete Fourier transform (DFT) implementation of a frequency-domain beamformer and finite impulse response (FIR) implementation of a time-domain beamformer. The former is implemented by a narrow-band decomposition structure, whereby the signal received at each sensor is transformed into the frequency domain using DFT and each narrow band of frequencies is treated as an independent narrow-band beamformer. The latter is implemented by FIR filter bank structure, in which each sensor feeds a transversal FIR (or tapped delay line) filter and the filter outputs are summed to produce the output time series. The relationship between the tapped delay line and DFT processing in adaptive arrays was presented by Compton,[5] in which it was shown that the signal-to-interference-plus-noise ratio (SINR) achieved by the two processors are identical when the number of taps in the delay lines is equal to that of samples used in the DFT.

The frequency-domain implementation is not suitable for some applications due to its associated time delay. The important advantage of the time-domain implementation is that we can update the beamformer when each new snapshot arrives. The Frost processor[6] and the generalized sidelobe canceller beamformer[7] are two types of traditional broadband FIR beamformers. However, the sidelobes of these beamformers are not controlled, and this can lead to severe performance degradation in the case of unexpected (i.e., suddenly appearing) interferers. Godara[8] has also developed frequency-domain techniques for computing the tap weights in the FIR beamformer. However, this algorithm may lead to large number of taps, and hence high computational cost, since it requires that the number of taps should be equal to the number of frequency bins. Moreover, the sidelobes cannot be controlled strictly. Recently, we proposed a FIR beamformer with strict sidelobe control using convex optimization approach.[9] However, the mainlobe beampattern was not taken into account in this approach.

In broadband beamforming, it is well known that as the frequency increases, the beamwidth of conventional beamformers decreases. This beamwidth variation as a function of frequency will subject the signals incident on the outer portions of the main beam to low-pass filtering and lead to distorted signal spectra. To deal with this problem, the so-called constant (frequency independent) mainlobe response (CMR)

[a)]Author to whom correspondence should be addressed. Electronic mail: sfyan@ieee.org

beamformer, i.e., beamformer whose mainlobe responses are constant over the frequency band of interest, is required as a remedy. The CMR beamformer is particularly useful and has been applied as a preprocessing structure in broadband beamspace DOA estimator[10–13] to improve its performance. The CMR beamformers employed in Refs. 10 and 11 are DFT implementations and the others are FIR implementations.

Several methods of designing a CMR beamformer have been proposed,[14–21] among them some were FIR implementation.[14–16,20,21] In Ref. 14, an analog technique based on approximating an ideal continuous aperture using the scale-frequency relationship of the array aperture was presented. The CMR beamformer is obtained by using dilation filters, which were designed in Ref. 15, at array elements before weighting. The sensor filters perform the role of beam shaping and ensure that the beam shape is constant as a function of frequency. However, the frequency-invariant beampatterns may not be achieved for arbitrary array geometry. The approaches presented in Refs. 16–18 also require that the arrays have a restricted geometry. More recent work presented spherical harmonics basis method[19] to construct a frequency-invariant response for arbitrarily chosen array geometry. In most of these approaches,[10,14–19] the array pattern synthesis error minimization is employed not only for the mainlobe but also for the sidelobes, which leads to suboptimal designs since the sidelobes only need to be guaranteed to remain below the prescribed threshold value.

Recently, we proposed an approach to designing FIR beamformer with constant mainlobe responses and controlled sidelobes.[20,21] A desired mainlobe beampattern is chosen, and then the FIR filters were optimally designed by minimizing the maximum synthesis error between the desired pattern and the synthesized pattern at each frequency bin within the frequency band. This approach provided high accuracy constant mainlobe response beamformer with low sidelobes. However, the main shortcoming of the approach is that it is not clear how to choose an optimal desired beampattern so that the resulting FIR beamformer has minimal mainlobe synthesis error and lower sidelobes.

In this paper, a new approach to FIR beamformer with constant mainlobe responses is developed. Several optimization criteria for this problem are presented and the corresponding convex second-order cone programming (SOCP) implementations are given. In this new approach, no preliminary desired beampattern is required. In formulating the optimization problem, no assumption has been made on the gain and phase response or characteristics of the individual array elements. Therefore, this proposed approach is applicable to arrays with arbitrary geometry. The element patterns can also be taken into account.

The paper is organized as follows. Some background of FIR beamformer is presented in Sec. II, where the earlier FIR beamforming techniques are overviewed. In Sec. III, we describe new formulations of FIR beamformer with constant mainlobe responses under several optimization criteria. In Sec. IV, the convex SOCP based implementations of our beamformer design problems are described. Section V presents simulation results where the performance of the proposed methods is compared with the existing algorithm and also with each other. In Sec. VI, the proposed methods are applied to real underwater acoustic data. Section VII contains our concluding remarks.

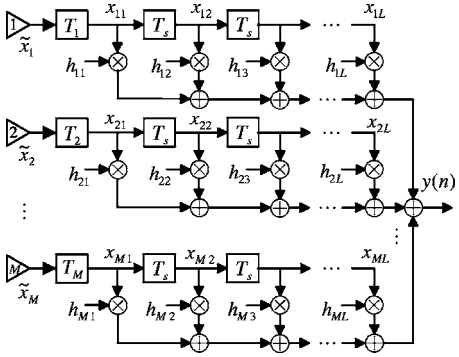


FIG. 1. FIR beamforming structure.

II. BACKGROUND

Consider an M-element array with a known arbitrary geometry. Assume that $D+1<M$ broadband point source plane waves impinge on the array from directions $\Phi=[\phi_0,\phi_1,\ldots,\phi_d,\ldots,\phi_D]$. The time series received at the mth element can be modeled as

$$\tilde{x}_m(t)=\beta s_0[t-\tau_m(\phi_0)]+\sum_{d=1}^{D} s_d[t-\tau_m(\phi_d)]+\tilde{\nu}_m(t), \quad (1)$$

$$m=1,\ldots,M,$$

where $\{s_d(t)\}_{d=0}^{D}$ are the $D+1$ source signals as observed at an arbitrarily chosen reference point, $\tau_m(\phi_d)$ is the propagation delay between the mth sensor and the reference point associated with the dth source, and $\tilde{\nu}_m(t)$ is the additive noise at the mth sensor. We assume that the first term in Eq. (1) corresponds to the signal of interest (SOI) and that of the second term to D interferences. β is a binary parameter indicating whether the SOI is present in the training snapshots.

Figure 1 illustrates the FIR beamforming structure consisting of M sensors and L taps per sensor. T_s is the time delay between any two adjacent taps, i.e., sampling interval. A group of predelays T_m, $m=1,\ldots,M$ are attached after the sensors.

For a conventional Frost beamformer,[6] the presteering delay is used to align the signals to the look direction, i.e., $T_m=-\tau_m(\phi_0)$. Here, the look direction is assumed to be equal to ϕ_0, the direction of arrival of the SOI. Since the predelays are usually not integer numbers of the sampling interval, mechanical or electronic presteering is applied.

The predelayed sampled data from the mth sensor is $x_m(n)=\tilde{x}_m(t-T_m)|_{t=nT_s}$. The input to the other tap weights is given by

$$x_{ml}(n)\triangleq x_m[n-(l-1)]$$
$$=\tilde{x}_m[t-T_m-(l-1)T_s]\big|_{t=nT_s}, \quad (2)$$

$$m=1,\ldots,M,\ l=1,\ldots,L.$$

All the data samples are weighted and summed together to obtain the desired output signal. Let h_{ml} be the adjustable weight of the lth tap following behind the mth sensor. We assume that the weights are real but not complex. Therefore, the algorithms in this paper are described by the operations on the real array data and real beamforming weights.

The time series of the beamformer output is given by

$$y(n) = \sum_{m=1}^{M} \sum_{l=1}^{L} h_{ml} x_m[n-(l-1)] = \sum_{m=1}^{M} \sum_{l=1}^{L} h_{ml} x_{ml}(n). \quad (3)$$

We can define two $M \times L$ matrices,

$$\mathbf{X}(n) = \begin{bmatrix} x_{11}(n) & x_{12}(n) & \cdots & x_{1L}(n) \\ x_{21}(n) & x_{22}(n) & \cdots & x_{2L}(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1}(n) & x_{M2}(n) & \cdots & x_{ML}(n) \end{bmatrix}, \quad (4)$$

and

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1L} \\ h_{21} & h_{22} & \cdots & h_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{ML} \end{bmatrix}. \quad (5)$$

Let us introduce the notations

$$\mathbf{x}(n) = \mathrm{vec}\{\mathbf{X}(n)\}, \quad (6)$$

$$\mathbf{h} = \mathrm{vec}\{\mathbf{H}\}, \quad (7)$$

where $\mathrm{vec}\{\cdot\}$ denotes the vectorization operator which stacks the columns of a matrix on top of each other, and $\mathbf{x}(n)$ represents the nth snapshot. Then Eq. (3) can be written as

$$y(n) = \mathbf{h}^T \mathbf{x}(n), \quad (8)$$

where $(\cdot)^T$ denotes the transpose.

For a broadband array, its beampattern is a function of both frequency and direction. The array response (frequency–wave-number response) of the FIR beamformer can be written as[9]

$$p(f,\theta) = \mathbf{h}^T \mathbf{u}(f,\theta) = \mathbf{u}^T(f,\theta)\mathbf{h}, \quad (9)$$

where $\mathbf{u}(f,\theta)$ is the $ML \times 1$ array manifold vector corresponding to the $ML \times 1$ stacked vector $\mathbf{x}$. It is given by

$$\mathbf{u}(f,\theta) = \mathbf{e}(f) \otimes [\mathbf{a}(f,\theta) \circ \boldsymbol{\kappa}(f)], \quad (10)$$

where $\otimes$ denotes the Kronecker product and $\circ$ denotes the Hadamard (i.e., element-wise) product of two vectors. $\mathbf{e}(f) = [1, e^{-j2\pi f T_s}, \ldots, e^{-j(L-1)2\pi f T_s}]^T$, $\boldsymbol{\kappa}(f) = [e^{-j2\pi f T_1}, \ldots, e^{-j2\pi f T_m}, \ldots, e^{-j2\pi f T_M}]^T$, $\mathbf{a}(f,\theta)$ is the array manifold vector at frequency f and angle θ, which is given by $\mathbf{a}(f,\theta) = [e^{-j2\pi f \tau_1(\theta)}, \ldots, e^{-j2\pi f \tau_M(\theta)}]^T$. Arbitrary element patterns can also be multiplied to elements of $\mathbf{a}(f,\theta)$ if necessary.

Thus, the $[m+(l-1)M]$th entry of $\mathbf{u}(f,\theta)$ is

$$[\mathbf{u}(f,\theta)]_{m+(l-1)M} = e^{-j2\pi f[T_m + \tau_m(\theta) + (l-1)T_s]} = e^{-j2\pi f \xi_{ml}}, \quad (11)$$

where $\xi_{ml} = T_m + \tau_m(\theta) + (l-1)T_s$.

In our earlier paper,[20] the predelay is chosen as

$$T_m = -\mathrm{int}[\tau_m(\phi_0)/T_s + (L-1)/2] \cdot T_s, \quad m = 1, \ldots, M, \quad (12)$$

where $\mathrm{int}[\cdot]$ denotes round toward nearest integer. From Eq. (12), we can see that the predelays are integer numbers of the sampling interval and can be implemented by a set of tapped delay lines, which is more convenient than the Frost processor.

Without loss of generality, we assume that the sensor data have finite frequency band $[f_l, f_u]$, where f_l and f_u are the lower and upper bound frequency, respectively, and the frequencies outside this band have been removed before array processing. We can discretize the frequency band using a finite number of frequency grids $f_k \in [f_l, f_u]$, $k = 1, \ldots, K$. The frequencies can be uniformly spaced and the narrow-band condition should be satisfied within each frequency bin:

$$B_s \Delta T_{\max} \ll 1, \quad (13)$$

where B_s is the bandwidth of each frequency bin and $\Delta T_{\max}$ is the maximum travel time between any two elements in the array.

A reasonable tradeoff should be made between the design accuracy and the complexity when choosing the grid size. For more grids, it would lead to better accuracy but longer computation time, and vice versa.

In array pattern synthesis, how well a synthesized pattern approximates a desired pattern can be measured in several ways. Typical measures are the peak synthesis error across θ, or the mean-squared error over θ. The CMR beamformer optimization problem in Ref. 20 is formulated as a minimax optimization problem,

$$\min_{\mathbf{h}} \max_{k,q} |p_d(\theta_q) - \mathbf{u}^T(f_k,\theta_q)\mathbf{h}|, \quad f_k \in [f_l, f_u],$$

$$k = 1,2,\ldots,K, \quad \theta_q \in \Theta_{\mathrm{ML}}, \quad q = 1, \ldots, Q,$$

$$\text{subject to } |\mathbf{u}^T(f_k,\theta_s)\mathbf{h}| \leqslant \epsilon, \quad f_k \in [f_l, f_u],$$

$$k = 1,2,\ldots,K, \quad \theta_s \in \Theta_{\mathrm{SL}}, \quad s = 1, \ldots, S, \quad (14)$$

where $\theta_s \in \Theta_{\mathrm{SL}}$, $s = 1, \ldots, S$ and $\theta_q \in \Theta_{\mathrm{ML}}$, $q = 1, \ldots, Q$ are the angular grids chosen that properly approximate the sidelobe region Θ_{SL} and mainlobe region Θ_{ML} by a finite number of directions, respectively. The choice of S and Q is determined by the required accuracy of approximation. $p_d(\theta_q)$, $q = 1, \ldots, Q$ is the preliminary desired beampattern within the mainlobe region Θ_{ML}. ϵ is a small positive value chosen to control the sidelobes.

The norm constraint

$$||\mathbf{h}|| \leqslant \Delta \quad (15)$$

can also be added to Eq. (14) to control the dynamic range of the filter coefficients to void large noise amplification at the array output, which is similar to the so-called white noise gain constraint employed in narrow-band robust adaptive beamforming.[22] Here $||\cdot||$ denotes the Euclidean norm, and Δ is a prescribed value to control the norm of FIR filter tap weights.

However, it can be seen from Eq. (14) that a desired beampattern should be chosen preliminarily, which is a main

shortcoming of this approach. It is not clear how to choose an optimal desired beampattern so that the resulting CMR beamformer has minimal mainlobe synthesis error and lower sidelobes.

III. PROPOSED BEAMFORMERS

We hereby present a new approach to CMR beamformer using several different criteria by introducing the mainlobe spatial response variation (MSRV), as a measure of the quality of the approximation among the synthesized patterns at different frequencies.

A. Peak sidelobes constrained minimax MSRV design

The maximum MSRV of a CMR beamformer is defined as

$$\sigma_{\max} = \max_{k,q} |\mathbf{u}^T(f_0,\theta_q)\mathbf{h} - \mathbf{u}^T(f_k,\theta_q)\mathbf{h}| = \max_{k,q} |[\mathbf{u}(f_0,\theta_q) - \mathbf{u}(f_k,\theta_q)]^T\mathbf{h}|, \quad f_k \in [f_l,f_u], \quad (16)$$

$$k = 1,2,\ldots,K, \quad \theta_q \in \Theta_{\mathrm{ML}}, \quad q = 1,\ldots,Q,$$

where f_0 is the reference frequency. It can be any frequency value selected within $[f_l,f_u]$.

Replacing the maximum mainlobe synthesis error in Eq. (14) by maximum MSRV $\sigma_{\max}$ defined in Eq. (16), and imposing the distortionless constraint on the signal of interest give the peak sidelobes constrained minimax MSRV beamformer design,

$$\min_{\mathbf{h}} \sigma_{\max}, \quad \text{subject to } \mathbf{u}^T(f_0,\phi_0)\mathbf{h} = 1,$$

$$|\mathbf{u}^T(f_k,\theta_s)\mathbf{h}| \leqslant \epsilon, \quad f_k \in [f_l,f_u], \quad \theta_s \in \Theta_{\mathrm{SL}}. \quad (17)$$

Introducing a new non-negative variable σ, Eq. (17) can be converted into the following form:

$$\min_{\mathbf{h}} \sigma, \quad \text{subject to } \mathbf{u}^T(f_0,\phi_0)\mathbf{h} = 1,$$

$$|[\mathbf{u}(f_k,\theta_q) - \mathbf{u}(f_0,\theta_q)]^T\mathbf{h}| \leqslant \sigma,$$

$$f_k \in [f_l,f_u], \quad \theta_q \in \Theta_{\mathrm{ML}},$$

$$|\mathbf{u}^T(f_k,\theta_s)\mathbf{h}| \leqslant \epsilon, \quad f_k \in [f_l,f_u], \quad \theta_s \in \Theta_{\mathrm{SL}}. \quad (18)$$

It is seen that no preliminary desired beampattern is required in this method, which implies that an optimal desired beampattern was chosen automatically. This means that the new method can obtain smaller mainlobe pattern synthesis error.

B. Peak sidelobes constrained least-square MSRV design

Consider the mean-squared MSRV relative to the response at the reference frequency f_0, it is defined as

$$\sigma_{\mathrm{rms}}^2 = \frac{1}{Q}\sum_{q=1}^{Q}\left\{\frac{1}{K}\sum_{k=1}^{K}[|\mathbf{u}^T(f_k,\theta_q)\mathbf{h} - \mathbf{u}^T(f_0,\theta_q)\mathbf{h}|^2]\right\} = \frac{1}{QK}\sum_{q=1}^{Q}\sum_{k=1}^{K}\{\mathbf{h}^T[\mathbf{u}(f_k,\theta_q) - \mathbf{u}(f_0,\theta_q)][\mathbf{u}(f_k,\theta_q) - \mathbf{u}(f_0,\theta_q)]^H\mathbf{h}\}, \quad f_k \in [f_l,f_u], \quad (19)$$

$$k = 1,2,\ldots,K, \quad \theta_q \in \Theta_{\mathrm{ML}}, \quad q = 1,\ldots,Q,$$

where the superscript H denotes the Hermitian transpose. It takes the form

$$\sigma_{\mathrm{rms}}^2 = \mathbf{h}^T\mathbf{\Omega}\mathbf{h}, \quad (20)$$

where $\mathbf{\Omega}$ is a $ML \times ML$ matrix,

$$\mathbf{\Omega} = \frac{1}{QK}\sum_{q=1}^{Q}\sum_{k=1}^{K}\{[\mathbf{u}(f_k,\theta_q) - \mathbf{u}(f_0,\theta_q)][\mathbf{u}(f_k,\theta_q) - \mathbf{u}(f_0,\theta_q)]^H\}. \quad (21)$$

Note that the matrix $\mathbf{\Omega}$ is Hermitian and the vector $\mathbf{h}$ is real-valued. Hence, it is easy to obtain

$$\sigma_{\mathrm{rms}}^2 = \mathbf{h}^T\mathbf{\Omega}_R\mathbf{h}, \quad (22)$$

where $\mathbf{\Omega}_R = \mathrm{Re}(\mathbf{\Omega})$, i.e., the real part of $\mathbf{\Omega}$.

By constraining the mean-squared MSRV to be less than a small value, we can restrict the synthesized pattern at different frequencies to be similar to some extent. The constraint is

$$\mathbf{h}^T\mathbf{\Omega}_R\mathbf{h} \leqslant \delta^2, \quad (23)$$

where δ is a small positive variable. This constraint can be rewritten as

$$||\mathbf{\Omega}_R^{1/2}\mathbf{h}|| \leqslant \delta. \quad (24)$$

Therefore, our peak sidelobes constrained least squared MSRV beamformer design problem can be stated as

$$\min_{\mathbf{h}} \delta, \quad \text{subject to } \mathbf{u}^T(f_0,\phi_0)\mathbf{h} = 1,$$

$$||\mathbf{\Omega}_R^{1/2}\mathbf{h}|| \leqslant \delta,$$

$$|\mathbf{u}^T(f_k,\theta_s)\mathbf{h}| \leqslant \epsilon, \quad f_k \in [f_l,f_u], \quad \theta_s \in \Theta_{\mathrm{SL}}, \quad (25)$$

By comparing the constraints in Eqs. (18) and (25), it can be noted that the latter reduced significantly the number of constraints.

C. Mean-squared MSRV constrained least-square sidelobes design

It is seen from Eq. (25) that there are still too many sidelobe constraints, which may still be computationally expensive for most applications. We hereby present a mean-squared MSRV constrained least-squared sidelobes beamformer design method, which is more computationally efficient than the methods presented in Secs. III A and III B, because the sidelobe constraints are reduced.

In this method, we control the mean-squared sidelobes instead of controlling the peak sidelobes. The mean-squared

sidelobes' power can be defined as (some non-negative weighting coefficients can also be used to obtain a weighted mean-squared sidelobes)

$$\sigma_{\text{rmsSL}}^2 = \frac{1}{S}\sum_{s=1}^{S}\left\{\frac{1}{K}\sum_{k=1}^{K}[|\mathbf{u}^T(f_k,\theta_s)\mathbf{h}|^2]\right\}, \quad f_k \in [f_l, f_u], \tag{26}$$

$$k = 1,2,\ldots,K, \quad \theta_s \in \Theta_{\text{SL}}, \quad s = 1,\ldots,S.$$

Similar to Eq. (20), it takes the form

$$\sigma_{\text{rmsSL}}^2 = \mathbf{h}^T\boldsymbol{\Psi}_R\mathbf{h}, \quad f_k \in [f_l, f_u], \quad \theta_s \in \Theta_{\text{SL}}, \tag{27}$$

where $\boldsymbol{\Psi}_R$ is a $ML \times ML$ matrix,

$$\boldsymbol{\Psi}_R = \frac{1}{SK}\sum_{s=1}^{S}\sum_{k=1}^{K}\text{Re}[\mathbf{u}(f_k,\theta_s)\mathbf{u}^H(f_k,\theta_s)]. \tag{28}$$

Hence, the mean-squared sidelobe constraint can be stated as

$$||\boldsymbol{\Psi}_R^{1/2}\mathbf{h}|| \leqslant \xi, \tag{29}$$

where ξ is a small positive variable.

Then, our broadband array pattern synthesis problem can be written as

$$\min_{\mathbf{h}} \xi, \quad \text{subject to } \mathbf{u}^T(f_0,\phi_0)\mathbf{h} = 1,$$

$$||\boldsymbol{\Omega}_R^{1/2}\mathbf{h}|| \leqslant \delta_0, \quad ||\boldsymbol{\Psi}_R^{1/2}\mathbf{h}|| \leqslant \xi, \tag{30}$$

where δ_0 is a prescribed value to control the MSRV.

It is worth noting that when the MSRV or mean-squared sidelobe is used as a constraint in the optimization, it is not necessary to use too small grid spacing in angles due to the correlations between responses at neighboring directions.

D. CMR beamforming with interference rejection property

If a beamformer is fixed, it cannot reject unexpected interfering sources with high strength. Adaptive beamforming can be used to solve this problem. Recently, we proposed a FIR adaptive beamformer with sidelobe control.[9] However, the resulting mainlobe beampattern is not constant over the frequency band. By imposing a MSRV constraint and replacing the peak sidelobe constraint with mean-squared sidelobe constraint, we get

$$\min_{\mathbf{h}} \mathbf{h}^T\mathbf{R}\mathbf{h}, \quad \text{subject to } \mathbf{u}^T(f_0,\phi_0)\mathbf{h} = 1,$$

$$||\boldsymbol{\Omega}_R^{1/2}\mathbf{h}|| \leqslant \delta_0, \quad ||\boldsymbol{\Psi}_R^{1/2}\mathbf{h}|| \leqslant \xi_0, \tag{31}$$

where ξ_0 is a prescribed value to control the mean-squared sidelobes, and $\mathbf{R}$ denotes the $ML \times ML$ covariance matrix of the stacked array output vector, i.e.,

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^T(n)], \tag{32}$$

where $E[\cdot]$ denotes the statistical expectation of the quantity in the brackets.

Assume that the signal sources and noise are statistically independent and the noise is uncorrelated from sensor to sensor. If the signals (interferences) have a flat spectrum, the theoretical covariance matrix of the stacked array output vector has the following form:

$$\mathbf{R} = \text{Re}[\tilde{\mathbf{R}}] = \text{Re}\left(\sum_{d=0}^{D}\tilde{\mathbf{R}}_d + \tilde{\mathbf{R}}_v\right), \tag{33}$$

where $\tilde{\mathbf{R}}_d$ and $\tilde{\mathbf{R}}_v$ are broadband complex covariance matrix corresponding to the dth source and the noise, whose $[m+(l-1)M, \tilde{m}+(\tilde{l}-1)M]$th entries (here $\tilde{m}=1,2,\ldots,M$, $\tilde{l}=1,2,\ldots,L$) are[9]

$$[\tilde{\mathbf{R}}_d]_{m+(l-1)M,\tilde{m}+(\tilde{l}-1)M} = \frac{\sigma_s^2}{B}\,\text{sinc}[\pi B(\xi_{ml} - \xi_{\tilde{m}\tilde{l}})]e^{-j2\pi f_c(\xi_{ml}-\xi_{\tilde{m}\tilde{l}})}, \tag{34}$$

and

$$[\tilde{\mathbf{R}}_v]_{m+(l-1)M,\tilde{m}+(\tilde{l}-1)M} = \frac{\sigma_n^2}{B}\delta_{m\tilde{m}}\,\text{sinc}[\pi B(l - \tilde{l})T_s]e^{-j2\pi f_c(l-\tilde{l})T_s}, \tag{35}$$

respectively. Here $\delta_{m\tilde{m}}$ is the Kronecker delta function and $\text{sinc}[\cdot]=\sin[\cdot]/[\cdot]$. σ_s^2 and σ_n^2 are the total power of the signal source and the noise. f_c and B are the corresponding center frequency and frequency bandwidth, respectively.

In fact, this beamformer is data dependent and does not require any *a priori* information about the spectral densities of signals and noise. In applications, the covariance matrix $\mathbf{R}$ is unavailable. Therefore, the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n) \tag{36}$$

is used instead of Eq. (32). Here N is the number of snapshots.

Clearly, we have

$$\mathbf{h}^T\mathbf{R}\mathbf{h} = ||\mathbf{R}^{1/2}\mathbf{h}||^2. \tag{37}$$

Noting that minimizing $||\mathbf{R}^{1/2}\mathbf{h}||^2$ is equivalent to minimizing $||\mathbf{R}^{1/2}\mathbf{h}||$ and introducing a new auxiliary variable γ, we can rewrite Eq. (31) as

$$\min_{\mathbf{h}} \gamma, \quad \text{subject to } \mathbf{u}^T(f_0,\phi_0)\mathbf{h} = 1,$$

$$||\mathbf{R}^{1/2}\mathbf{h}|| \leqslant \gamma, \quad ||\boldsymbol{\Omega}_R^{1/2}\mathbf{h}|| \leqslant \delta_0, \quad ||\boldsymbol{\Psi}_R^{1/2}\mathbf{h}|| \leqslant \xi_0. \tag{38}$$

IV. CONVEX OPTIMIZATION BASED IMPLEMENTATION

All the broadband array pattern synthesis problems (18), (25), (30), and (38) can be formulated as the convex SOCP problem, which can be solved efficiently using a SOCP solver such as SeDuMi.[23] Convex optimization was first used in array pattern synthesis by Lebret and Boyd.[24] SOCP has also been used in the optimal design of digital FIR filters[25] and narrow-band beamformer.[26–28] The global optimal solution of a SOCP problem is guaranteed if it exists. Another important advantage of SOCP is that it is very convenient to

include additional linear or convex quadratic constraints, such as the norm constraint of the variable vector, to the design problem. A review of the applications of SOCP was presented by Lobo.[29]

A. Review of second-order cone programming

SOCP is a subclass of the well-structured convex programming problems where a linear function is minimized subject to a set of second-order cone constraints and possibly a set of linear equality constraints. The problem can be described as

$$\min_{\mathbf{y}} \mathbf{b}^T\mathbf{y}, \quad \text{subject to } \|\mathbf{A}_i\mathbf{y}+\mathbf{b}_i\| \leqslant \mathbf{c}_i^T\mathbf{y}+d_i,$$

$$i=1,2,\ldots,I,$$

$$\mathbf{F}\mathbf{y}=\mathbf{g}, \tag{39}$$

where $\mathbf{b}\in C^{\alpha\times 1}$, $\mathbf{y}\in C^{\alpha\times 1}$, $\mathbf{A}_i\in C^{(\alpha_i-1)\times\alpha}$, $\mathbf{b}_i\in C^{(\alpha_i-1)\times 1}$, $\mathbf{c}_i\in C^{\alpha\times 1}$, $\mathbf{c}_i^T\mathbf{y}\in\Re$, $d_i\in\Re$, $\mathbf{F}\in C^{g\times\alpha}$, $\mathbf{g}\in C^{g\times 1}$ with $\Re$ and C being the set of real and complex numbers (or matrix).

B. Convex formulations of the array pattern synthesis problems

In this section, we only reformulated the optimization problem (38) in the form of SOCP. The others can be reformulated and solved with a similar processing procedure.

By stacking together γ and the coefficients of $\mathbf{h}$, we define $\mathbf{y}=[\gamma,\mathbf{h}^T]^T$. Let $\mathbf{b}=[1,\mathbf{0}_{1\times ML}]^T$ so that $\gamma=\mathbf{b}^T\mathbf{y}$. Here $\mathbf{0}_{1\times ML}$ is a $1\times ML$ zero vector. The optimization problem (38) can be rewritten in the following form:

$$\min_{\mathbf{y}} \mathbf{b}^T\mathbf{y},$$

$$\text{subject to } \|[\mathbf{0}_{ML\times 1} \quad \mathbf{R}^{1/2}]\mathbf{y}\| \leqslant [1 \quad \mathbf{0}_{1\times ML}]\mathbf{y},$$

$$\|[\mathbf{0}_{ML\times 1} \quad \mathbf{\Omega}_R^{1/2}]\mathbf{y}\| \leqslant \delta_0,$$

$$\|[\mathbf{0}_{ML\times 1} \quad \mathbf{\Psi}_R^{1/2}]\mathbf{y}\| \leqslant \xi_0,$$

$$\begin{bmatrix} 0 & \mathrm{Re}[\mathbf{u}^T(f_0,\phi_0)] \\ 0 & \mathrm{Im}[\mathbf{u}^T(f_0,\phi_0)] \end{bmatrix}\mathbf{y}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{40}$$

where $\mathbf{I}_{ML\times ML}$ is a $ML\times ML$ identity matrix, and $\mathrm{Im}[\cdot]$ represents the imaginary part of the quantity in the brackets.

The optimization problem (40) is now in the form of a standard SOCP as in Eq. (39), which can be easily solved using highly efficient interior point method software tools such as SeDuMi.

C. Computational complexity

When interior-point methods are used to solve the SOCP problem in Eq. (39), the number of iterations to decrease the duality gap to a constant fraction of itself is bounded above by $O(\sqrt{I+1})$ (see Ref. 30, here the term "1" is due to the equality constraint), and the amount of computation per iteration is $O[\alpha^2(\Sigma_i\alpha_i+g)]$.[29]

TABLE I. Computational complexity of the five broadband array pattern synthesis methods.

Method	Number of iterations	Amount of computation per iteration
Eq. (14)	$O(\sqrt{K(Q+S)})$	$O\{(ML)^2[3K(Q+S)]\}$
Eq. (18)	$O(\sqrt{K(Q+S)+1})$	$O\{(ML)^2[3K(Q+S)+2]\}$
Eq. (25)	$O(\sqrt{KS+2})$	$O\{(ML)^2[3KS+ML+3]\}$
Eq. (30)	$O(1)$	$O\{(ML)^2[2ML+4]\}$
Eq. (38)	$O(1)$	$O\{(ML)^2[3ML+5]\}$

The major computational demand of the broadband array pattern synthesis comes from solving the optimization problem in variant criteria. The computational complexities of optimization problems (14), (18), (25), (30), and (38) are calculated and listed in Table I.

From Table I, it is seen that the computational complexity of our method (18) is comparable with that of the earlier method (14), while the new method does not require the preliminary desired beampattern. The computational complexity is reduced by the new method (25), and further reduced significantly by the method (30). The computational cost of the proposed FIR adaptive beamformer (38) only increases a little by comparing to the fixed beamformer (30). However, the former can suppress the interferences arrived from the sidelobe regions and thus improve the output SINR.

V. SIMULATION RESULTS

Several numerical examples are provided to illustrate the performance of our methods for broadband array pattern synthesis.

In all of the examples considered in the following, we assume a linear array of $M=12$ uniformly spaced elements, with a half-wavelength spacing at the upper frequency f_u. The normalized design band is $[f_l,f_u]=[0.16,0.32]$ (the normalized sampling frequency was 1), which is decomposed into $K=33$ uniform frequency bins. The SOI direction of arrival is $\phi_0=10^\circ$ with respect to the broadside of the array. The length of the FIR filters is $L=15$. A uniform grid of 2° is used to discretize the field of view $[-90^\circ,90^\circ]$. Unless otherwise stated, we assume the mainlobe region is $\Theta_{\mathrm{ML}}=[-8^\circ:2^\circ:28^\circ]$ and the sidelobe regions are $\Theta_{\mathrm{SL}}=[-90^\circ:2^\circ:-12^\circ]\cup[32^\circ:2^\circ:90^\circ]$.

A. Example 1

We consider the effect of the desired beampattern on the mainlobe synthesis accuracy of the resulting beamformer for the existing CMR beamformer design method (14).

We choose the mainlobe pattern of a conventional (delay-and-sum) beamformer as the desired beam pattern and consider two cases: the reference frequency $f_0=(f_l+f_u)/2=0.24$ and $f_0=f_l=0.16$. Consider that as the frequency increases, the beamwidth of conventional beamformers decreases, a narrower mainlobe region $\Theta_{\mathrm{ML}}=[-2^\circ:2^\circ:22^\circ]$ is chosen for the former. The desired sidelobe level is set to be -30 dB. The desired mainlobe patterns and the prescribed sidelobe level are shown in Fig. 2(a) for the two cases.

We solve the optimization problem (14) to obtain the CMR beamformer. The beam response as a function of fre-

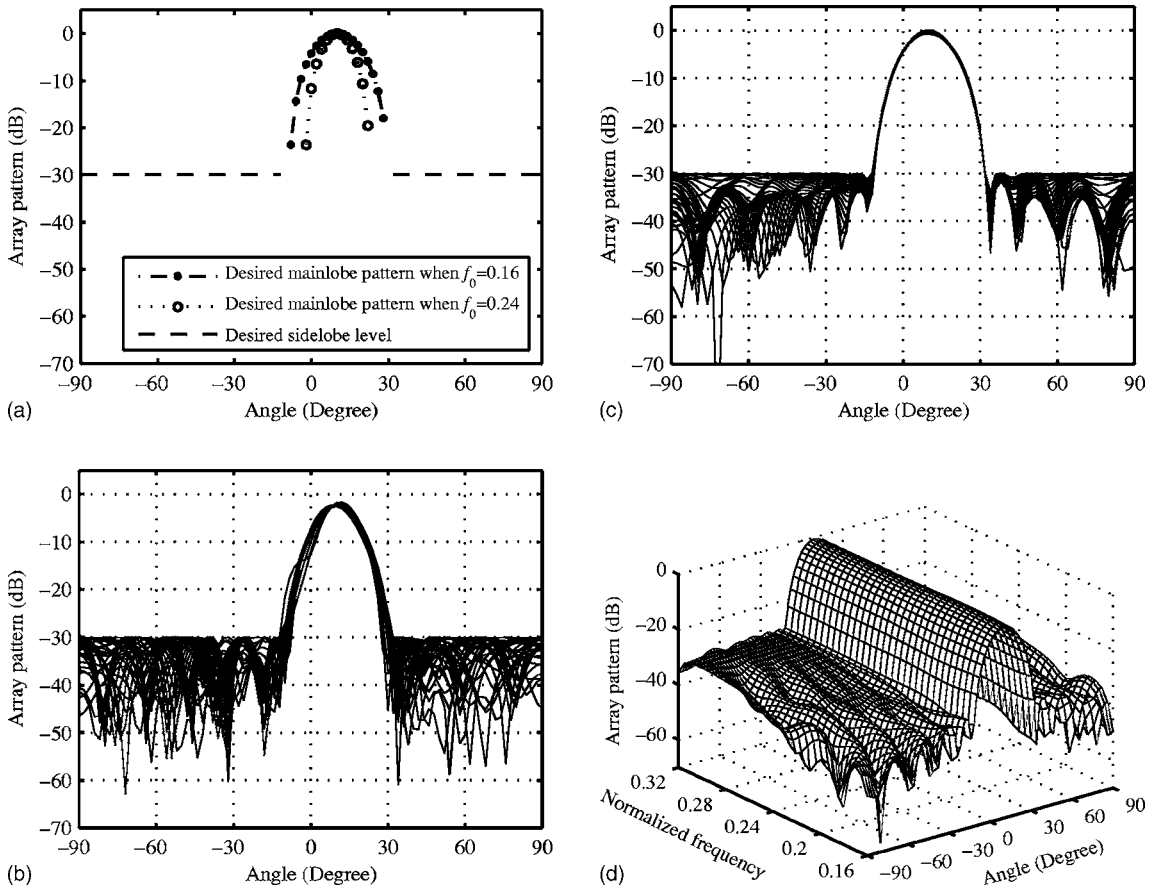


FIG. 2. Results of example 1 using the existing method (14). (a) Desired beam responses and sidelobe level when choosing f_0=0.24 and f_0=0.16, respectively; Superposition of all 33 beampatterns for the case (b) f_0=0.24 and (c) f_0=0.16. (d) The variation of beampattern with frequency and angle for the case f_0=0.16.

quency and angle can be calculated on a grid of points spaced every 0.05 in frequency and 2° in angle using Eq. (9).

Figures 2(b) and 2(c) show the superposition of the achieved broadband array patterns for the case f_0=0.24 and f_0=0.16, respectively. The variation of beampattern with frequency for the case f_0=0.16 is shown in Fig. 2(d). It is seen that the mainlobe pattern is severely distorted for f_0=0.24 and the mainlobe synthesis error is very large. For f_0=0.16, the resulting beampattern within the mainlobe is approximately constant over the frequency band and the sidelobes are strictly guaranteed to be below −30 dB.

From this example, we can conclude that the performance of this array pattern synthesis method depends on the choosing of desired mainlobe pattern. An arbitrarily chosen desired mainlobe pattern can result in very poor synthesis pattern. Therefore, how to choose an optimal desired beampattern is the main limitation of this method.

B. Example 2

We design the CMR beamformer using the first proposed method, i.e., solving the optimization problem (18). In this example, the center frequency is chosen as the reference frequency, i.e., f_0=0.24. The other parameters are fixed. The resulting array pattern is shown in Fig. 3(a). It is seen that the mainlobe beampattern is approximately constant and the sidelobes are strictly guaranteed to be below −30 dB.

We calculate the maximum MSRV of the array patterns as shown in Fig. 2(c) and in Fig. 3(a) using Eq. (16). The maximum MSRV in decibels (i.e., $20\lg\sigma_{\max}$) of the former is −21.4 dB and that of the latter is −41.8 dB, which is much small than that of the former. As expected, our new method (18) outperforms the existing method (14).

C. Example 3

We consider the second proposed method. With all the parameters fixed, the CMR beamformer is designed by solving optimization problem (25). The resulting array pattern is shown in Fig. 3(b). Using Eq. (19), the root-mean-squared MSRV of the array pattern in decibels (i.e., $10\lg\sigma^2_{\text{rms}}$) is −51.2 dB.

D. Example 4

We consider our third method as described in Eq. (30). δ_0 is chosen to make the array pattern have the same root-mean-squared MSRV as in Fig. 3(b), i.e., $20\lg\delta_0$ =−51.2 dB.

The resulting array pattern is shown in Fig. 3(c). Using Eq. (27), the resulting root-mean-squared sidelobe of the array pattern is $10\lg\sigma^2_{\text{rmsSL}}$=−36.2 dB. We see that the peak sidelobe level is −23.9 dB, which is excellent. In fact, a lower sidelobe level can be obtained if we increase the mainlobe beamwidth by adjusting Θ_{ML} and Θ_{SL}.

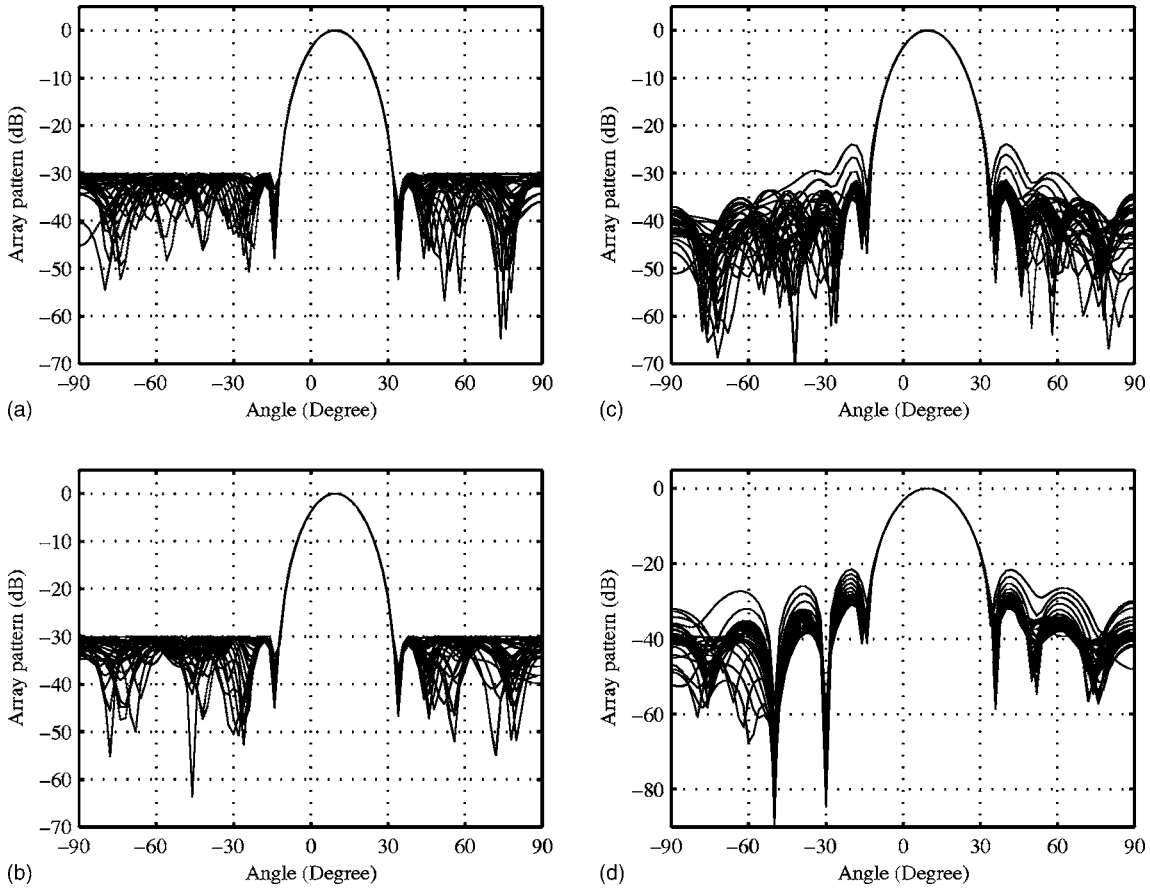


FIG. 3. Superposition of all 33 beampatterns for a 12-element uniform linear array by computer simulation calculated at 33 frequencies within the normalized band [0.16, 0.32] using (a) peak sidelobes constrained minimax MSRV design (18), (b) peak sidelobes constrained least-square MSRV design (25), (c) mean-squared MSRV constrained least-square sidelobes design (30), and (d) CMR beamforming with interference rejection method (38), respectively. The resulting mainlobe synthesis accuracy and sidelobes by these methods are listed in Table II.

E. Example 5

We use the fourth method (38). A signal and two interferers are assumed to impinge on the array from 10°, −30°, and −50° with the signal(interferer)-to-noise ratio (SNR) 0, 30, and 30 dB at each sensor, respectively. The signal, interferers, and noise are assumed to be bandpass white Gaussian process with flat spectral density over the frequency band $[f_l, f_u]$.

The theoretical array covariance matrix $\mathbf{R}$ is used. Let δ_0 be the same as in example 4. We set $20\lg\xi_0 = -33.2$ dB, which is 3 dB higher than the resulting root-mean-squared sidelobe of the fixed beamformer in Fig. 3(c). What we obtain is some nulls in the directions of strong interferers. Our algorithm will provide a suitable compromise among these conflicting objectives (i.e., high output SINR, low sidelobes, and small synthesis error).

By solving optimization problem (38), the obtained FIR broadband array patterns are shown in Fig. 3(d). As expected, the array patterns have constant mainlobe responses and deep nulls in the directions of arrival of the interferers.

F. Comparison of the methods

To compare the performance of the broadband array pattern synthesis methods, Table II lists the maximum MSRV $20\lg\sigma_{\max}$(dB), root-mean-squared MSRV $20\lg\sigma_{\mathrm{rms}}$(dB), peak sidelobe level $20\lg\sigma_{\mathrm{maxSL}}$(dB), root-mean-squared

TABLE II. Performance comparison of the existing beamformer (14) and the proposed beamformers using different array pattern synthesis criteria. Simulation results.

Method	Array pattern	$20\lg\sigma_{\max}$ (dB)	$20\lg\sigma_{\mathrm{rms}}$ (dB)	$20\lg\sigma_{\mathrm{maxSL}}$ (dB)	$20\lg\sigma_{\mathrm{rmsSL}}$ (dB)	CPU time (s)
Eq. (14)	Fig. 2(c)	−21.4	−23.9	−30.0	−33.1	154.0
Eq. (18)	Fig. 3(a)	**−41.8**	−44.8	−30.0	−32.3	90.6
Eq. (25)	Fig. 3(b)	−35.6	**−51.2**	−30.0	−32.1	52.2
Eq. (30)	Fig. 3(c)	−35.8	−51.2	−23.9	**−36.2**	1.70
Eq. (38)	Fig. 3(d)	−37.0	−51.2	−21.5	−33.2	1.78

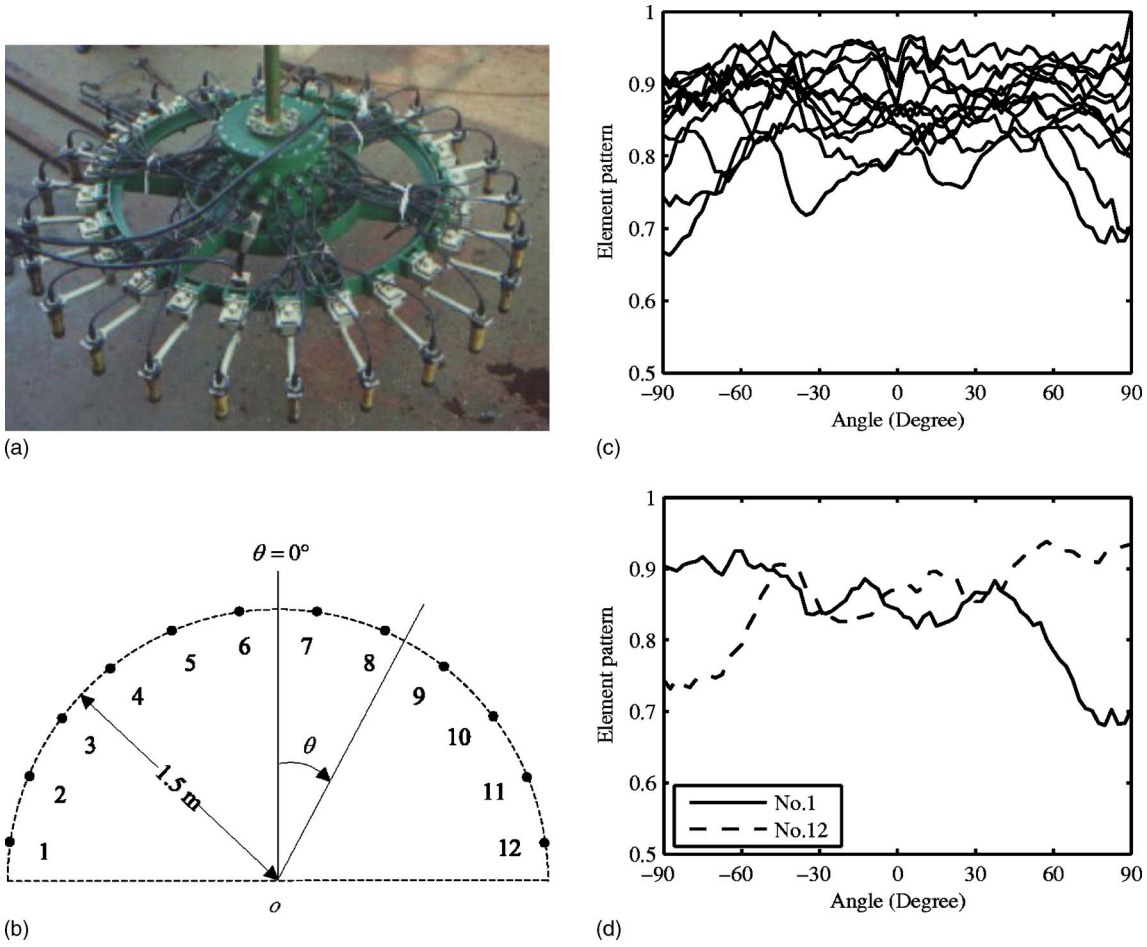


FIG. 4. (Color online) Measure of the actual array manifold vectors. (a) Photograph of the 24-element actual circular hydrophone array. (b) The construction drawing of the 12 elements on the sunny side that are used in designing the CMR beamformer with pointing direction $\phi_0=0°$. (c) The measured magnitude response of the 12-element semicircular array as a function of angle. (d) Element patterns of hydrophones, 1 and 12, from which the structural scattering and shadowing phenomena can be observed clearly.

sidelobe level $20\lg\sigma_{\mathrm{rmsSL}}$(dB), and the CPU time spent in solving the optimization problem using SeDuMi for the above-mentioned examples.

From Table II, it is seen that our method (18) has both smaller $\sigma_{\max}$ and σ_{rms} than the existing method (14) with the same σ_{maxSL}. Obviously, the new method outperforms the existing method. As expected, our method (25) has smaller σ_{rms} but a little larger $\sigma_{\max}$ than method (18) with the same σ_{maxSL}, and also spend shorter CPU time. Our method (30) has smaller σ_{rmsSL} than method (25) with the same σ_{rms}, and spend the shortest CPU time, although it has a little higher σ_{maxSL}. We observe that the CPU time of method (30) is reduced by 90 times to that of the existing method (14). Our method (38) can provide good interference rejection ability at the cost of little increased sidelobes and computation load when comparing with the method (30).

VI. EXPERIMENTAL RESULTS ON A PRACTICAL ARRAY

The discussion in the simulation has assumed that each element had an isotropic response. In practical scenarios, each element will have a nonisotropic pattern, and the element patterns are not identical from sensor to sensor. This can be caused by sensor sensitivity mismatch, channel gain and phase mismatch, element position perturbation, structural scattering, shadowing, mutual coupling between the sensors, and so on. As a result, the actual array manifold vector is not a constant function of angle.

In this section, the proposed algorithms are applied to real underwater acoustic data collected by an actual hydrophone array. A 24-element horizontal hydrophone uniform circular array with radius 1.5 m is shown in Fig. 4(a). The design band and the sampling frequency are $[f_l,f_u]=[960$ Hz, 1920 Hz$]$ and $f_s=6400$ Hz, respectively. The design band is decomposed into $K=16$ uniform frequency bins with a grid of 64 Hz. To design the CMR beamformer with pointing direction $\phi_0=0°$, only the 12 elements on the sunny side are used, as shown in Fig. 4(b).

The main goal of this experiment is to estimate the array manifold vector of the actual circular hydrophone array. A far-field omnidirectional source transmits a high SNR sinusoidal signal. By fixing the position of the source and rotating the circular array around its center axis, we recorded the received wave form at different bearing angles. Using FFT spectrum analysis method, the actual array manifold vector at each bearing angle can be estimated (see Ref. 21).

The actual element patterns (magnitude) of the 12-element semicircular array, as a function of the bearing angle

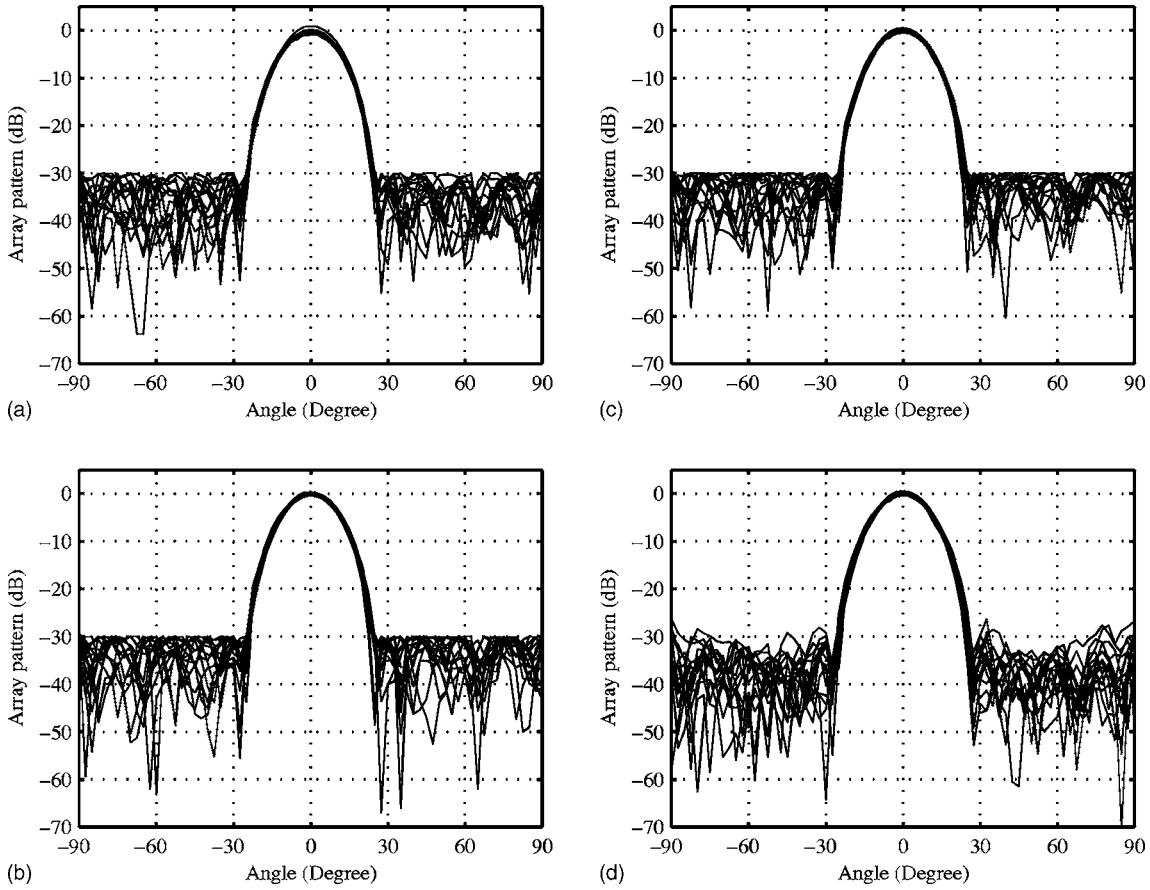


FIG. 5. Superposition of all 16 beampatterns for an actual hydrophone array calculated at 16 frequencies within the frequency band [960 Hz, 1920 Hz] using (a) the existing method (14), (b) peak sidelobes constrained minimax MSRV design (18), (c) peak sidelobes constrained least-square MSRV design (25), and (d) mean-squared MSRV constrained least-square sidelobes design (30), respectively. The resulting mainlobe synthesis accuracy and sidelobes by these methods are listed in Table III.

θ, can be plotted by using all measured array manifold vectors, as shown in Fig. 4(c). In the case of no array errors, all these curves should be identical and take a constant value 1.0 over all bearing angle, i.e., omnidirectional. The element patterns of two typical hydrophones, 1 and 2, are shown in Fig. 4(d), from which the structural scattering and shadowing phenomena can be observed clearly. The sensor has a higher amplitude response in the direction toward the source.

By changing the frequency of the transmitting sinusoidal signal, the measured array manifold vectors at the frequencies grids (960 Hz, 1024 Hz, ..., 1920 Hz) were obtained.

We design the CMR beamformer for the actual hydrophone array using the existing method (14), and the proposed methods (18), (25), and (30), respectively. In all these methods, the measured array manifold vectors are used instead of the ideal ones. For the method (14), the conventional mainlobe pattern at the lower bound frequency is chosen as the desired beampattern. The other parameters are kept the same as in the simulations except the mainlobe width is set to be 6° wider, since the aperture of this semicircular array is smaller than that of the linear array used in the simulations.

The beam response as a function of frequency and angle are calculated on a grid of points in frequency and angle. The resulting array patterns by the four methods are shown in Figs. 5(a)–5(d), respectively. The resulting mainlobe synthesized errors and resulting sidelobes of the four array patterns synthesized by the four methods are listed in Table III.

It can be seen that the conclusions from Table III by the actual array agree with those from Table II by the ideal array, i.e., the method (18) gets the smallest $\sigma_{\max}$, method (25) gets the smallest σ_{rms}, and method (30) gets the smallest σ_{rmsSL}. All three methods outperform the existing method (14). It is also noted that the resulting errors from the actual array are a little larger than those obtained from the ideal array, but in most cases acceptable.

VII. CONCLUSION

A new array pattern synthesis approach to designing broadband FIR beamformer with constant mainlobe re-

TABLE III. Performance comparison of the existing beamformer (14) and the proposed beamformers using different array pattern synthesis criteria. Experimental results.

Method	Array pattern	$20\lg\sigma_{\max}$ (dB)	$20\lg\sigma_{\mathrm{rms}}$ (dB)	$20\lg\sigma_{\mathrm{maxSL}}$ (dB)	$20\lg\sigma_{\mathrm{rmsSL}}$ (dB)
Eq. (14)	Fig. 5(a)	−17.5	−21.4	−30.0	−34.0
Eq. (18)	Fig. 5(b)	**−25.2**	−29.1	−30.0	−33.0
Eq. (25)	Fig. 5(c)	−22.3	**−30.9**	−30.0	−33.2
Eq. (30)	Fig. 5(d)	−20.7	−30.9	−26.3	**−35.9**

sponses and controlled sidelobes has been proposed. Our approach optimally designs the FIR filters of the beamformer by imposing constraints on the mainlobe spatial response variation over frequency and the sidelobes. Under this framework, several optimization criteria are presented and the corresponding convex second-order cone programming implementations are given. For our new approach, no preliminary desired beampattern is required and the beam responses' variation constraint is employed only in the mainlobe region while controlling the sidelobes, which improves the beamformer mainlobe synthesis accuracy. Our approach is applicable to arrays with an arbitrary geometry. Computational complexities of these criteria of our approach are also given, which is helpful for us to choose suitable criteria for various application requirements. We illustrate the proposed approach by applying it to an ideal uniform linear array and also to an actual semicircular hydrophone array. Results of simulations and experimental data processing show good performance of the proposed approach. The resulting beamformer by our new approach provides higher mainlobe synthesis accuracy as compared to the existing method with the same sidelobe level.

For simplicity, plane wave approximation is used in the simulations. For the case of multipath that occurs with vertical arrays and horizontal arrays near endfire in ocean waveguides, our approach also works if we replace the array manifold vectors with the matched-field replica vectors.

ACKNOWLEDGMENTS

S.Y. would like to thank Professor Jens M. Hovem and Professor Hefeng Dong at Norwegian University of Science and Technology for their discussions. This work was supported by the National Natural Science Foundation of China under Grant No. 60602055.

[1]B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE Trans. Acoust., Speech, Signal Process. **5**, 4–24 (1988).
[2]H. L. Van Trees, *Optimum Array Processing*, Detection, Estimation, and Modulation Theory, Part IV (Wiley, New York, 2002).
[3]J. Li and P. Stoica, *Robust Adaptive Beamforming* (Wiley, New York, 2005).
[4]T. Do-Hong and P. Russer, "Signal processing for wideband smart antenna array applications," IEEE Microw. Mag. **5**, 57–67 (2004).
[5]R. T. Compton, Jr., "The relationship between tapped delay-line and FFT processing in adaptive arrays," IEEE Trans. Antennas Propag. **36**, 15–26 (1988).
[6]O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935 (1972).
[7]L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **30**, 27–34 (1982).
[8]L. C. Godara, "Application of the fast Fourier transform to broadband beamforming," J. Acoust. Soc. Am. **98**, 230–240 (1995).
[9]S. Yan, C. Hou, X. Ma, and Y. Ma, "Convex optimization based time-domain broadband beamforming with sidelobe control," J. Acoust. Soc. Am. **121**, 46–49 (2007).
[10]T. Lee, "Efficient wideband source localization using beamforming invariance technique," IEEE Trans. Signal Process. **42**, 1376–1387 (1994).
[11]S. Yan and C. Hou, "Broadband DOA estimation using optimal array pattern synthesis technique," IEEE Antennas Wireless Propag. Lett. **5**, 88–90 (2006).
[12]D. B. Ward, Z. Ding, and R. A. Kennedy, "Broadband DOA estimation using frequency invariant beamforming," IEEE Trans. Signal Process. **46**, 1463–1469 (1998).
[13]S. Yan, "Broadband beamspace DOA estimation: Frequency-domain and time-domain processing approaches," EURASIP Journal on Advances in Signal Processing, see http://www.hindawi.com/journals/asp/. Last viewed 9/26/07.
[14]D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," J. Acoust. Soc. Am. **97**, 1023–1034 (1995).
[15]D. B. Ward, R. A. Kennedy, and R. C. Williamson, "FIR filter design for frequency invariant beamformers," IEEE Signal Process. Lett. **3**, 69–71 (1996).
[16]T. Sekiguchi and Y. Karasawa, "Wideband beamspace adaptive array utilizing FIR fan filters for multibeam forming," IEEE Trans. Signal Process. **48**, 277–284 (2000).
[17]Z. S. Wang, J. Li, P. Stoica, T. Nishida, and M. Sheplak, "Constant-beamwidth and constant-powerwidth wideband robust Capon beamformers for acoustic imaging," J. Acoust. Soc. Am. **116**, 1621–1631 (2004).
[18]S. C. Chan and H. H. Chen, "Uniform concentric circular arrays with frequency-invariant characteristics—Theory, design, adaptive beamforming and DOA estimation," IEEE Trans. Signal Process. **55**, 165–177 (2007).
[19]L. C. Parra, "Steerable frequency-invariant beamforming for arbitrary arrays," J. Acoust. Soc. Am. **119**, 3839–3847 (2006).
[20]S. Yan and Y. Ma, "Design of FIR beamformer with frequency invariant patterns via jointly optimizing spatial and frequency responses," in *Proceedings ICASSP 2005*, IEEE, Philadelphia, 2005, Vol. **4**, pp. 789–792.
[21]S. Yan, "Optimal design of FIR beamformer with frequency invariant patterns," Appl. Acoust. **67**, 511–528 (2006).
[22]H. Cox, R. M. Zeskind, and M. H. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **35**, 1365–1376 (1987).
[23]J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," Optim. Methods Software **11–12**, 625–653 (1999).
[24]H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," IEEE Trans. Signal Process. **45**, 526–532 (1997).
[25]S. Yan and Y. Ma, "A unified framework for designing FIR filters with arbitrary magnitude and phase response," Digit. Signal Process. **14**, 510–522 (2004).
[26]S. A. Vorobyov, A. B. Gershman, and Z. Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," IEEE Trans. Signal Process. **51**, 313–324 (2003).
[27]F. Wang, V. Balakrishnan, P. Y. Zhou, J. J. Chen, R. Yang, and C. Frank, "Optimal array pattern synthesis using semidefinite programming," IEEE Trans. Signal Process. **51**, 1172–1183 (2003).
[28]J. Liu, A. B. Gershman, Z. Q. Luo, and K. M. Wong, "Adaptive beamforming with sidelobe control: A second-order cone programming approach," IEEE Signal Process. Lett. **10**, 331–334 (2003).
[29]M. Lobo, L. Vandenberghe, S. Boyd, and H. Hebret, "Applications of second-order cone programming, Linear Algebra Applicat.," Numer. Linear Algebra Appl. **284**, 193–228 (1998).
[30]Y. Nesterov and A. Nemirovsky, "Interior-point polynomial methods in convex programming," in *Studies in Applied Mathematics* (SIAM, Philadelphia, 1994), Vol. **13**.

Noise reduction combining time-domain ε-filter and time-frequency ε-filter

Tomomi Abe[a)]
Major in Pure and Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Mitsuharu Matsumoto[b)] and Shuji Hashimoto[c)]
Department of Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

(Received 17 April 2007; revised 21 August 2007; accepted 23 August 2007)

A time-domain ε-filter (TD ε-filter) is a nonlinear filter that can reduce noise while preserving a signal that varies drastically, such as a speech signal. Although the filter design is simple, it can effectively reduce noise. It is applicable not only to stationary noise but also to nonstationary noise. It cannot, however, be applied when the amplitude of noise is relatively large. This paper introduces an advanced method for noise reduction that applies an ε-filter to complex spectra, namely a time-frequency ε-filter (TF ε-filter). This paper also introduces noise reduction combining a TD ε-filter and a TF ε-filter. An advanced method called a variable time-frequency ε-filter is also proposed. First, the algorithm of the TD ε-filter is explained to clarify the problem. Then, the algorithms of the proposed methods are explained. By utilizing an ε-filter in the frequency domain, the proposed method can reduce not only noise that has a relatively small amplitude but also noise that has a relatively large amplitude. Experimental results are also given to demonstrate the performance of the proposed methods in comparison to the results of some conventional methods. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2785038]

PACS number(s): 43.60.Hj, 43.60.Wy, 43.60.Ac [EJS] Pages: 2697–2705

I. INTRODUCTION

Noise reduction is an important aspect of human hearing and understanding of sounds. There are many studies regarding noise reduction. A microphone array is one useful approach to reducing noise. It can reduce noise utilizing the phase difference or amplitude difference of each microphone.[1–3] Many methods utilizing microphone arrays have been reported, such as the delay-sum type microphone array[4] and the adaptive microphone array.[5,6] However, as a microphone array utilizes multiple inputs, it requires a large system size. It also needs a precise estimation of the position and directivity of the microphones. Independent component analysis (ICA) is also a well-known method for extracting the objective sound.[7,8] ICA can separate mixed sounds utilizing the statistical independence of the sounds. However, calculation cost is large because higher order statistics are used. It also needs multiple signals such as a microphone array to separate sounds. Some authors have reported other methods of extracting objective sounds called sparseness approaches.[9–11] Sparseness means that the sources rarely overlap in the frequency domain. Under this assumption, it is possible to extract each signal using a time-frequency binary mask. However, due to this assumption, sparseness approaches cannot handle sounds which overlap in the frequency domain. They also need multiple signals to extract sounds.

If we can effectively reduce the noise of a monaural sound, various applications can be realized. It will also be easy to miniaturize system size because only one signal is required. Spectral subtraction (SS) is a well-known approach to reducing the noise of a monaural sound.[12–14] (SS) can reduce noise effectively in spite of the simplicity of its procedure. However, it can only handle stationary noises and noise needs to be estimated in advance. Some authors have reported a model-based approach to noise reduction.[15] In this approach, we can extract the objective sound by determining the sound model in advance. However, as with SS, this approach is not applicable to signals with unknown noise. There are some approaches that utilize a comb filter.[16] In these approaches, the pitch of the speech signal is estimated, and noise is reduced with a comb filter. However, estimation error results in the degradation of noise reduction. Although noise reduction utilizing sophisticated filters such as the Kalman filter and the iterative Wiener filter[17–20] has also been reported, the calculation cost to estimate signal model is large. In image processing, a median filter is often utilized for noise reduction.[21,22] However, when we apply a median filter to a speech signal, the attack and decay of the speech signal are also smoothed because speech signals often vary drastically, unlike images.

To solve these problems, we pay attention to a filter called the ε-filter. A conventional ε-filter is a nonlinear filter in the time domain, which can reduce noise while preserving a speech signal.[23,24] In spite of its simple design, a conventional ε-filter has some desirable features for noise reduction. It does not need an advance noise estimate. It can reduce not

[a)]Electronic mail: tomomi@shalab.phys.waseda.ac.jp
[b)]Electronic mail: matsu@shalab.phys.waseda.ac.jp
[c)]Electronic mail: shuji@waseda.jp

only stationary noise, but also nonstationary noise. It is easy to design and calculation costs are small. However, in principle, it can reduce only small amplitude noise. We therefore aim to develop an improved ε-filter that can handle not only small stationary noise, but also large nonstationary noise.

To realize this aim, we paid attention to the characteristics of the signal and noise in the frequency domain, and we extended the ε-filter in the frequency domain. In this paper, we propose an algorithm for noise reduction applying an ε-filter to the time change of a complex spectrum in the frequency domain, namely a time-frequency ε-filter (TF ε-filter). We also propose noise reduction combining a conventional ε-filter, namely a time-domain ε-filter (TD ε-filter), and a TF ε-filter.

Utilizing TF ε-filter, we can reduce not only noise that has relatively small amplitude but also noise that has relatively large amplitude. The essential difference is caused by the different properties of the wave in the time domain and the complex spectra in the frequency domain. The proposed method can also handle not only stationary noise but also nonstationary noise. It need not estimate the noise as well as the TD ε-filter. In Sec. II, we first describe the algorithm of the conventional ε-filter and explain the problems of the conventional ε-filter. Second, we describe the algorithm of the TF ε-filter. We also propose a method for noise reduction combining a TF ε-filter and a TD ε-filter. An advanced method called a variable time-frequency ε-filter (VTF ε-filter) is also introduced. In Sec. III, we show our experimental results. To compare the performance of the proposed methods, we also show experimental results utilizing SS and a component separating ε-filter (CS ε-filter),[24] which is an improved TD ε-filter. Moreover, to show the simplicity of the proposed method in comparison to other existing sophisticated filters, we compare the calculation cost of the proposed method to that of the iterative Wiener filter. A discussion and conclusions follow in Sec. IV.

II. A NOISE REDUCTION METHOD USING AN ε-FILTER

A. Time-frequency ε-filter

To clarify the problems of a TD ε-filter, we first explain the TD ε-filter algorithm. Let us define $x(k)$ as the input signal at time k. Let us also define $y(k)$ as output signal of the ε-filter at time k as follows:

$$y(k) = x(k) + \sum_{i=-N}^{M} a(i) F(x(\kappa + i) - x(k)), \tag{1}$$

where $a(i)$ represents the filter coefficient. $a(i)$ is usually constrained as follows:

$$\sum_{i=-N}^{M} a(i) = 1. \tag{2}$$

The window size of the ε-filter is $N+M+1$. $F(x)$ is the nonlinear function described as follows:

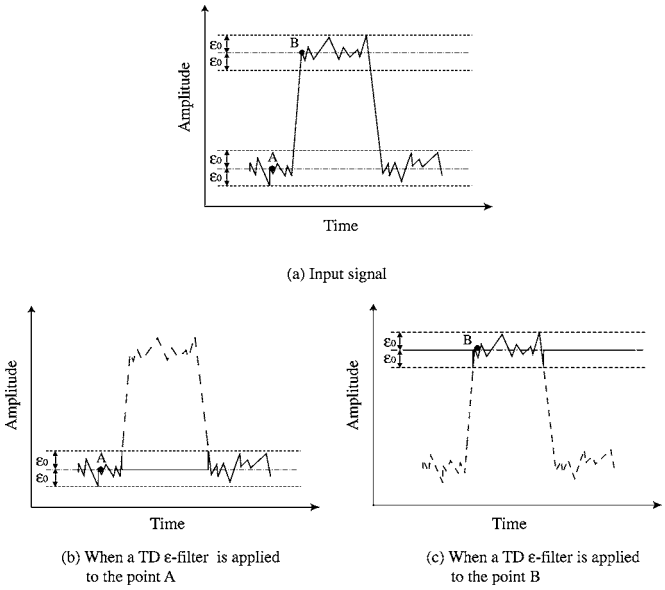


FIG. 1. The basic concept of a TD ε-filter.

$$|F(x)| \leqslant \varepsilon_0 \ (-\infty \leqslant x \leqslant \infty), \tag{3}$$

where ε_0 is a constant. This method can reduce small amplitude noise while preserving the speech signal. For example, we can set the nonlinear function $F(x)$ as follows:

$$F(x) = \begin{cases} x & (-\varepsilon_0 \leqslant x \leqslant \varepsilon_0) \\ 0 & (\text{else}). \end{cases} \tag{4}$$

Figure 1 shows the basic concept of an ε-filter when Eq. (4) is utilized as $F(x)$. Figure 1(a) shows the wave form of the input signal. Executing the ε-filter at point A in Fig. 1(a), we first replace all the points where the distance from A is more than ε_0 by the value of point A. We then add the signals in the same window. Figure 1(b) shows the basic concept of this procedure. In Fig.1(b), the dotted line represents the points where the distance from A is more than ε_0. In Fig. 1(b), the continuous line represents the values replaced through this procedure. As a result, if the points are far from A, the points are ignored. On the other hand, if the points are close to A, the points are smoothed. Due to this procedure, the ε-filter reduces noise while preserving the precipitous attack and decay of the speech signal. In the same way, by executing the ε-filter at point B in Fig. 1(a), we replace all the points where the distance from B is more than ε_0 by the value of the point B. The points are ignored if they are far from B, while the points are smoothed if the points are close to B. Consequently, we can reduce small amplitude noise near by the processed point while preserving the speech signal.

An ε-filter can reduce small amplitude noise in the time domain. However, due to the procedure, it is not applicable to large amplitude noise. To solve this problem, we pay attention to the distribution of the speech signal and the noise in the frequency domain. We make the following assumptions regarding the sound sources:

Assumption 1. The speech signal varies more drastically than the noise signal in the time-frequency domain.

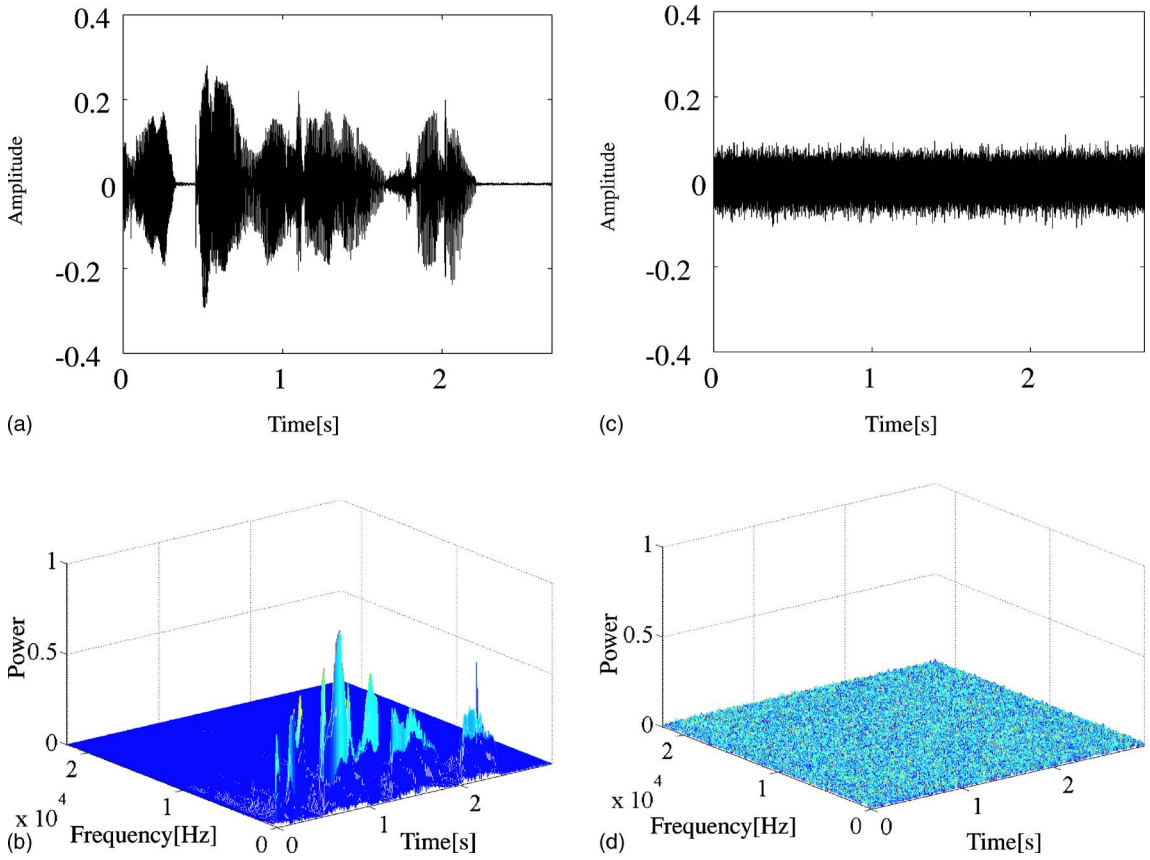


FIG. 2. (Color online) A speech signal or noise signal in the time and frequency domains. (a) Speech signal (in the time domain). (b) Speech signal (in the frequency domain). (c) Noise signal (in the time domain). (d) Noise signal (in the frequency domain).

Assumption 2. The noise signal distributes more widely than the speech signal in the frequency domain.

Figure 2 depicts the speech signal and the white noise signal in the time and frequency domains. As shown in Fig. 2, assumptions 1 and 2 are fulfilled in the case of various noise such as white noise and natural noise such as the sound of a cooling fan. In Figs. 2(b) and 2(d), the power is normalized using the maximal power of the speech signal. When we consider frequency bins where there are signals, the ratio of noise power to signal power is smaller than the ratio of noise amplitude to signal amplitude in the time domain. Our aim is to utilize this feature to apply an ε-filter to large-amplitude noise.

Figure 3 illustrates the proposed method with a block diagram. As shown in Fig. 3(1), we first transform the input signal $x(k)$ to the complex amplitude $X(\kappa,\omega)$ by short-term Fourier transformation (STFT) as follows:

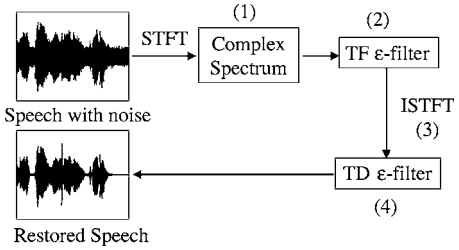


FIG. 3. Block diagram of the proposed method.

$$X(\kappa,\omega)=\int_{-\infty}^{\infty}x(k)W(k-\kappa)e^{-i\omega k}dk, \tag{5}$$

where W and κ represent the window function and the time in the frequency domain, respectively. Next we execute a TF ε-filter, which is an ε-filter executed in the time-frequency domain, as shown in Fig. 3(2). In this procedure, $V(\kappa,\omega)$ is obtained as follows:

$$V(\kappa,\omega)=\sum_{i=-N}^{M}a(i)X'(\kappa+i,\omega), \tag{6}$$

where

$$X'(\kappa+i,\omega)=\begin{cases}X(\kappa,\omega) & (\|X(\kappa,\omega)|-|X(\kappa+i,\omega)\|>\varepsilon_F)\\ X(\kappa+i,\omega) & (\|X(\kappa,\omega)|-|X(\kappa+i,\omega)\|\leqslant\varepsilon_F)\end{cases} \tag{7}$$

and ε_F is a constant. Figure 4 illustrates the differences in performance when we apply a TF ε-filter to the speech signal and the noise. The horizontal axis and the vertical axis represent the real axis and the imaginary axis, respectively. In Fig. 4, an asterisk and a cross represent the processed point and the other signal points in the same window as the processed point, respectively. Point A in Fig. 4(a) and point B in Fig. 4(b) represent the complex amplitude of the processed point. A' and B' represent the complex amplitudes of the outputs when we apply the TF ε-filter to points A and B,

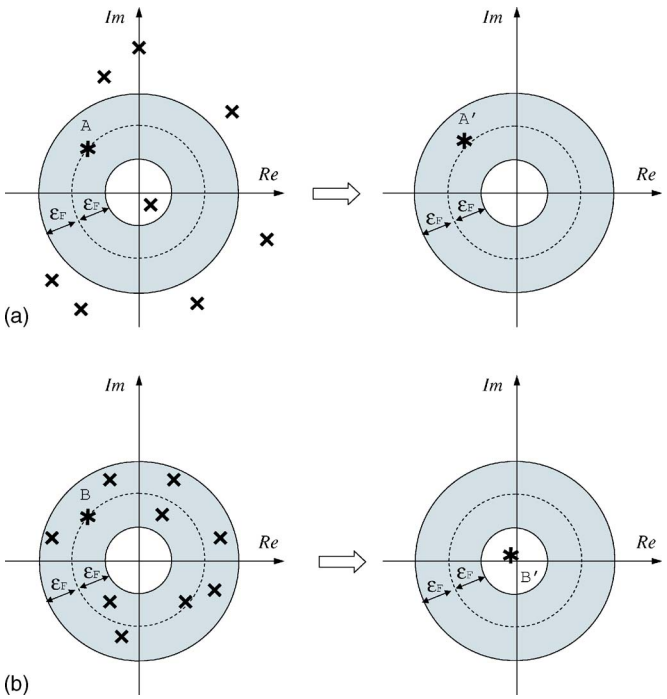


FIG. 4. (Color online) Differences in performance when a TF ε-filter is applied to the speech signal and noise. (a) Speech signal. (b) Noise signal.

respectively. Executing the TF ε-filter, we first replace the complex amplitude of the signal outside of the shadow area by that of A. We then add the complex spectra of all the points in the same window. Due to handling complex spectra, when we have many signals that have similar amplitudes but different phases, the real part and imaginary part cancel each other. In other words, even if the amplitude of the noise is large, the noise is reduced because they cancel each other. Note that the noise is reduced not only when the amplitude of the noise is small but also when the amplitude of the noise is large because of this procedure. Figure 4(a) represents the basic concept in the case that the power varies drastically, like a speech signal. When we consider a signal whose power varies drastically, the difference between the absolute value of A and that of the other signals is large as shown in Figure 4(a). For this reason, many signals in the same window as point A are replaced by A. As a result, when we handle the speech signal, the complex amplitude of the processed point is intact. Figure 4(b) represents the basic concept in the case that the power does not vary so much, like a noise signal. When we consider a noise signal, the difference between the absolute value of B and that of the other signals is relatively small compared with the speech signal. Hence, few signals in the same window as point B are replaced by B. In other words, when handling noise, the complex amplitude of the processed point becomes smaller when the TF ε-filter is applied. Based on these aspects, we can reduce noise while preserving the signal by setting ε_F appropriately. Hence, the TF ε-filter is effective when the power of the noise to signal is large. Additionally, under assumption 2, the TF ε-filter becomes more effective. When assumption 2 is satisfied, the variation of the noise to the signal in the frequency domain becomes smaller than that in the time domain. As a consequence, even if the noise varies drastically in the time domain, the ε-filter can be applied in the frequency domain. Next, we transform $V(\kappa,\omega)$ to $v(k)$ by inverse STFT as shown in Fig. 3(3). To reduce the remaining noise, we additionally apply the ε-filter in the time domain to $v(k)$ as shown in Fig. 3(4). Note that the ε-filter in the time domain can be utilized because large amplitude noise has already been reduced in the previous procedure. The output $y(k)$ can be obtained as follows:

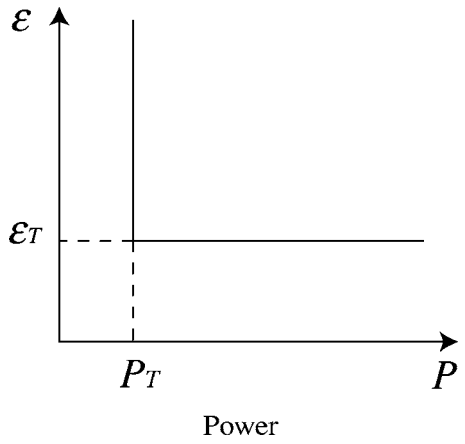


FIG. 5. Relationship between the power of the input signal P and ε.

$$y(k) = \sum_{i=-N}^{M} a(i)v'(k+i), \tag{8}$$

where

$$v'(k+i) = \begin{cases} v(k) & (|v(k+i) - v(k)| > \varepsilon_t) \\ v(k+i) & (|v(k+i) - v(k)| \leqslant \varepsilon_t). \end{cases} \tag{9}$$

We label this method of combining a TF ε-filter and a TD ε-filter "proposed method 1."

B. Variable time frequency ε-filter

The TF ε-filter described earlier can reduce noise effectively. However, the signal is also distorted when the power change of the signal is small. To solve this problem, we make the following assumption, in addition to assumptions 1 and 2:

Assumption 3. The power of the speech signal is larger than that of the noise in the frequency domain where the speech signal exists.

Under assumption 3, the power of the frequency components is large when there is not only noise but also a speech signal, while the power of the frequency components is small when there is not a speech signal but there is noise. Therefore, we regard ε as function of P. $\varepsilon(P)$ is defined as follows:

$$\forall P_1, P_2 \; P_1 \leqslant P_2 \rightarrow \varepsilon(P_1) \geqslant \varepsilon(P_2) \quad (0 < P \leqslant \infty), \tag{10}$$

where

$$\lim_{P \rightarrow \infty} \varepsilon(P)$$

represents the linear filter. In this paper, as shown in Fig. 5,

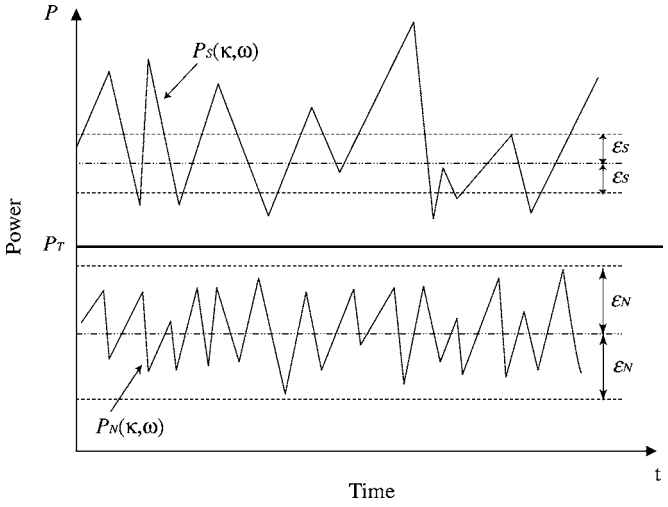


FIG. 6. Relationship among speech signal, noise signal, and P_T.

we set $\varepsilon(P)$ as follows:

$$\varepsilon(P) = \begin{cases} \varepsilon_T & (P > P_T) \\ \infty & (P \leqslant P_T), \end{cases} \tag{11}$$

where ε_T is a constant. Figure 6 illustrates the basic concept of the relationship among P_S, P_N, and P_T when assumption 3 is satisfied. P_S and P_N represent the power of the speech signal and the power of the noise signal. Let us define $\min(P_S(\kappa,\omega))$ and $\max(P_N(\kappa,\omega))$ as the minimal power of the speech signal and the maximal power of the noise signal at each frequency ω, respectively. Assume that $\min(P_S(\kappa,\omega))$ and $\max(P_N(\kappa,\omega))$ are constrained as follows:

$$\max(P_N(\kappa,\omega)) < \min(P_S(\kappa,\omega)). \tag{12}$$

Under this assumption, this method becomes effective by setting P_T as follows:

$$\max(P_N(\kappa,\omega)) < P_T < \min(P_S(\kappa,\omega)). \tag{13}$$

In this paper, we label this type of ε-filter as a VTF ε-filter. This method, combining a VTF ε-filter and a TD ε-filter, is called "proposed method 2."

III. EXPERIMENT

A. Experimental conditions

We conducted the experiments utilizing a speech signal with a noise signal. As the sound source, we utilized "Japanese Newspaper Article Sentences" edited by the Acoustical Society of Japan. We also prepared three kinds of noise signals: stationary noise, nonstationary noise, and natural noise. The signal and the noise are mixed in the computer. To compare the effectiveness of the proposed methods compared to other methods, we conducted the experiments utilizing four methods, such as a component separating (CS) ε-filter, spectral subtraction (SS), proposed method 1, and proposed method 2. A component separating ε-filter (CS ε-filter) is an improved TD ε-filter. In this method, we apply the ε-filter in two stages to make the TD ε-filter more effective. Utilizing a CS ε-filter, we first apply a TD ε-filter to the signal with the noise to extract the primary shape of the signal. We then apply a TD ε-filter to the difference between the input signal and the output of the first TD ε-filter to extract the detailed shape of the signal. Finally, we summate the outputs of the TD ε-filters. Table I shows the value of common parameters for all the experiments. The details of implementation are as follows: We first transformed the signal in the complex spectrum by utilizing STFT. Block size and hop size were set to 512 and 256 as shown in Table I, respectively. A Hanning window is utilized as the window function. A TF ε-filter or VTF ε-filter was then executed in the frequency domain. The window size of the TF ε-filter or VTF ε-filter was set to 101. The window was shifted point by point. After the procedure, we transformed the output of the TF ε-filter or VTF ε-filter into the signal in the time domain by ISTFT. The TD ε-filter is executed in the time domain. The window size of the TD ε-filter was set to 7. We utilized a computer with an Intel Pentium M processor 1.73 GHz CPU. All programs were implemented by MATLAB. To evaluate the performance of the proposed methods quantitatively, we utilized the noise reduction ratio (NRR) signal-to-distortion ratio (SDR). Signal-to-noise ratio (SNR) is defined as follows:

TABLE I. Common parameters.

Parameter	Value
Sampling frequency	44 100
STFT Block size	512
Hop size	256
Window function	Hanning window
Window size of TF ε-filter	101
Window size of VTF ε-filter	101
Window size of TD ε-filter	7

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sum_{k=1}^{L} s(k)^2}{\sum_{k=1}^{L} n(k)^2} \right), \tag{14}$$

where $s(k)$, $n(k)$, and L represent the speech signal at time k, the noise signal at time k, and the length of signal, respectively.

NRR is defined as follows:

$$\mathrm{NRR} = \mathrm{SNR}_{\mathrm{out}} - \mathrm{SNR}_{\mathrm{in}}, \tag{15}$$

where $\mathrm{SNR}_{\mathrm{out}}$ and $\mathrm{SNR}_{\mathrm{in}}$ represent the SNR after the process and before the process, respectively. SDR can be represented as follows:

$$\mathrm{SDR} = 10 \log_{10} \left(\frac{\sum_{k=1}^{L} s_{\mathrm{in}}(k)^2}{\sum_{k=1}^{L} (s_{\mathrm{in}}(k) - s_{\mathrm{out}}(k))^2} \right), \tag{16}$$

where $s_{\mathrm{in}}(k)$ and $s_{\mathrm{out}}(k)$ represent the input signal at time k and the output signal at time k.

NRR represents how much the method reduces the noise. SDR represents how much the signal is distorted by reducing the noise.

TABLE II. SNR and SDR when a signal with stationary noise is utilized.

	SNR	SDR
Input signal	8.26	···
CS ε-filter	18.5	17.7
SS	15.9	16.4
Proposed method 1	29.0	17.4
Proposed method 2	30.8	17.9

B. Experimental results in the case of stationary noise

We first conducted the experiment utilizing a signal with stationary noise. We prepared a speech signal and white noise as the signal and the stationary noise, respectively. We set ε_F and ε_t in proposed method 1 to 0.6 and 0.01, respectively. In addition, we set ε_T, P_T, and ε_t in proposed method 2 to 0.4, 0.5, and 0.02, respectively. Table II shows the results of the experiments for stationary noise. The parameters of the CS ε-filter and SS are set optimally.

As shown in Table II, the SNR of the proposed methods is superior to other methods by 12 dB. SDR of the proposed methods is also better than that of the other methods.

It should be noted that there is a tradeoff between noise reduction performance and distortion. However, in the proposed method, noise reduction performance is relatively large, while distortion is relatively small compared to conventional methods. In this paper, to clarify the effectiveness of the proposed method compared to conventional methods, noise reduction performance and distortion were compromised such that the results were superior to the best results of the conventional methods. This means that the proposed method is at least potentially more effective than conventional methods.

Figure 7(a) depicts the wave form of the original signal. Figure 7(b) depicts the wave form of the signal with stationary noise. Figures 7(c)–7(f) depict the waveform of the output of CS ε-filter, SS, proposed method 1, and proposed method 2, respectively. As shown in Fig. 7, the proposed methods can reduce noise effectively, while noise remains when the CS ε-filter and SS are applied.

C. Experimental results in case of nonstationary noise

The experiment was conducted using a signal with nonstationary noise. We used the same speech signal as in Sec. III B. We prepared white noise with an amplitude that sometimes varied. We set ε_F and ε_t in proposed method 1 to 0.9 and 0.03, respectively. In addition, we set ε_T, P_T, and ε_t in proposed method 2 to 0.4, 0.5, and 0.03, respectively. Table III shows the results of the experiment when we utilized nonstationary noise. The parameters of the CS ε-filter and SS are optimally set. The SNR in the proposed methods is superior to other methods by 12 dB, while SDR in the proposed methods is almost the same as that in the other methods. In other words, the proposed methods reduce noise while preserving the signal, even when the noise is nonstationary. Figure 8(a) depicts the wave form of the original signal. Figure 8(b) depicts the wave form of the signal with

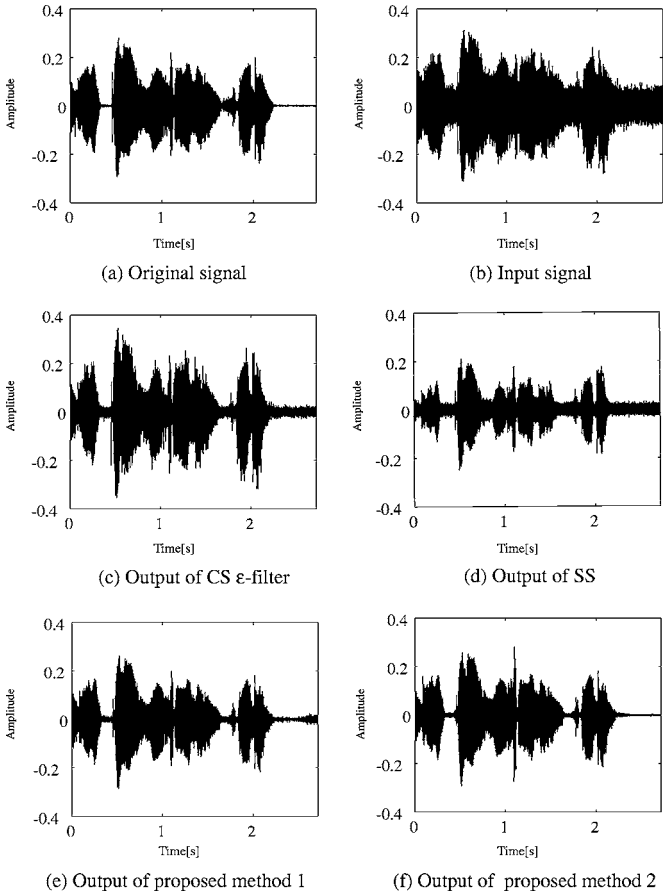


FIG. 7. Experimental results when a signal with stationary noise is utilized.

nonstationary noise. Figure 8(c)–8(f) depict the wave form of the output of the CS ε-filter, SS, proposed method 1, and proposed method 2, respectively. As shown in Fig. 8, when we utilize the proposed method, the noise can be reduced in spite of nonstationary noise.

D. Experimental results in case of natural noise

To evaluate the performance of the proposed methods for natural noise, we conducted the experiment utilizing a speech signal and noise generated from the cooling fan of a personal computer. We set ε_F and ε_t in proposed method 1 to 0.7 and 0.6, respectively. In addition, we set ε_T, P_T, and ε_t in proposed method 2 to 0.7, 0.6, and 0.02, respectively. Table IV shows the results of the experiment when we utilized the signal with natural noise. The parameters of the CS ε-filter and SS are set to their optimum value. The distribution of natural noise in the frequency domain is as narrow as that of

TABLE III. SNR and SDR when a signal with nonstationary noise is utilized.

	SNR	SDR
Input signal	8.27	···
CS ε-filter	18.4	15.9
SS	12.5	15.0
Proposed method 1	30.6	14.6
Proposed method 2	30.4	15.1

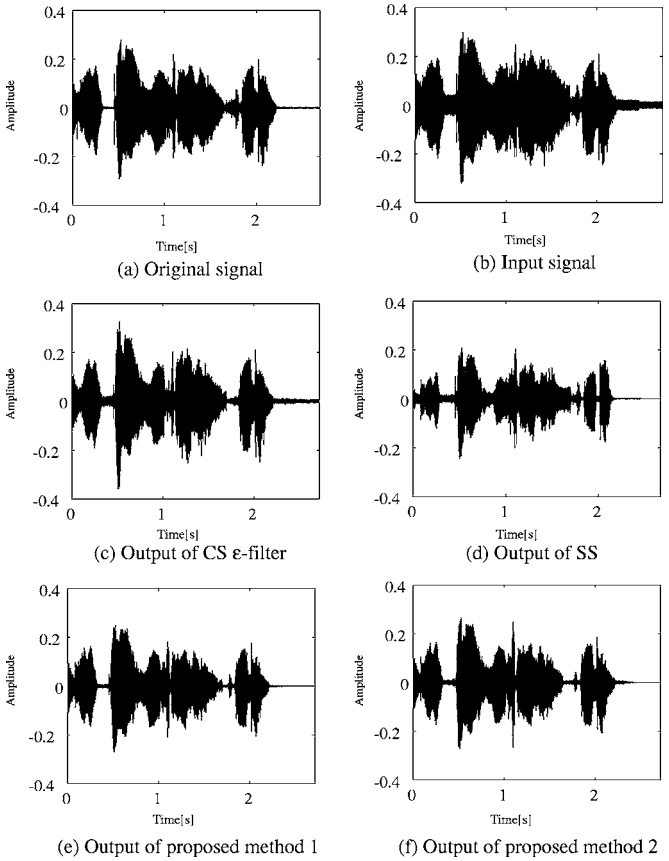


FIG. 8. Experimental results when a signal with nonstationary noise is utilized.

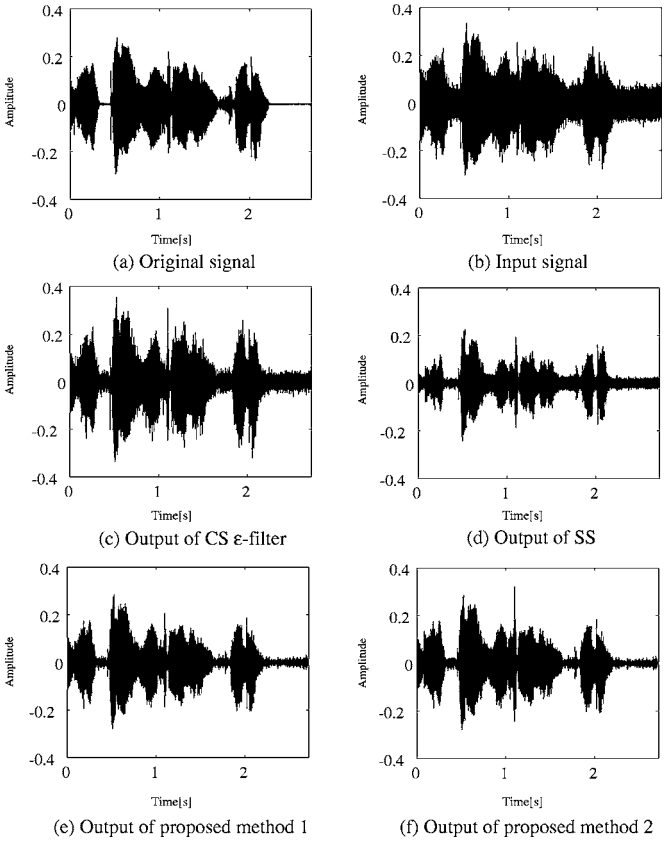


FIG. 9. Experimental results when a signal with natural noise is utilized.

the speech signal, unlike white noise. However, as shown in Table IV, the proposed methods reduce noise more effectively than other methods. Figure 9(a) depicts the wave form of the original signal. Figure 9(b) depicts the wave form of the signal with natural noise. Figures 9(c)–9(f) depict the wave form of the output of the CS ε-filter, SS, proposed method 1, and proposed method 2, respectively. The performance of the proposed methods is better than that of conventional methods as shown in Fig. 9.

E. Robustness for various SNR

We also conducted the experiments utilizing the signal with various noise to confirm that the proposed methods can be applied not only to large noise but also to small noise. We use three signals whose SNR is 0, 5, 10, 15, and 20 dB, respectively. Figure 10 shows the experimental results on NRR. As shown in Fig. 10, the NRR of the CS ε-filter or SS is only about 10 dB, while proposed methods 1 and 2 can reduce much more noise than a CS ε-filter or SS. Figure 11 shows the experimental results for the SDR, NRR and SDR of proposed methods 1 and 2 are better than those of a CS ε-filter or SS in all the cases.

TABLE IV. SNR and SDR when a signal with natural noise is utilized.

	SNR	SDR
Input signal	8.27	···
CS ε-filter	12.2	17.8
SS	15.5	15.0
Proposed method 1	19.1	15.6
Proposed method 2	19.5	15.6

F. Robustness for the change of ε_T and P_T

We also conducted the experiments to confirm robustness for the change of ε_T and P_T of the proposed method, because it seems more useful if we can reduce the noise sufficiently even when we do not set ε_T and P_T most appropriately. We utilized a signal with the stationary noise utilized in Sec. III B. The ε_T is changed from 0.1 to 0.9 by 0.1. The P_T is also changed from 0.3 to 1.5 by 0.3. Figure 12 shows the experimental results for NRR. As shown in Fig.

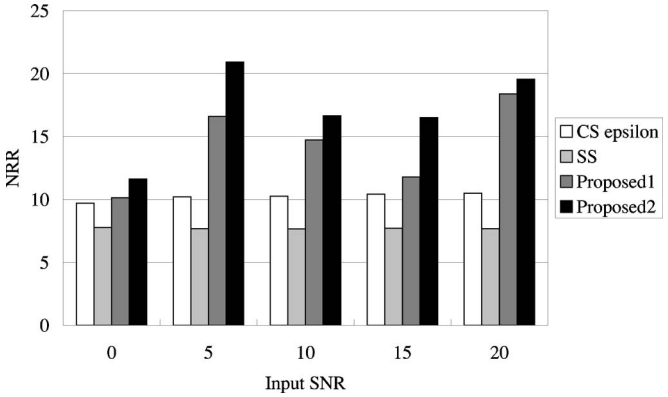


FIG. 10. Experimental results for NRR.

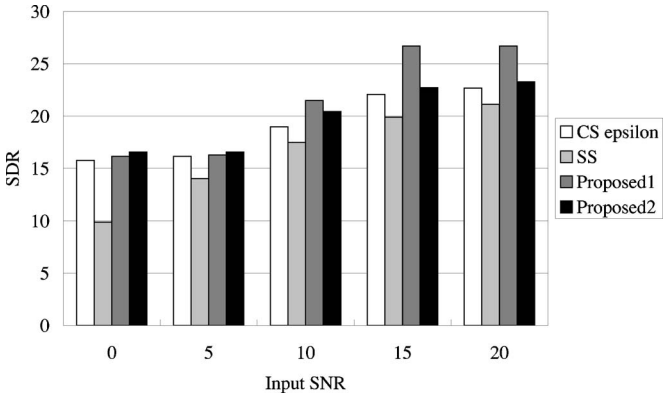


FIG. 11. Experimental results for the SDR of the output.

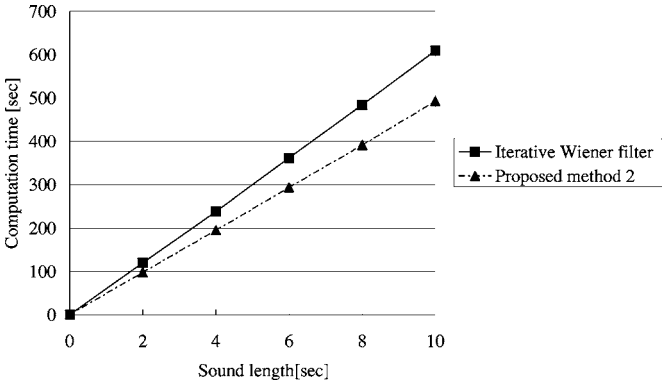


FIG. 13. Comparison of the calculation cost of the proposed method and the iterative Wiener filter.

12, the NRR hardly changes when the P_T in larger than 0.3. Even if ε_T and P_T change variously, NRR is more than 10 dB for any ε_T or P_T, while in the conventional methods it is less than 10 dB. In other words, the results of the proposed method were superior to those of the conventional methods even if ε_T and P_T change to some extent.

G. Experimental results on calculating cost

We also conducted the experiments to confirm that the calculation cost of the proposed method is smaller than that of the more sophisticated filter. As the filter, we utilized an iterative Wiener filter[20] that estimates the noise spectrum to perform adaptive Wiener filtering. We prepared a signal with stationary noise whose SNR is 0 dB. The iteration count of the iterative Wiener filter was set to 7 in order to obtain the same auditory property as the proposed method. Figure 13 depicts the relationship between computation time and sound length for proposed method 2 and the iterative Wiener filter. As shown in Fig. 13, the calculation cost of proposed method 2 is smaller than that of the iterative Wiener filter.

IV. DISCUSSION AND CONCLUSION

In this paper, we introduced an algorithm for noise reduction that applies an ε-filter to the time change of a complex spectrum. We also proposed noise reduction combining a TD ε-filter and a TF ε-filter. Although the approach is simple, the proposed methods can reduce not only stationary noise that has small amplitude but also nonstationary noise that has relatively large amplitude. The essential difference is caused by the different properties of the wave in the time domain and the complex spectra in the frequency domain. The reason the proposed method can cope with large amplitude noise is that the signal is handled as a complex spectra. When we consider the frequency bins where there are signals, the ratio of noise power to signal power is smaller than the ratio of noise amplitude to signal amplitude in the time domain. Moreover, even when noise power is large, the complex spectra are canceled by applying ε-filter to the complex spectra in the frequency domain. In other words, the noise power itself does not have much effect on the effectiveness of the proposed method. The experimental results show that the proposed methods may be applied to various kinds of noise. The proposed methods can reduce large noise better than conventional methods such as CS ε-filter and SS. Although in our method, the signals are processed in the frequency domain, musical noise was relatively small compared to spectral subtraction even when we only utilized a TF ε-filter. Moreover, when we combined a TF ε-filter and TD ε-filter, musical noise was more inconspicuous aurally, because impulsive noise in the frequency domain is transformed into a wave in the time domain and the TD ε-filter smoothes the musical noise in the time domain. Furthermore, the calculation cost of the proposed method is less than that of the iterative Wiener filter. For future works, we would like to determine each parameter adaptively. To preserve the intelligibility of the signal, we also aim at developing an advanced method by utilizing speech formant.

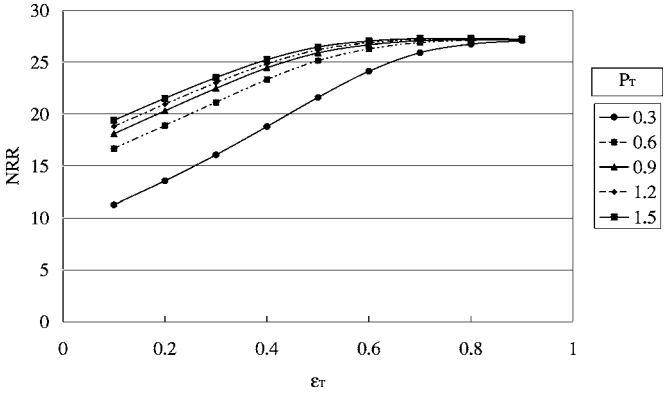


FIG. 12. Alteration of SNR due to change of the ε_T.

ACKNOWLEDGMENTS

This work was supported in part by "Establishment of Consolidated Research Institute for Advanced Science and Medical Care," Encouraging Development Strategic Research Centers Program, the Special Coordination Funds for Promoting Science and Technology, MEXT, and the 21st Century Center of Excellence Program, "The innovative research on symbiosis technologies for human and robots in the elderly dominated society," Waseda University. This research is also supported by CREST project "Foundation of technology supporting the creation of digital media contents" of JST.

[1]K. Sasaki and K. Hirata, "3D-localization of a stationary random acoustic source in nearfield by using 3 point-detectors," Trans. SICE **34**, 1329–1337 (1998).
[2]Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by the closely located four point microphone method," J. Acoust. Soc. Jpn. **10**, 101–110 (1990).
[3]M. Mastumoto and S. Hashimoto, "A minimized adaptive microphone array under directional constraint utilizing aggregated microphones," J. Acoust. Soc. Am. **119**, 352–359 (2006).
[4]K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, Germany, 1997, pp. 215–218.
[5]Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**, 1391–1400 (1986).
[6]K. Takao, M. Fujita, and T. Nishi, "An adaptive antenna array under directional constraint," IEEE Trans. Antennas Propag. **24**, 662–669 (1976).
[7]A. J. Bell and T. J. Sejowski, "An information maximization approach to blind seperation and blind deconvolution," Neural Comput. **7**, 1129–1159 (1995).
[8]H. Saruwatari, S. Kurita, and K. Takeda, "Blind source seperation combining frequency-domain ICA and beamforming," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, USA, 2001, pp. 146–157.
[9]T. Ihara, M. Handa, T. Nagai, and A. Kurematsu, "Multi-channel speech seperation and localization by frequency assignment," IEICE Trans. Fundamentals **J86-A**, 998–1009 (2003).
[10]S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, Florida, USA, 2002, pp. 529–532.
[11]M. Aoki, Y. Yamaguchi, K. Furuya, and A. Kataoka, "Modifying SAFIA: seperation of the target source close to the microphones and noise sources far from the microphones," IEICE Trans. Fundamentals **J88-A**, 468–479 (2005).
[12]S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**, 113–120 (1979).
[13]K. Yamashita, S. Ogata, and T. Shimamura, "Improved spectral subtraction utilizing iterative processing," IEICE Trans. Fundamentals **J88-A**, 1246–1257 (2005).
[14]J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 354–358 (1978).
[15]P. Daniel, W. Ellis, and R. Weiss, "Model-based monaural source separation using a vector-quantized phase-vocoder representation," *Proceedings of the IEEE International on Acoustics, Speech, and Signal Processing*, Toulouse, France, 2006, pp. V–957–960.
[16]J. S. Lim, A. V. Oppenheim, and L. D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 419–423 (1978).
[17]R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME, J. Appl. Mech. **82**, 35–45 (1960).
[18]M. Fujimoto and Y. Ariki, "Speech recognition under noisy environments using speech signal estimation method based on Kalman filter," IEICE Trans. Inf. Syst. **J85-D-II**, 1–11 (2002).
[19]S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 2nd ed. (Wiley, New York, 2000).
[20]J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 197–210 (1978).
[21]M. Meguro, A. Taguchi, and N. Hamada, "Data-dependent weighted median filtering with robust motion information for restoring image sequence degraded by additive Gaussian and impulsive noise," IEICE Trans. Fundamentals **E84-A**, 432–440 (2001).
[22]X. Wang and D. Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images," IEEE Trans. Circuits Syst., II: Analog Digital Signal Process. **46**, 78–80 (2002).
[23]H. Harashima, K. Odajima, Y. Shishikui, and H. Miyakawa, "ε-separating nonlinear digital filter and its applications," IEICE Trans. Fundamentals **J65-A**, 297–303 (1982).
[24]K. Arakawa, K. Matsuura, H. Watabe, and Y. Arakawa, "A method of noise reduction for speech signals using component separating ε-filters," IEICE Trans. Fundamentals **J85-A**, 1059–1069 (2002).

Comparison of a subrank to a full-rank time-reversal operator in a dynamic ocean

Geoffrey F. Edelmann,[a] Joseph F. Lingevitch, Charles F. Gaumond, David M. Fromm, and David C. Calvo
Acoustics Division, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC 20375-5320

(Received 30 March 2006; revised 8 August 2007; accepted 17 August 2007)

This paper investigates the application of time-reversal techniques to the detection and ensonification of a target of interest. The focusing method is based on a generalization of time-reversal operator techniques. A subrank time-reversal operator is derived and implemented using a discrete set of transmission beams to ensonify a region of interest. In a dynamic ocean simulation, target focusing using a subrank matrix is shown to be superior to using a full-rank matrix, specifically when the subrank matrix is captured in a period shorter than the coherence time of the modeled environment. Backscatter from the point target was propagated to a vertical 64-element source-receiver array and processed to form the sub-rank time-reversal operator matrix. The eigenvector corresponding to the strongest eigenvalue of the time-reversal operator was shown to focus energy on the target in simulation. Modeled results will be augmented by a limited at-sea experiment conducted on the New Jersey shelf in April–May 2004 measured low-frequency backscattered signal from an artificial target (echo repeater). [DOI: 10.1121/1.2783127]

PACS number(s): 43.60.Tj, 43.60.Fg, 43.30.Vh, 43.30.Re [DRD] Pages: 2706–2714

I. INTRODUCTION

The performance of active acoustic systems for target detection, tracking, and classification is degraded by: ocean variability, target motion, noise, and reverberation. One solution to the problems of noise and reverberation is time-reversal focusing. By focusing sound on the target, the target echo is increased, whereas reverberation from the surface and bottom is decreased; thus, increasing the signal-to-reverberation ratio.[1–3] Time reversal methods are still susceptible to ocean variability, particularly time-reversal operator methods for which the acquisition time of the data can exceed the coherence time of the medium. This paper will describe a subrank beam-based time-reversal operator technique that requires no environmental knowledge to find and focus sound on a target in a changing heterogeneous waveguide. The time-reversal operator is constructed with a few snapshots that are sufficient to capture the target, yet be measured within the period of temporal coherence. In addition to averaging over multiple snapshots, the target signal-to-noise-ratio was increased by utilizing a beam-based decomposition of the time-reversal operator (DORT) method.[1]

Time-reversal focusing of sound has been demonstrated in ultrasonic laboratory experiments[4,5] and at-sea experiments.[6–8] In these demonstrations, time-reversal techniques were applied to focus acoustic energy back to a probe source in a heterogeneous waveguide without requiring any knowledge of the propagation environment. The governing principle is the time-reversal invariance of the wave equation which implies that an outgoing wave will refocus to the original source location when the wavefront is reversed in time. In practice, a time-reversed wave is approximately realized using a time-reversal mirror (TRM) implemented by a source/receiver array (SRA) of collocated transducers. The TRM discretely samples a wavefront in space and time and retransmits a time-reversed version of the sampled signal.

Time-reversal mirrors have been used to detect and selectively focus acoustic energy on scatterers in a waveguide.[9,11] The DORT algorithm[11–13] considers a set of point scatterers. A response matrix is measured by sequentially transmitting impulses from individual source elements of the TRM and recording the backscattered echo from all targets on each receiver element. Ideally, the response matrix is symmetric and contains the measured Green's functions between the targets and every transducer of the TRM. In the DORT algorithm, this matrix is transformed into the frequency domain and factored by singular value decomposition. The resulting singular vectors approximate array weights associated with individual point targets. Retransmission of the conjugate array vectors yields a focus at the corresponding target location. This original theory has been expanded to include non-isotropic point scattering,[14] multiple scattering,[15] non-ideally separated targets,[13] scattering by extended objects,[16,17] and object imaging.[12,18]

In shallow water, the quality of the time-reversal focus is degraded by ocean change, target motion, SRA deformation, and noise.[6,19,20] In order to mitigate these effects, it is useful to formulate the DORT method in terms of a sum of covariance matrices over multiple snapshots of the gradually changing environment.[13] This formulation emphasizes the correspondence between time reversal and matched field processing (MFP) techniques which have been extensively developed in passive ocean acoustics.

In this paper, a subrank beam-based DORT technique for target detection in shallow water ocean environments is derived and demonstrated in simulation. The simulation is augmented with limited experimental data taken during the

[a]Electronic mail: geoffrey.edelmann@nrl.navy.mil

TREX-04 time-reversal focusing experiment. Sound propagation was modeled in a changing ocean waveguide as was measured between April and May 2004 on the New Jersey shelf. The drift of the echo repeater during the acquisition of the response matrix was the primary source of variability in time. Backscatter measurements of a drifting echo repeater acting as an artificial target were made with the Naval Research Laboratory's 64-channel SRA in the 500–600 Hz band. In Sec. II, the theory for the subrank beam-based time-reversal operator (TRO) is developed. In Sec. III, the theory is applied using a combination of numerical simulation and experimental data. It is shown that the subrank TRO is beneficial in cases where target drift or environmental fluctuations are significant during the time required to acquire the full-rank TRO.

II. THEORY

A. The time-reversal operator

The TRO

$$\mathbf{T} = \mathbf{K}^*\mathbf{K} \tag{1}$$

has been derived previously by Prada *et al.*[9] in the context of an acoustic iterative time-reversal process where $K_{ij}(\omega)$ is the frequency-domain transfer function between source element j and receive element i of a time-reversal mirror (complex conjugation is denoted by "an asterisk"). In this section, the basic theory relating the eigenvectors of the time-reversal operator to the one-way Green's functions (from individual targets to the TRM) is briefly summarized. Then, using the covariance matrix interpretation of the TRO (Sec. II C), we formulate a subrank time-reversal operator (Sec. II D).

The analysis of Prada treats the detection of a set of M isotropic point scatterers (targets) by an N element array ($M \leqslant N$) when multiple scattering between the targets can be neglected. Under these assumptions, the $\mathbf{K}$ matrix can be written in terms of the Green's function as,

$$\mathbf{K} = \mathbf{G}\boldsymbol{\Omega}\mathbf{G}^T A \tag{2}$$

where $G_{ij}(\omega)$ is a $N \times M$ matrix of normalized one-way Green's functions between source i and scatterer j, Ω_{ij} is a $M \times M$ diagonal matrix that includes the target scattering coefficients and spreading loss between the TRM and the targets, and $A(\omega)$ is the source function. Substituting Eq. (2) into Eq. (1) gives the TRO in terms of the one way Green's function as

$$\mathbf{T} = \mathbf{G}^*(\boldsymbol{\Omega}^*\mathbf{G}^H\mathbf{G}\boldsymbol{\Omega})\mathbf{G}^T|A|^2. \tag{3}$$

In order to relate the eigenvectors of $\mathbf{T}$ to the one-way Green's functions, we require that the TRM spans the water column with sufficient resolution to adequately sample the propagating modes, and that the targets are ideally separated (by more than a wavelength), so that the columns of $\mathbf{G}$ are approximately orthogonal,[6,9] that is

$$\mathbf{G}^H\mathbf{G} \approx \mathbf{I}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix. In this case, the grouped term in Eq. (3) is a diagonal matrix and the TRO simplifies to

$$\mathbf{T} \approx \mathbf{G}^*|\boldsymbol{\Omega}|^2\mathbf{G}^T|A|^2 = \sum_{k=1}^{M} |\Omega_{kk}A|^2 G_{ik}^* G_{jk}. \tag{5}$$

From Eq. (5) and the orthogonality relationship, Eq. (4), we see that the nonzero eigenvalues of the TRO are $|\Omega_{kk}A|^2$ and the corresponding eigenvectors are given by the columns of $\mathbf{G}^*$. Thus, decomposition of the TRO yields the one way Green's functions from the time-reversal mirror to the scatterers.

B. Measurement of the time-reversal operator

The time-reversal operator can be obtained from a direct measurement of the response matrix, $\mathbf{K}$, on an N-element array using Eq. (1). This is the method used in previous ultrasonic tank experiments. In principle, $\mathbf{K}$ is a square symmetric matrix due to the reciprocity relationship between source and receiver locations.[21] In practice, the experimentally measured data matrix, $\mathbf{K}_d$, is acquired over a time interval involving multiple source transmissions; therefore, in applications where reciprocity is degraded over time by variability in the water column, source/receiver motion, and noise,[6,19,20] the direct measurement of the full response matrix becomes problematic. When considering these factors in the measurement of the $\mathbf{K}_d$ matrix, the question arises of how to construct a meaningful TRO in the case of low signal level and/or time varying conditions.

The main theoretical result of this paper is a generalized TRO derived below and given by

$$\mathbf{T}_d = \mathbf{K}_d^*\mathbf{K}_d^T, \tag{6}$$

where the T superscript denotes the transpose of the matrix. In this formulation, $\mathbf{T}_d$ is a Hermitian matrix regardless of the symmetry of $\mathbf{K}_d$. Additionally, it is not necessary that $\mathbf{K}_d$ be a square matrix. In the special case of a symmetric and square $\mathbf{K}_d$ matrix, Eq. (6) reduces to Eq. (1). We will show in Sec. II C that, for the case of a non-symmetric $\mathbf{K}_d$ matrix, Eq. (6) is interpreted as a sum over a set of covariance matrices.

C. The time-reversal operator and the covariance matrix

MFP is a common method for target detection and localization in the field of underwater acoustics. It is useful to frame the discussion of the time-reversal operator in the well developed terminology, known strengths, and limitations of MFP.[22] Most MFP techniques operate on a covariance matrix which is generally unknown and must be estimated. An ensemble of data snapshots (in time) are used to estimate the covariance matrix $\mathbf{C}$ as discussed in (Ref. 21, Chap. 10) and (Ref. 23, Chap. 15.4). These data contain signal and uncorrelated noise and are measured along an array of receivers. For an array with N hydrophones, $\mathbf{C}$ is a $N \times N$ matrix. In order to estimate $\mathbf{C}$, snapshots of recorded data are windowed, overlapped, and averaged with each other. If $\mathbf{k}_n$ is the nth snapshot of the complex pressure measured along the array at an angular frequency ω, then with L realizations we define the covariance matrix (suppressing the $1/L$ normalization constant) as

$$\mathbf{C} = \sum_{n=1}^{L} \mathbf{k}_n \mathbf{k}_n^H. \tag{7}$$

The covariance matrix $\mathbf{C}$ has L degrees of freedom and here we assume that the propagation environment has not changed appreciably during the period of data acquisition. Assuming that the noise and signal are uncorrelated, averaging a number of snapshots will improve the signal-to-noise ratio of the measurement. Additionally, as $\mathbf{C}$ is Hermitian, its eigenvalues are positive and its eigenvectors are orthogonal. With sufficient signal-to-noise ratio, the strongest eigenvalues correspond to any strong point scatterers or sources present in the data. The remaining eigenvalues represent noise. The conjugate of the first eigenvector of $\mathbf{C}$ is a steering vector that points in the direction of the strongest target. Transmitting this vector is equivalent to a single frequency version of time reversal and is a technique commonly referred to as eigenvector beamforming (Ref. 21, Chap. 10.3).

If the N measurements of $\mathbf{k}_n$ are made by transmitting a probe pulse from the N individual TRM transducers (or N orthonormal beams, Sec. II D), then the measured $\mathbf{k}_n$ vectors are the columns of the response matrix $\mathbf{K}$. If we have $L=N$ number of snapshots, then we can define an ensemble of the data snapshots as

$$\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_N], \tag{8}$$

in which case the covariance matrix is equal to

$$\mathbf{C} = \mathbf{K}\mathbf{K}^H. \tag{9}$$

Equation (9) is simply the conjugate of the time-reversal operator as written in Eq. (6). Therefore, we define the TRO as

$$\mathbf{T} = \mathbf{K}^*\mathbf{K}^T = \mathbf{C}^*. \tag{10}$$

The TRO is the conjugate of the MFP covariance matrix.

D. Beam-space methods and the subrank time-reversal operator

The beam space time-reversal operator is a generalization of the element space TRO introduced by Prada.[9–11] In ocean experiments, measuring the backscatter from targets is hindered by ambient noise levels and small target scattering strengths. The generalization to beam space increases the total transmitted power by broadcasting from all transducers simultaneously instead of transmitting from individual transducer elements.

We formulate a beam representation of the response matrix by defining the amplitude and phase of the transducer elements on an array as the columns of a complex matrix $\mathbf{E}$. For an N-element array, N beams are chosen to satisfy the orthogonality condition $\mathbf{E}^H\mathbf{E}=\mathbf{E}\mathbf{E}^H=N\mathbf{I}$, where the superscript H denotes complex conjugate transpose and $\mathbf{I}$ is the identity operator. In this paper, we choose

$$E_{jk} = \exp(i2\pi j(k - N/2 - 1)/N), \tag{11}$$

where k is the index of the beam and j is the index of the transducer. With this phasing, beam index $k=N/2+1$ corresponds to a broadside transmission, beam indices $1 \leqslant k < N/2+1$ correspond to up-going plane waves, and indices $N/2+1 < k \leqslant N$ correspond to down-going plane waves. The beam-space representation is simply related to the element-space in terms of a discrete Fourier transform. Also, note that the incoming beams are given by the columns of $\mathbf{E}^*$ and they satisfy orthogonality, $\boldsymbol{E}^T\boldsymbol{E}^* = \boldsymbol{E}^*\boldsymbol{E}^T = N\boldsymbol{I}$ where the superscript T denotes transpose. The beam-space response matrix is related to the element-space matrix by the transformation,

$$\tilde{\mathbf{K}} = \mathbf{E}^T\mathbf{K}\mathbf{E}, \tag{12}$$

where the elements of $\tilde{K}_{jk}$ correspond to the response on the jth incoming beam due to an impulsive transmission on the kth outgoing beam. Note that $\tilde{\mathbf{K}}$ is symmetric if $\mathbf{K}$ is a symmetric matrix. The beam space TRO, $\tilde{\mathbf{T}}$, is defined by simply replacing $\mathbf{K}$ with $\tilde{\mathbf{K}}$ in Eq. (10) and interpreting the eigenvectors of this operator in a beam-space basis.

The original DORT theory was formulated in terms of a response matrix measured in the element-space basis. The beam-space basis, described above, is also a complete orthonormal set. We emphasize here that a complete, orthonormal basis is not a necessary requirement. In fact, backscatter from a single beam which ensonifies a target is sufficient to form the Green's function to that target. In many cases of interest, neither the target location nor the propagation characteristics of the beams will be known in advance. Therefore, it is advantageous to build the TRO from an ensemble of beams that will be more likely to successfully ensonify an arbitrarily located target. Compared to repeatedly transmitting the same beam (e.g., broadside) N times, transmitting a set of unique beams will illuminate different propagation paths between array and target. Additionally, each $\mathbf{k}_n$ records a different realization of the ambient noise field. However, due to the dynamic nature of the propagation medium in ocean acoustic applications, it may be impractical to measure the full matrix $\tilde{\mathbf{K}}$ in a sufficiently short period of time. In this case, a covariance matrix is formed using a subset of transmit beams. We refer to this as a subrank time-reversal operator.

The covariance matrix formulation of the TRO in Eq. (10) offers several advantages over the original theory. First, this convention forces the TRO to be a Hermitian matrix, even if the $\mathbf{K}_d$ matrix is no longer symmetric (changing medium) or $\mathbf{K}_d$ is deficient because not all of its elements were measured. Second, a subrank version of the $\mathbf{K}_d$ matrix can quickly characterize a target by transmitting only a subset of the possible beams $\mathbf{E}$, thereby mitigating the well-known "snapshot problem" in MFP literature by measuring the target before the medium changes.[22] Last, $\mathbf{K}_d$ can be overdetermined by measuring multiple realizations of a given transmit beam; this is especially useful if one beam is believed to be, through experience or simulation, superior at ensonfying the desired target (therefore $\mathbf{K}$ can be larger than $N \times M$). This paper will demonstrate that, in a dynamic environment, quickly measuring a coherent subset of beams can produce a superior TRO matrix than a TRO formed from a complete yet incoherent set of beams.

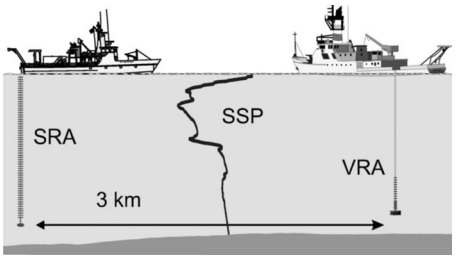


FIG. 1. The R/V Henlopen was moored in approximately 95-m deep water and is depicted with the source-receiver array (SRA) deployed from the A-frame. The array focused sound on the echo-repeater/probe source which was married beneath a short vertical receive array (VRA) deployed from the R/V ENDEAVOUR, which was slowly drifting away in 85-m deep water. A typical soundspeed profile (SSP) is shown.

III. SIMULATION AND EXPERIMENT

In this section, the main results of this paper are demonstrated numerically. Some supporting experimental measurements are also described. A time-reversal experiment (TREX-04) was conducted in April–May 2004 by the Naval Research Laboratory to test TRO methods for detection of an artificial target. The prevailing experimental conditions were utilized in the modeled environment. The ship geometry for the experiment is shown in Fig. 1.

A. Experimental setup

The following is a description of the equipment used and the environmental conditions present during TREX-04 from April 22 to May 4. The measured ocean conditions and experimental geometry were utilized for the simulation of acoustic propagation. Acoustic measurements were made in the region southwest of the Hudson Canyon off the coast of New Jersey. The water depth in this relatively flat region spanned from about 95 to 85 m over a range of approximately 3 km. The R/V HENLOPEN, moored in 95-m deep water, deployed a vertically suspended SRA. The 64-element SRA had 1.25 m interelement spacing for a total vertical aperture of 78.75 m. During the TREX-04 event described in this paper, the transmitted signals were 500–600 Hz LFM sweeps corresponding to a center frequency wavelength of about 2.7 m. The R/V ENDEAVOUR deployed a single Raytheon XF4 transducer below a 16-element, 8 m aperture dual nested vertical receive array (VRA). Although a longer aperture VLA would have been desirable, the short monitoring VLA does provide a glimpse at the spatial distribution of the acoustic field near the target. The XF4 was used both as a probe source and as an echo-repeater. When used as an echo-repeater, the VRA element closest to the XF4 (2 m separation) was echoed on the XF4. The R/V ENDEAVOUR drifted slowly away from the R/V HENLOPEN at approximately 0.6 km/h (0.3 knot) during acoustic transmissions. The echo repeater is considered to be the target in this paper. More details of TREX-04 are presented in Refs. 24–27.

Environmental measurements were made during this TREX-04 event as shown in Fig. 2. Twenty-one thermistor elements, which were physically coupled to the SRA, made continuous measurement of ocean temperature. Their record is shown in Fig. 2(a) over the two and a half days on interest. Conductivity, temperature, and depth (CTD) measurement were taken three times daily from the starboard winch. Overlapping temperature measurements recorded by the CTD casts and thermistor elements were in excellent agreement. The salinity profiles inferred from the ocean conductivity are shown in Fig. 2(b). Sound speed calculated from salinity and temperature appears in Fig. 2(c). From the point of view of measuring strong reverberation, the sound speed structure

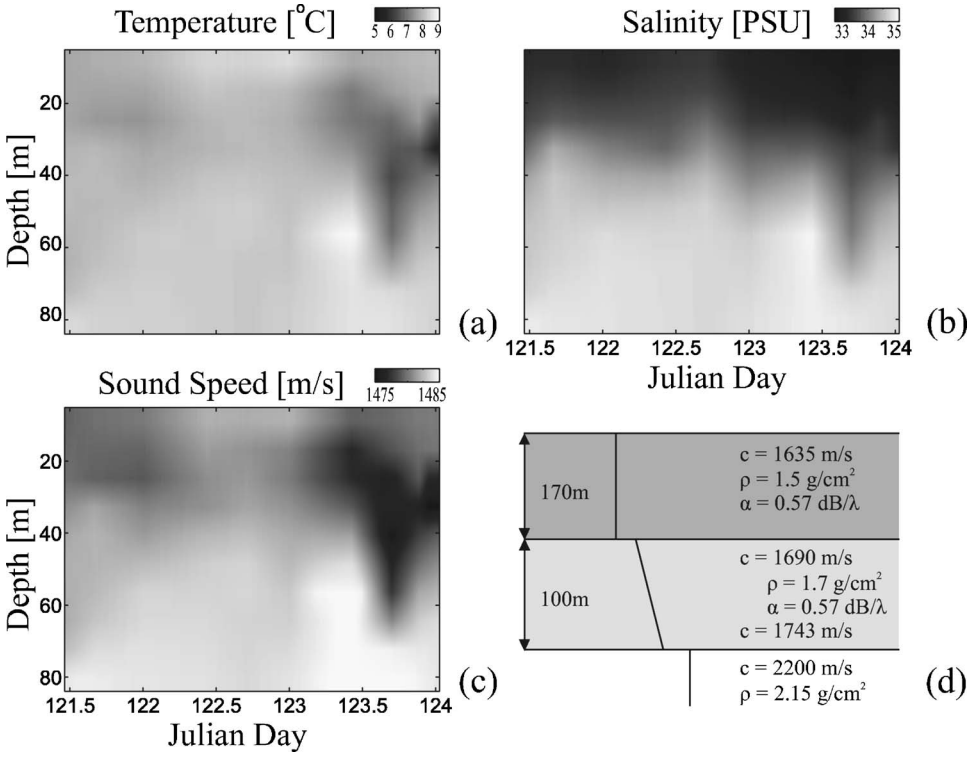


FIG. 2. (a) Twenty-one thermistors sampled the temperature field along the aperture of the SRA every 10 s. (b) Salinity measurements were inferred three times daily with CTD casts. (c) Calculated sound speed showing an upward refracting profile that traps energy in the water column. (d) The geoacoustic model for the experimental site area off the coast of New Jersey.

unfortunately created a sound channel that trapped acoustic energy in the water column thus preventing it from interacting with the bottom. The geoacoustic parameters in Fig. 2(d) are used for acoustic propagation modeling in this paper. These values are obtained from previous site surveys in the region.[26] The sediment layer has a sound speed that is greater than that of the water column. Measured reverberation was weak especially in light of the prevailing sound speed and relatively low acoustic transmission levels. Finally, the range-dependent bathymetry between the two ship positions was measured with an echo-sounder.

The coherence time of the channel was estimated to be at least 12 min. The point-to-point estimate was made by transmitting probe source pulses from the XF4 and cross-correlating the receptions at the 80-m depth SRA hydrophone in the half-hour prior to the TRO experiment. It includes environmental fluctuations as well as ship drift.

The data analyzed in this paper were measured during at a single event of the TREX-04 experiment conducted on May 1, 2004. Unfortunately, active acoustic transmissions were limited; therefore, this paper reinforces numerical simulations based upon the measured environment with acoustic data. These simulations are shown to agree with the available measured data and are also used to extrapolate when measured data were unavailable.

B. Time reversal of a probe source signal

This section will discuss at-sea measurements of time-reversal focusing and compare those measurements to simulations created by backpropagating the measured probe source signal data using computer models. By successfully simulating the measured time-reversal focus, we validate the use of both the propagation model and our detailed understanding of the environment. These simulations are then used to further elaborate on time-reversal operator methods.

On May 1st (Julian Day 122), a series of time-reversal foci were measured. The experimental setup of time-reversal focusing is shown in Fig. 1. The slowly drifting R/V ENDEVOUR deployed the XF4 probe source (PS) 2-m below the short VRA. The PS transmitted a Tukey windowed 1 s up-sweep chirp from 500 to 600 Hz. The PS chirp propagated about 2.8 km and was received on the SRA. The top element of the vertical SRA was suspended 3-m below the surface from the A-frame of the moored R/V HENLOPEN. The SRA digitally recorded, time-reversed, and retransmitted the PS transmission.[6] The time-reversed signal focused back at the original PS position. There was a 2 min turn around time from PS capture and time-reversal focus. Three of the foci measured on the VRA are shown in Fig. 3. Note that the VRA is 2-m above the original probe source position therefore the VRA is capturing only the upper half of the focal region. At these frequencies, the 3 dB down width of the focus was observed at 3 m as expected.[28]

The measured time-reversal foci are compared with simulated time-reversal foci created by backpropagating measured PS data using a computer model. In order to make accurate simulations of acoustic propagation, the *in situ* environment (bathymetry, thermistor temperature, salinity, and

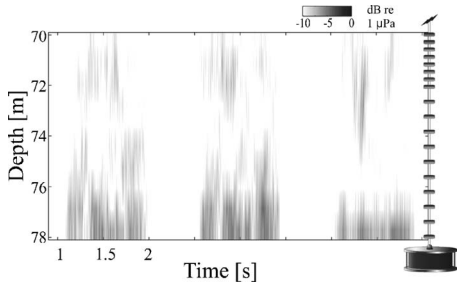


FIG. 3. A gray scale plot of three measured time-reversal foci are shown over depth and time. A diagram to the right of the time series shows the nested vertical-receiver array and its placement above the probe source. The probe source transmitted a Tukey windowed 1 s pulse which was measured on the source-receiver array about 2.8-km away. The signal was windowed, time-reversed, and retransmitted. The vertical-receiver array is measuring only the upper part of the focus.

GPS estimated ship locations) was used as described in Sec. III A. A wide-angle range-dependent parabolic equation model (RAM)[29] was used to simulate the broadband propagation. Time evolution of the water column and ship locations is incorporated into the simulations by using sound speed and ship location data that is 1 min older than the probe source reception. The environmental change includes temperature and salinity, but even more importantly incorporates slow range drifting of the PS. The R/V ENDEAVOUR was drifting away from the moored R/V HENLOPEN at approximately 0.6 km/h (0.3 knot). The drift was not fast enough to introduce Doppler spread into the PS signal.

Acoustic transmission loss simulated by backpropagating recorded PS data signals in the measured environment can be seen in Fig. 4. The upper left-hand panel shows the single-frequency (550 Hz) simulated backpropagation of a captured data PS ping in range and depth. The recorded PS ping used in this simulation was one of the three that were time-reversed, retransmitted, and focused on the VLA shown in Fig. 3. The time-reversal signal focuses at the same range as, and 2-m shallower than, the original PS position. Simulated backpropagation of a measured time-reversed PS signal is related to Bartlett matched-field processing.[22] Figure 4(c) shows the synthesised time series versus depth at the VRA that is simulated using the measured PS data from the SRA. These time-reversal foci simulations used the same PS signals recorded, time-reversed, and retransmitted to form the foci shown in Fig. 3. A side-by-side comparison of Figs. 3 and 4(c) shows that the backpropagated time series of the PS data are qualitatively identical to the measured time-reversal foci.

Next, MFP tracking of the PS position over time is used to verify that *in situ* ocean variability has been accurately modeled. Over the half-hour period of interest, the SRA recorded a PS signal almost every minute. Figures 4(b) and 4(d) depict a track of MFP results. A MFP ambiguity surface is made every minute over the course of half an hour. Range tracking at a fixed 80-m depth is displayed in Fig. 4(b). The most-likely range of the PS corresponds to the darkest peak. The global positioning system estimated range is overlaid as a black and white line. Using measured PS ping data, MFP is able to resolve accurately the PS in range. Note that no PS

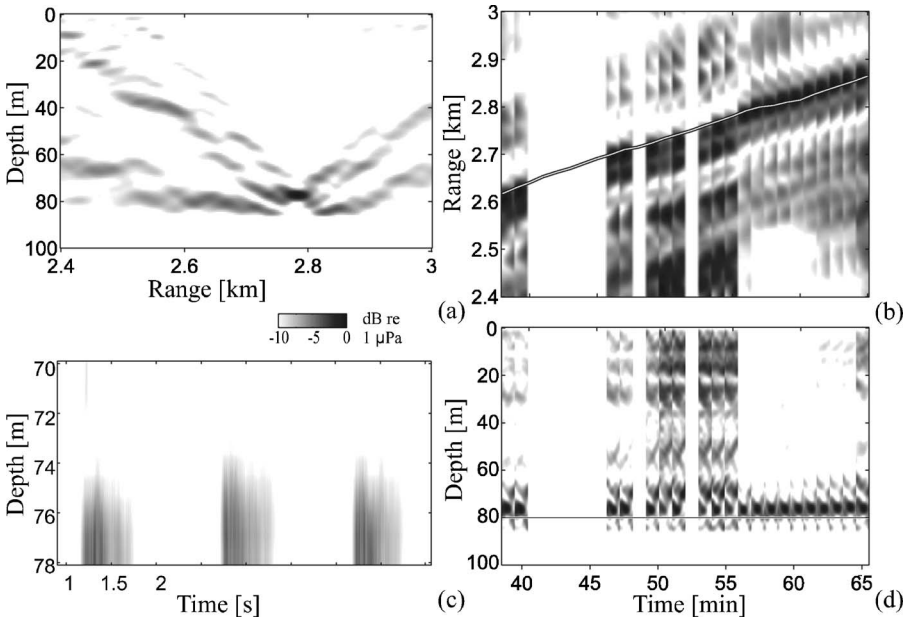


FIG. 4. Modeled backpropagation of probe-source signal data recorded on the SRA using the environmental measurements described in Fig. 1. (a) Single frequency ambiguity surfaces simulated at 550 Hz by backpropagating a measured probe-source data signal. The nominal range/depth of the probe-source is 2.8 km/80 m. (c) Coherent broadband simulation of the received time-reversal focus as it would be received on the VRA (three different realizations of the sweep are shown). (b) and (d) Tracking the target range and depth versus time. (b) Stacked plot of horizontal slices of the ambiguity surfaces at the nominal probe source depth. The solid line indicates the GPS estimated distance between the ships, one of which was freely drifting. (d) Vertical slices of the ambiguity surfaces at the nominal probe-source range. The solid line shows the measured depth of the probe source, which was different by 2 m.

signals were transmitted during the blank period (41–44 min). The MFP ambiguity surface of source depth over time appears in Fig. 4(d). Again, simulations reveal clear and successful MFP resolution over an extended period of time. Although accurate, the predicted PS depth is 2-m shallower than the depth sensor recordings (black and white line). This could be the result of a slight mismatch in bathymetric depth or the positioning of the depth sensor.

The correspondence between measured data and simulation demonstrates sufficiently an accurate understanding of the propagation environment between the two ships and a reliable method to simulate sound propagation during the TREX-04 experiment. These same simulation techniques will be used to verify and predict results in Sec. III C using time-reversal operator methods described Sec. II D.

C. Beam space methods and the time-reversal operator

This section will show simulations and limited results from a TRO target focusing portion of TREX-04. The degradation of TRO focusing due to environmental change and compensation via a subrank TRO measured over a shorter time frame shown. Additionally, the simulated TRO will be compared to the experimentally measured TRO and shown to be in excellent agreement.

We apply this approach to detecting targets in the ocean as follows: (1) a set of orthogonal transmit-beams is selected from Eq. (11) and the phase delays are applied to a 1.0 s LFM sweep (500–600 Hz) to form beams, (2) the transmit-beams are individually broadcast and 10 s of backscattered energy is recorded on all channels, (3) the backscattered signal is match filtered with the LFM and segmented in time to obtain the beam-element responses $k_{ij}(t)$ as a function of range from the array, (4) the beam-element responses are transformed to the frequency domain and singular vectors for the data TRO are formed, (5) the singular vector amplitude and phase delays are applied to the LFM sweep and the focusing pulse is transmitted from the SRA.

1. Simulated full-rank time-reversal operator focusing in a static ocean

In a static environment, measurement of the full TRO is desirable since this improves both the resolution of the focus and the signal-to-noise ratio. Ideal TRO focusing of a target at 80-m depth is displayed in Figs. 5(a) and 5(b). In this simulation, a complete set of 64 beams were transmitted and echoes from the target are recorded on the 64 element SRA. The TRO matrix was formed at the center frequency of the array (550 Hz) and decomposed into eigenvalues and eigenvectors. Only one strong eigenvalue was present, and the associated eigenvector corresponded to a weighting vector pointing acoustic energy back to the target. Ideal focusing of the first vector of the TRO is shown in Fig. 5(a). For the given geometry, attenuation, and frequency, this is the best focusing possible using TRO methods. A broadband signal can be sent using the eigenvector as an array weight. The incoherent average of the focus in Fig. 5(b), from 500 to 600 Hz, shows that an eigenvector made at the center frequency will successfully focus over the entire band. Applying the single frequency weight to a broadband signal smears the focus in range and depth but this can be an advantage if the target is mobile. In both Figs. 5(a) and 5(b) the energy is focused upon the target. The cylindrical spreading term has been removed from the transmission loss.

2. Simulated full-rank time-reversal operator focusing in a dynamic ocean

The ocean is a dynamic medium and environmental change can degrade TRO focusing. In this case, the simulated environment evolves as was recorded during the TRO portion of the TREX-04 experiment. Temperature, salinity, and, most importantly, ship-drift play a part in degrading the

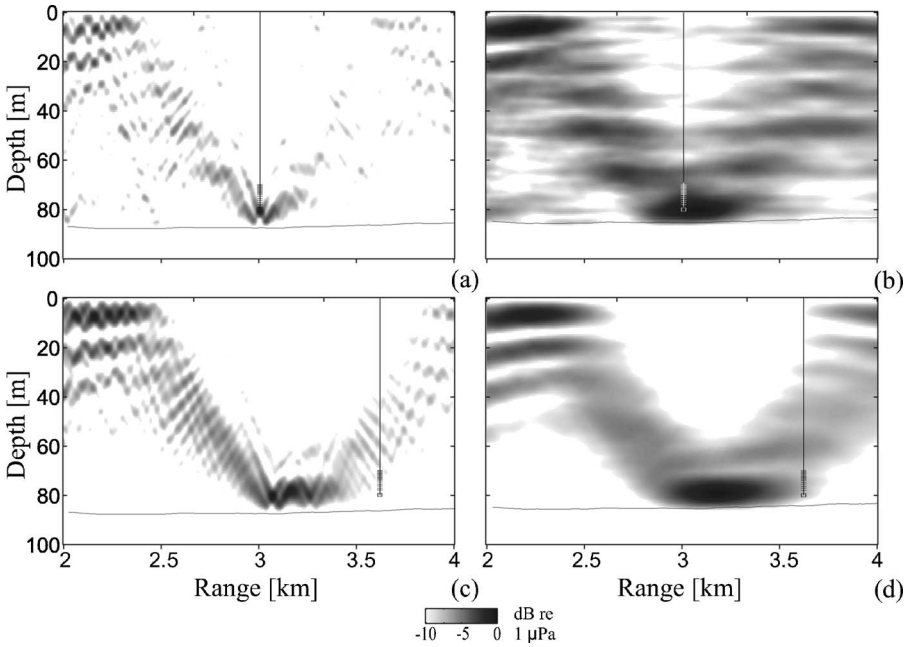


FIG. 5. Modeled TRO focusing simulation using 64 simulated beams. (a) Single frequency TRO focus of the first eigenvector in an inhomogeneous, range dependent, and static version of the typical ocean environment during the TREX-04 experiment and excluding ship drift. (b) Incoherent average of the broadband (500–600 Hz) TRO focus. (c) Single frequency TRO focus in a dynamically changing ocean including estimated ship drift. (d) Incoherent average of the dynamic broadband TRO focus. The black line represents the target collocated with the VRA which is drifting. Measuring all 64 beams, at one beam a minute, takes more than a hour which leads to a broadening of the focus. Note that the ship has drifted out of the center of the focal region.

TRO focus at the target position. The TRO is simulated by transmitting one beam per minute over the course of 64 min. The target was deployed from the R/V ENDEAVOUR which drifted almost 600 m over that time period in a changing ocean. Target focusing by weighting the SRA with the strongest eigenvector is shown in Fig. 5(c). The target has drifted beyond the focal position and is not successfully ensonified. Note how the focus is no longer sharp and has broadened in range and depth. The broadening is due to the averaging the Green's function over time and space and is the equivalent to the multiple constraint matching processor in MFP.[30] The broadband TRO method Fig. 5(d) peripherally ensonifies the target.

In a dynamic environment, one target can be associated with multiple eigenvalues. As the ocean changes the target back-scatter is incoherently distributed amongst multiple eigenvalues. In the static case, only one strong eigenvalue was present, the rest were associated with noise. In the dynamic case, the target has been spread over five eigenvalues. Practically speaking, unless there is concern about sending a coherent signal to the target (e.g., a communication signal), transmitting simultaneously multiple eigenvectors may prove advantageous. Transmission of the second TRO eigenvector [Fig. 6(a)] ensonifies the general target area and appears to be associated with the back-scatter simulated at later ranges. Transmitting both the first and second eigenvectors weighted by their respective eigenvalues appears in Fig. 6(b). Transmitting the first two eigenvectors has tightened the focus in range when compared to Fig. 5(c).

For the purpose of broadening the TRO focus and ensonifying the target, transmitting multiple eigenvectors did not prove advantageous during the environmental conditions present at the TREX-04 experiment.

3. Simulated subrank time-reversal operator focusing in a dynamic ocean

Measuring a subrank TRO can produce superior target ensonification in a dynamic ocean. During the TREX-04 experiment only a subset of the 64 beams were used to form the TRO; specifically, starting with the first beam, every 4th beam was transmitted ending on the 61st. These beams were

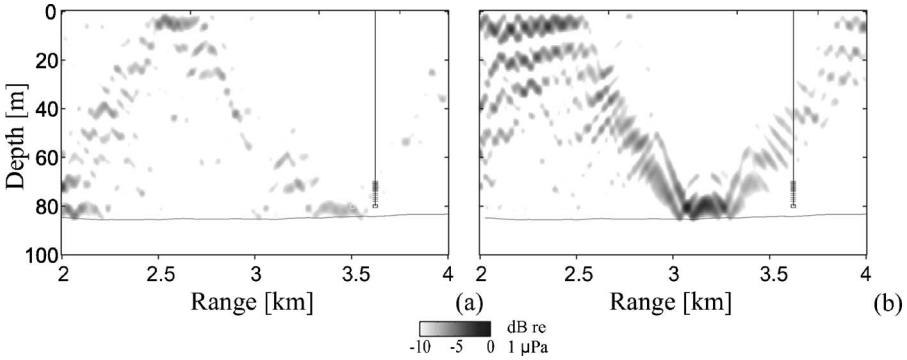


FIG. 6. Simulated single frequency TRO focusing using multiple eigenvectors in a dynamic ocean. (a) TRO focus using the second eigenvector and (b) TRO focusing using both the first and second eigenvectors weighted by their respective eigenvalues. No significant enhancement is seen. Over the 64 min used to measure the full-rank TRO the target has been spread into five significant incoherent eigenvalues (the other eigenvalues are noise). In the static environment [Figs. 5(a) and 5(b)], only one strong eigenvalue was present and corresponded to the target.

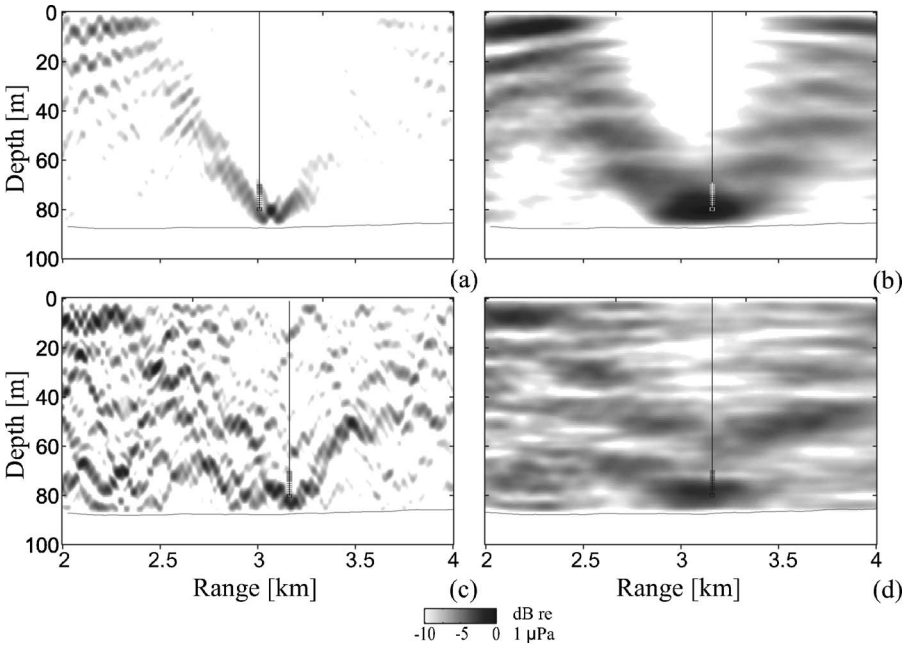


FIG. 7. Comparison of modeled backpropagation of simulated sub-rank TRO and measured subrank TRO. The subrank TRO was measured over a 15 min period using every fourth beam during the TREX-04 experiment. Backpropagation was modeled using a TRO constructed with simulated target response in (a) and (b) and from measured at-sea target response in (c) and (d). Subrank TRO focusing of simulated scatter: (a) Single frequency TRO focus of the first eigenvector in an inhomogeneous range-dependent and dynamic ocean with a drifting target. (b) Incoherent average of the broadband (500–600 Hz) TRO focus. Sub-rank TRO focusing of experimentally measured scatter: (c) Single frequency TRO focus of the first eigenvector. (d) Incoherent average of the broadband (500–600 Hz) TRO focus. In both cases, when the subrank TRO was estimated over a 15 min (instead of 64 min) period, the target was incoherently smeared into just two eigenvalues. In this dynamic case, the focus associated with the subrank TRO is superior to the focus of the full-rank TRO shown in Fig. 6.

not optimally selected; the only criterion was the desire to measure the TRO in a relatively short period of time (15 min). To ensure that the target was ensonified by at least a few beams, a broad range of launch angles were included. It should be noted that the steepest beams were absorbed by the sediment before reaching the target. Backscatter was present for only beams 21–41 in both the simulation and measurements made during TREX-04. Although it may have been optimal to select a narrower subset of beams between 21 and 41, such presumption would have required *a priori* estimates of acoustic propagation in the environment.

Backpropagation is modeled with a simulated subrank TRO matrix are shown in Fig. 7. Focusing of the first eigenvector at 550 Hz is shown in Fig. 7(a). The incoherent average of the broadband focus Fig. 7(b) ensonifies the desired target position. Before the first eigenvector was transmitted, 1 min is allowed to elapse in the simulated environment. Compared with the full rank matrix discussed in Sec. III C 2, the target focusing is significantly improved and is comparable to the TRO of the static ocean. The eigenvalue spread of the simulated TRO indicated that the target is only associated with two eigenvalues, with the majority of energy contained in the first value. Transmitting a subset of the beam matrix **E** greatly improved the simulated target focusing when that subset is measured within the coherence time of the channel.

4. Comparison of simulated to data derived subrank time-reversal operator focusing

In this section, the subrank TRO measured during the TREX-04 experiment is discussed and shown to focus on the desired target via propagation modeling. During the TREX-04 experiment a subset of 64 beams were transmitted and the back-scatter measured from an artificial target (echo repeater). As mentioned in Sec. III C 3, only beams 21–41 had sufficient signal level to trigger the echo repeater. The TRO was created using the measured back-scatter from all 15 beams. As no assumptions were made about which back-scatter events may or may not contain signal, the **K** matrix included ambient noise vectors.

Analysis of the measured TRO matrix is shown in Fig. 7. The single frequency backpropagation of the first TRO eigenvector is displayed in Fig. 7(c). The energy focuses in range and depth to almost the exact position as the purely simulated TRO in Fig. 7(a). The TRO target focusing appears more clearly in the broadband case, Fig. 7(d). Acoustic energy is incident upon the location of the drifting echo repeater.

Only two strong eigenvalues were present in the eigenvalue distribution of the measured TRO. The distribution is almost identical to the simulated TRO (not shown) with slightly more energy incoherently transferred into the second eigenvector. Surface, bottom, and volume scattering may have contributed to this slight decrease in target coherency. The coherence time of the channel was estimated to be approximately 12 min in the half-hour prior to the formation of the TRO matrix. The TRO matrix was created with 15 beams over 15 min, of which only 6 beams contributed to the signal. Thus, it is not surprising that the sub-rank TRO has little eigenvalue spread.

Comparing respectively Figs. 7(a) and 7(b) to Figs. 7(c) and 7(d) confirms that computer simulations are in agree-

ment with the measured data, and that acoustic energy was focused back on the drifting target during TREX-04. Additionally these results demonstrate that, in a dynamic medium, it is more important to measure a subrank TRO than a full-rank TRO because the target becomes incoherently distributed amongst multiple eigenvalues.

Focusing sound on the target not only increased the echo-repeater response, but also decreased the reverberation from the surface and bottom.[1–3] Figure 7(d) shows that the focal intensity at the target is about 7 dB higher than the surface (at the target range) and from the bottom (before and after target range). Thus, the self-generated reverberation that masks the target was greatly reduced.

IV. CONCLUSION

This paper describes a practical application of the TRO to target detection and ensonification. The TRO technique is based on the DORT algorithm and is capable of ensonifying a target while requiring neither a probe source, prior environmental knowledge, nor acoustic modeling predictions. The subrank TRO method was shown, in simulation, to focus acoustic energy back on an unknown drifting target. The modeled acoustic focus increased the target echo, while it simultaneously reduced reverberation from the surface and bottom; thereby, increasing detection, tracking, and classification capabilities.

The ocean is a time-dependent heterogeneous waveguide, where environmental fluctuations and target drift may be appreciable over the TRO acquisition window. Under these conditions, a target may be incoherently spread into multiple eigenvalues if a full-rank TRO matrix is measured. A simulated subrank TRO, acquired in a shorter time frame than the coherence time of the ocean, was shown to produce a superior focus. The subrank TRO is closely related to covariance matrix techniques of matched field processing.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors would like to thank Richard Menis and Elisabeth Kim.

[1]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).
[2]H. C. Song, S. Kim, and W. A. Kuperman, "Environmentally adaptive reverberation nulling using a time reversal mirror," J. Acoust. Soc. Am. **116**, 762–768 (2004).
[3]S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal, "Echo-to-reverberation enhancement using a time reversal mirror," J. Acoust. Soc. Am. **115**, 1525–1531 (2004).
[4]M. Fink, "Time reversed acoustics," Phys. Today **50**, 34–40 (1997).
[5]M. Fink, "Time reversed acoustics" Sci. Am. (November 1999), 91–97 (1999).
[6]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[7]W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. Jackson, "A long range and variable focus phase-conjugation experiment in shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[8]H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[9]C. Prada, J. L. Thomas, and M. Fink, "The iterative time reversal process: Analysis of the convergence," J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[10]E. Kerbrat, C. Prada, D. Cassereau, and M. Fink, "Ultrasonic nondestructive testing of scattering media using the decomposition of the time reversal operator," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**(8), 1103–1113 (2002).
[11]C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Detection and selective focusing on two scatters," J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[12]N. Mordant, C. Prada, and M. Fink, "Highly resolved detection and selective focusing in a waveguide using D.O.R.T method," J. Acoust. Soc. Am. **105**, 2634–2642 (1999).
[13]C. Prada and J. L. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," J. Acoust. Soc. Am. **114**, 235–243 (2003).
[14]D. H. Chambers and A. K. Gautesen, "Time reversal for a single spherical scatterer," J. Acoust. Soc. Am. **109**, 2616–2624 (2001).
[15]A. J. Devaney, E. A. Marengo, and F. K. Gruber, "Time-reversal-based imaging and inverse scattering of multiply scattering point targets," J. Acoust. Soc. Am. **118**, 3129–3138 (2005).
[16]D. H. Chambers, "Analysis of the time-reversal operator for scatterers of finite size," J. Acoust. Soc. Am. **112**, 411–419 (2001).
[17]S. Hou, K. Solna, and H. Zhao, "Imaging of location and geometry for extended targets using the response matrix," J. Comput. Phys. **199**, 317–338 (2004).
[18]T. Folegot, C. Prada, and M. Fink, "Resolution enhancement and separation of reverberation from target echo with the time reversal operator decomposition," J. Acoust. Soc. Am. **113**, 3155–3160 (2003).
[19]K. G. Sabra, S. R. Khosla, and D. R. Dowling, "Broadband time-reversing array retrofocusing in noisy environments," J. Acoust. Soc. Am. **111**, 823–830 (2002).
[20]K. G. Sabra and D. R. Dowling, "Effects of time-reversing array deformation in an ocean wave guide," J. Acoust. Soc. Am. **115**, 2844–2847 (2004).
[21]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[22]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[23]S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, (Prentice–Hall, Upper Saddle River, N.J., 1993), Vol. **1**.
[24]D. M. Fromm, C. F. Gaumond, J. F. Lingevitch, R. Menis, D. C. Calvo, G. F. Edelmann, and E. Kim, "The effect of coherence and noise on the decomposition of the time reversal operator," OCEANS'05. MTS/IEEE TECHNO-OCEAN'05 (2005).
[25]C. F. Gaumond, D. M. Fromm, J. Lingevitch, R. Menis, G. Edelmann, D. Calvo, and E. Kim, "Application of DORT to active sonar," OCEANS'04. MTS/IEEE TECHNO-OCEAN'04, (2004), Vol. **4**, pp. 2230–2235.
[26]C. F. Gaumond, D. M. Fromm, J. F. Lingevitch, R. Menis, G. Edelmann, D. C. Calvo, and E. Kim, "Demonstration at sea of the decomposition-of-the-time-reversal-operator technique," J. Acoust. Soc. Am. **2**, 976–990 (2006).
[27]D. Calvo, C. F. Gaumond, D. M. Fromm, R. Menis, J. F. Lingevitch, G. F. Edelmann, and E. Kim, "Detection enhancement using multiple time-reversed guide sources in shallow water," OCEANS'05. MTS/IEEE TECHNO-OCEAN'05 (2005).
[28]S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Spatial resolution of time-reversal arrays in shallow water," J. Acoust. Soc. Am. **110**, 820–829 (2001).
[29]M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[30]H. Schmidt, A. Baggeroer, and W. A. Kuperman, "Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch," J. Acoust. Soc. Am. **88**, 1851–1862 (1990).

Optimal adaptive focusing through heterogeneous media with the minimally invasive inverse filter

François Vignon,[a] Julien de Rosny, Jean-François Aubry, and Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, 10 rue Vauquelin, 75231 Paris cedex 05, France

(Received 8 September 2006; revised 3 August 2007; accepted 17 August 2007)

The inverse filter is a technique used to adaptively focus waves through heterogeneous media. It is based on the inversion of the Green's functions matrix between the *M* transducers of a focusing array and *N* control points in the focal area. The inverse filter minimizes the pressure deposited around the focal point. However it is highly invasive, requiring the presence of *N* transducers or hydrophones in the focal area at the control points' locations to measure the Green's functions. This paper presents a way of reaching the inverse filter's focusing quality with a minimally invasive setup: only one transducer (at the desired focal location) is needed. This minimally invasive inverse filter takes advantage of the fact all the information about the propagation medium can be retrieved from the signals backscattered by the medium towards the focusing array, if the propagation medium is lossless. A numerical simulation is performed to test this minimally invasive inverse filter through a scattering, lossless medium. The focusing quality equals the conventional, highly invasive inverse filter's. The average spatial and temporal contrast is increased by up to 10 dB compared to the time reversal focusing. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783128]

PACS number(s): 43.60.Tj, 43.60.Mn, 43.60.Fg, 43.60.Pt [DRD] Pages: 2715–2724

I. INTRODUCTION

Spatio temporal focusing of waves is necessary in a wide range of applications, from medical imaging to telecommunications. In the past 2 decades in the Laboratoire Ondes et Acoustique (LOA), two methods have been developed for adaptive focusing through complex media: the time reversal (TR)[1] and the Spatio Temporal Inverse Filter (STIF).[2]

In order to focus on a point by TR, the first requirement is to receive, on the focusing array, a signal emanating from this point. It is thus necessary to have a source at this point. Emitting from the focusing array the time-reversed version of what has been received allows one to achieve a good focusing. Indeed, frequency-dependent phase aberrations suffered by the wave during propagation are naturally compensated if the propagation medium is nonabsorbing. TR ensures energy maximization at the focus as the signals coming from the different transducers of the focusing array all arrive in phase at the focal point's location. However, no condition is imposed on the neighboring points so that significant amounts of energy can be deposited around the focal point.

In order to minimize the energy deposited around the focal point, a better choice is to use the spatio temporal inverse filter. Based on the measurement of the Green's functions between the transducers of the focusing array and an array of control points in the focal area, the STIF ensures that low pressure levels will be deposited on the control points around the desired focal point.[2,3] The price to pay to reach such an optimal contrast is a much more invasive setup than for the TR: transducers or hydrophones are needed at all the control points' locations in order to measure the Green's functions.

This paper presents a technique that allows, in a lossless medium, to achieve a focusing quality as good as the highly invasive STIF's (contrast optimization) but with a setup as minimally invasive as the TR's (only one source at the desired focal position is needed).

The strategy is as follows: a source at the focal point emits a signal that is received by the focusing array, and time reversed. Instead of simply re-emitting these time-reversed signals from the focusing array—as TR would do—a filter is first applied to them, that changes them into the STIF focusing signals. This filter is derived from the knowledge of the Green's functions measured between the transducers of the focusing array. Information about the transmission of waves through the medium can be deduced from these backscattered Green's functions as there is a direct link between waves backscattered by and transmitted through a lossless medium, as highlighted by Weaver.[4] As the measure of these Green's functions is noninvasive (they are measured using only the transducers of the focusing array, no additional transducer is needed around the desired focal point), such a strategy to compute the STIF signals from the TR signals does not require a more invasive setup than the one used for TR: this way of calculating the STIF signals requires only one source at the focal point's location. It leads to a focusing equal to the one achieved by the conventional, highly invasive STIF described in Ref. 2.

The minimally invasive approach presented here is an adaptation to acoustic waves of an idea introduced first by Pastawski *et al.* in quantum mechanics.[5] Their work suggests that it is possible to actively control the shape and evolution of a wave packet in a closed region of space by introducing locally a carefully calculated time-dependent source term to

[a]Author to whom correspondence should be addressed. Present address: Philips Research North America, 345 Scarborough Road, Briarcliff Manor, NY 10510. Electronic mail: francois.vignon@philips.com

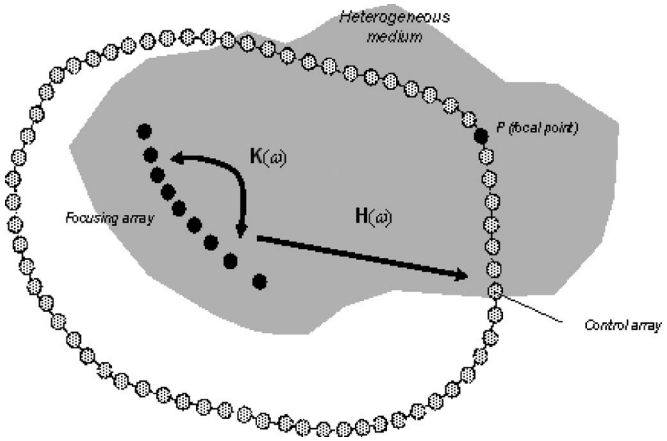


FIG. 1. Setup of a typical focusing experiment. The aim is to focus a wave on a point P embedded in or located behind a heterogeneous medium, from a focusing array of transducers (black points). The array of control points making up the time-reversal cavity surrounding the focusing array is also represented (gray points). Point P is one of the control points. The forward propagator $\mathbf{H}(\omega)$ and the backscatter propagator $\mathbf{K}(\omega)$ are also schematically represented.

the Shrödinger equation. They term their approach the perfect inverse filter (PIF) and demonstrate the benefits of this method over time reversal. That same team also applied the PIF to a microscopic model for one-dimensional sound propagation: a semi-infinite chain of oscillating masses coupled by strings.[6] Earlier on, Dorn noticed that applying the right filter to the TR focusing signals would lead to the STIF focusing signals.[7] In this approach the filter was computed through a series of time-reversal experiments between the focusing array and the array of control points (eqs. (77)–(81) in Ref. 7): the possibility to compute that filter noninvasively was missed.

The outline of the paper is as follows: in Sec. II the time reversal and spatio-temporal inverse filter focusing techniques are briefly recalled, and the formalism is introduced. The noninvasive derivation of the STIF focusing signals from the TR focusing signals is explained. In Sec. III a numerical simulation is presented to illustrate the theory. Section IV discusses the results as well as the practical limitations of the method. Section V concludes.

II. TIME REVERSAL, THE INVERSE FILTER, AND THE MINIMALLY INVASIVE INVERSE FILTER

In this section, the principles of TR and STIF focusing are recalled. The matrix formalism that will be used throughout this paper is introduced. The main differences, relative advantages, and drawbacks of these two adaptive focusing techniques are highlighted. The noninvasive derivation of the STIF focusing signals from the TR focusing signals using the signals backscattered by the medium to the focusing array is then explained.

A. Presentation of the formalism

The aim is to focus a wave from a focusing array of M transducers onto a point P embedded in, or laying behind, a heterogeneous medium (Fig. 1). A "control" array of N control points is set up: it contains the focal point P and makes up a closed time-reversal cavity that surrounds the focusing array. The notion of "closed cavity" implies that: (i) the control points completely surround the focusing array; (ii) the spatial pitch between two adjacent control points is equal to or less than half the smallest wavelength of the used signals; and (iii) the propagation medium is nonabsorbing.[8] This control array will be used for the conventional, highly invasive implementation of the STIF.

Let $\mathbf{H}(\omega)$ be the "forward propagator" that links the M transducers of the focusing array to the N control points of the control array. It is a $N \times M$ matrix whose element $\mathbf{H}_{ji}(\omega)$ is the Green's function (in Fourier domain) between transducer i and control point j. $\mathbf{H}(\omega)$ requires, for its measurements, to place transducers or hydrophones at all the control points' locations in order to measure the Green's functions.

$\mathbf{H}(\omega)$ formalizes the propagation of a wave from the focusing array to the control array. If the emission vector $\mathbf{e}(\omega)$ is sent from the focusing array [the emission vector $\mathbf{e}(\omega)$ is a vector of size M that contains, in position i, the signal emitted by transducer i], the reception vector $\mathbf{f}(\omega)$ (vector of size N that contains, in its jth position, the signal received by control point j) received on the control array is

$$\mathbf{f}(\omega) = \mathbf{H}(\omega)\mathbf{e}(\omega). \tag{1}$$

Let $\mathbf{K}(\omega)$ be the "backscatter" propagator that links together the M transducers of the focusing array: it is an $M \times M$ matrix whose element $\mathbf{K}_{i'i}(\omega)$ is the Green's function between transducers i and i' of the focusing array.[9] This signal is made of the signal that directly propagates from i to i' followed by a signal backscattered by the medium. Hence, $\mathbf{K}(\omega)$ contains information about the propagating medium, through its backscattering properties. Unlike $\mathbf{H}(\omega)$, its measurement is noninvasive: it is not necessary to use transducers or hydrophones embedded in the medium we want to focus through in order to measure it. It only requires the focusing array's transducers for its measurement.

B. Time reversal focusing

The desired focal point (point P in the control array) emits a pulse. It is equivalent to say that the whole control array emits a vector $\mathbf{s}(\omega)$

$$\mathbf{s}(\omega) = [0,0 \ldots 0,1,0, \ldots 0]^T. \tag{2}$$

$\mathbf{s}(\omega)$ is a vector of size N (the number of control points) that contains, in position j, the signal sent by control point j. $S(\omega)$ is zero on all points except on the emitting point P. $(.)^T$ denotes the transpose of a vector or a matrix, $(.)^*$ its complex conjugate, and $(.)^H$ its conjugate transpose.

The wave propagates through the heterogeneous medium and leads to the vector $\mathbf{r}(\omega)$ recorded by the focusing array. These signals are then time reversed to compute the time-reversal focusing signals $\mathbf{e}_{\mathbf{TR}}(\omega)$ that are sent back into the medium to adaptively focus on P. They can be expressed as

$$\mathbf{e}_{\mathbf{TR}}(\omega) = \mathbf{r}(\omega)^* = \mathbf{H}(\omega)^H \mathbf{s}(\omega) \tag{3}$$

[if $\mathbf{H}(\omega)$ formalizes propagation from the focusing to the control array, its transpose formalizes propagation in the re-

verse sense, and the complex conjugation accounts for the TR operation].

The focused signals $\mathbf{f_{TR}}(\omega)$ that are received on the control array after propagation of $\mathbf{e_{TR}}(\omega)$ can be expressed as

$$\mathbf{f_{TR}}(\omega) = \mathbf{H}(\omega)\mathbf{e_{TR}}(\omega) = \mathbf{H}(\omega)\mathbf{H}(\omega)^H\mathbf{s}(\omega). \quad (4)$$

The time reversal operator $\mathbf{H}(\omega)\mathbf{H}(\omega)^H$ is not necessarily close to the identity matrix. It may have strong nondiagonal elements. This means that when focusing, significant amounts of energy can be sent to control points other than the desired focal point: TR does not optimize contrast.[2] One practical advantage of TR is that it is a minimally invasive technique: in order to focus on a point P it is only necessary to have an ultrasound source at that point's location. In other words, although the whole $\mathbf{H}(\omega)$ matrix—that requires transducers or hydrophones at all the control points' locations for its measurement—has been used to formalize the TR operation, the measurement of only one row of $\mathbf{H}(\omega)$ (the row corresponding to the control point we want to focus on) is necessary for the practical implementation of TR.

C. Spatio temporal inverse filter focusing

The signals $\mathbf{e_{STIF}}$ to be emitted from the focusing array to focus by STIF on the target $\mathbf{s}$ on the control array are obtained by regularized inversion of the propagator $\mathbf{H}$ (the frequency dependence of the vector or matrix variables is no longer made explicit in order to have more compact equations)

$$\mathbf{e_{STIF}} = \hat{\mathbf{H}}^{-1}\mathbf{s}. \quad (5)$$

$\hat{\mathbf{H}}^{-1}$ is the regularized inverse of $\mathbf{H}$,[2] or equivalently its Moore–Penrose pseudoinverse.[10] It can also be expressed in the following way:

$$\hat{\mathbf{H}}^{-1} = (\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}\mathbf{H}^H \quad (6)$$

where $(\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}$ is the regularized inverse of $\mathbf{H}^H\mathbf{H}$. From Eqs. (3), (5), and (6), one can see that the STIF focusing signals can be expressed as a function of the TR focusing signals:

$$\mathbf{e_{STIF}} = (\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}\mathbf{e_{TR}}. \quad (7)$$

The STIF signals can be deduced from the TR signals by the application of the filter $(\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}$.

The STIF signals are emitted by the focusing array and propagate through the propagator $\mathbf{H}$ toward the control array. The focused signals $\mathbf{f_{STIF}}$ eventually received by the control array are

$$\mathbf{f_{STIF}} = \mathbf{H}\mathbf{e_{STIF}} = \mathbf{H}(\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{s}. \quad (8)$$

If L is the rank of $\mathbf{H}$ $[L \leqslant \min(M,N)]$, it is possible to write

$$\mathbf{H}(\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}\mathbf{H}^H = \mathbf{P_L} \quad (9)$$

where $\mathbf{P}_L$ is the projector on the L first eigenvectors of the matrix $\mathbf{H}$.[9]

$\mathbf{f_{STIF}}$ still differs from the optimal focusing that would be achieved by focusing through a homogeneous medium. Indeed, the wave field radiated by the source cannot be completely recorded on the focusing array because of the finite extent of that array. Information about the propagation medium is thus lost. However, it can be shown that $\mathbf{f_{STIF}}$ is the focusing pattern achievable with the focusing array that most closely approaches the target focusing $\mathbf{s}$ in the least-squares sense.[3] As a consequence, the STIF approach leads to lower sidelobes than the TR (less energy is received by the control points other than the focal point).[9] In other words, $\mathbf{P_L}$ has lower nondiagonal values than the time-reversal operator $\mathbf{H}\mathbf{H}^H$.

The drawback of the STIF technique is that the full measurement of the propagator $\mathbf{H}$ is necessary to compute the STIF focusing signals. It is thus a far more invasive approach than the TR that only requires the measurement of one row of $\mathbf{H}$.

D. The minimally invasive STIF

1. Principle of the minimally invasive STIF (miSTIF)

Equation (7) shows that the STIF signals can be deduced from the TR signals by the application of the filter $(\mathbf{H}^{\hat{H}}\mathbf{H})^{-1}$. This filter is expressed as a function of the propagator $\mathbf{H}$, it must thus *a priori* be measured invasively through the invasive measurement of $\mathbf{H}$. Although, the analysis of the physical meaning of $\mathbf{H}^H\mathbf{H}$ as a time reversal operator shows that it can be evaluated noninvasively. Indeed, one can show that, in a nonabsorbing (lossless) medium

$$\mathbf{H}^H\mathbf{H} \approx \mathbf{K} + \mathbf{K}^* \quad (10)$$

where $\mathbf{K}$ is the backscatter propagator that links the transducers of the focusing array between them. The "$\approx$" sign in Eq. (10) stands for "approximate equality." The meaning and extent of this approximation will be made clear in the following section.

Equation (10) suggests that, instead of invasively measuring $\mathbf{H}$ and inverting $\mathbf{H}^H\mathbf{H}$, one could in practice noninvasively measure $\mathbf{K}$ and invert $\mathbf{K}+\mathbf{K}^*$ to compute the STIF from the TR signals. These minimally invasive STIF focusing signals (miSTIF signals) would be expressed as

$$\mathbf{e_{miSTIF}} = (\mathbf{K}\hat{+}\mathbf{K}^*)^{-1}\mathbf{e_{TR}} \quad (11)$$

2. The relation $\mathbf{H}^H\mathbf{H} \approx \mathbf{K}+\mathbf{K}^*$

This section aims to show that the propagators $\mathbf{H}^H\mathbf{H}$ and $\mathbf{K}+\mathbf{K}^*$ can be identified for practical purposes, for any nonabsorbing propagation medium. Elements of the demonstration of the relationship $(\mathbf{H}^H\mathbf{H} \approx \mathbf{K}+\mathbf{K}^*)$ can be found in Ref. 4. Reference 11 also provides some insight, with a formalism and physical explanations that are closer to the present paper. Another demonstration of this relation with a more mathematical approach can be found in the Appendix of this paper.

The element $\mathbf{K}_{i'i}(\omega)$ of the backscatter propagator $\mathbf{K}(\omega)$ is the Fourier transform at frequency ω of the Green's function $\mathbf{k}_{i'i}(t)$ that links transducers i and i' of the focusing array. This signal can be obtained directly: transducer i emits a pulse, and $\mathbf{k}_{i'i}(t)$ is the signal then received by transducer i'. This signal can also be measured by simulating the pres-

ence of a source at the location of transducer i. This can be done by focusing a wave on transducer i by TR using the control array as a TR cavity:

(1) In the first step, transducer i emits a pulse, the vector emitted by the focusing array is

$$\mathbf{s}(\omega) = [0,0,\ldots,1,\ldots,0]^T \tag{12}$$

(this vector of size M is 0 on all transducers except on transducer i).

(2) The signals then recorded by the control array cavity after propagation through the heterogeneous medium formalized by the propagator $\mathbf{H}(\omega)$ are time reversed and sent back. The vector finally received on the focusing array is

$$\mathbf{f}(\omega) = \mathbf{H}(\omega)^T\mathbf{H}(\omega)^*\mathbf{s}(\omega) \Leftrightarrow \mathbf{f}(\omega) = [\mathbf{H}(\omega)^H\mathbf{H}(\omega)]^*\mathbf{s}(\omega). \tag{13}$$

As the control array is a closed TR cavity, the original impulsion is recreated at point i at $t=0$ (the focusing moment is arbitrarily taken as the origin of time). The focused wave continues its propagation and a diverging wave appears emanating from point i, as if point i had emitted a pulse. When this diverging wave arrives at point i', the measured signal at this point is thus the Green's function $\mathbf{k}_{i'i}(t)$ [Fourier transform $\mathbf{K}_{i'i}(\omega)$] between i and i'.

A signal is also received by transducer i' before the wave focuses ($t<0$). This signal is due to the focusing wave that is to converge at point i during the backwards phase of the time-reversal experiment. It is thus the time-reversed version of the signal that is sensed by i' in the forward phase of the time-reversal experiment, where i sends out a pulse that propagates toward the time-reversed cavity. In other words, it is the time-reversed version $\mathbf{k}_{i'i}(-t)$ [Fourier transform $\mathbf{K}_{i'i}(\omega)^*$] of the Green's function between i and i'. $\mathbf{f}(\omega)_{i'}$ is thus equal to the causal $\mathbf{K}_{i'i}(\omega)$ plus the anticausal $\mathbf{K}_{i'i}(\omega)^*$ Green's function between i and i'.

$\mathbf{f}(\omega)_{i'}$ is element i' of $[\mathbf{H}(\omega)^H\mathbf{H}(\omega)]^*\mathbf{s}(\omega)$, it is also element i',i of the matrix $[\mathbf{H}(\omega)^H\mathbf{H}(\omega)]^*$ as $\mathbf{s}(\omega)$ is the ith canonical vector. It comes that $[\mathbf{H}(\omega)^H\mathbf{H}(\omega)]^*_{i'i}$ is equal to $\mathbf{K}_{i'i}(\omega)+\mathbf{K}_{i'i}(\omega)^*$. As the same reasoning can be held for every pair (i',i) of transducers of the focusing array, it then comes that

$$[\mathbf{H}(\omega)^H\mathbf{H}(\omega)]^* = \mathbf{K}(\omega) + \mathbf{K}(\omega)^*. \tag{14}$$

Note that, from the above discussion, the temporal signal received by any tranducer i' during a time-reversed experiment on any transducer i is symmetrical in time, provided that a real time-reversal cavity is used. All the elements of $\mathbf{H}(\omega)^H\mathbf{H}(\omega)$ are thus real and Eq. (14) can be simplified to

$$\mathbf{H}(\omega)^H\mathbf{H}(\omega) = \mathbf{K}(\omega) + \mathbf{K}(\omega)^* \tag{15}$$

[note that in reality, there is a proportionality factor between the left-hand side and the right-hand side of Equation (15) (see Appendix). However, this proportionality factor is of no importance in practice].

In reality, the impulsion $\mathbf{f}(\omega)$ recreated on the focusing array to create a virtual source there (step 2) is not a perfect spatial Dirac function. Because of the diffraction limit, it is a two-dimensional sinc function whose main lobe width is equal to λ, the wavelength of the waves that are used.[8] The two ways (direct or indirect) of measuring the Green's function presented above are not rigorously equivalent because the virtual transducer cannot have the same characteristics as the real transducer. Thus, one can only write Eq. (15) with a "$\approx$" sign

$$\mathbf{H}(\omega)^H\mathbf{H}(\omega) \approx \mathbf{K}(\omega) + \mathbf{K}(\omega)^*. \tag{16}$$

The amount of the approximation depends on how close to a real transducer the virtual transducer can be considered to be:

If the medium is lossless, the amplitude received by any tranducer i' of the focusing array when doing a time-reversal experiment on transducer i is proportional to $\mathrm{sinc}(k|\mathbf{r}_i-\mathbf{r}_{i'}|)$, where $\mathbf{r}_i$ and $\mathbf{r}_{i'}$ denote the spatial positions of the transducers i and i', respectively.[8] If the focusing array is sampled at $\lambda/2$, this amplitude reduces to $\delta^2(\mathbf{r}_i-\mathbf{r}_{i'})$ where δ^2 stands for the two-dimensional Kronecker delta function. In that particular case, an equal sign can be used between the left-hand side and the right-hand side of Eq. (16). However, the focusing array cannot be sampled precisely at $\lambda/2$ for all frequencies within the useful bandwidth of the transducers. In general, the equality is thus not strict.

If the medium is absorbing, then the time-reversal experiment is no longer perfect. The control array can no longer be considered to be a "time-reversal cavity." The spatial extent of the focusing on the focusing array achieved by using the control array as a time-reversal device can no longer be described by a sinc function. Most likely, its mainlobe will spread out, and if the medium is not isotropic, its mainlobe and sidelobes will be assymetric. The "$\approx$" sign in Eq. (16) becomes even more of an approximation.

Now, it is instructing to compare the behavior of the STIF and the TR focusing techniques in the simple case of a homogeneous lossless medium. In that case, and if the focusing array is sampled at $\lambda/2$, it comes that $\mathbf{H}(\omega)^H\mathbf{H}(\omega)=\mathbf{I}$ ($\mathbf{I}$ stands for the identity matrix) and from Eq. (7), $\mathbf{e}_{\mathrm{STIF}}=\mathbf{e}_{\mathrm{TR}}$. The time reversal and the inverse filter focusing techniques yield the same focusing vector.

Equation (16) suggests that the *a priori* invasive evaluation of $\mathbf{H}(\omega)^H\mathbf{H}(\omega)$ through the measurement of $\mathbf{H}(\omega)$ may be in practice replaced by the noninvasive evaluation of $\mathbf{K}(\omega)+\mathbf{K}(\omega)^*$ through the measurement of $\mathbf{K}(\omega)$, for the derivation of the STIF focusing signals from the TR focusing signals. This hypothesis is tested in the next section with a numerical example.

III. NUMERICAL SIMULATION

In this section, the different focusing methods (time reversal, the conventional highly invasive STIF and the minimally invasive STIF) are tested using a numerical simulation. The propagation of waves is simulated using a bidimensional finite differences acoustic code developed in house.

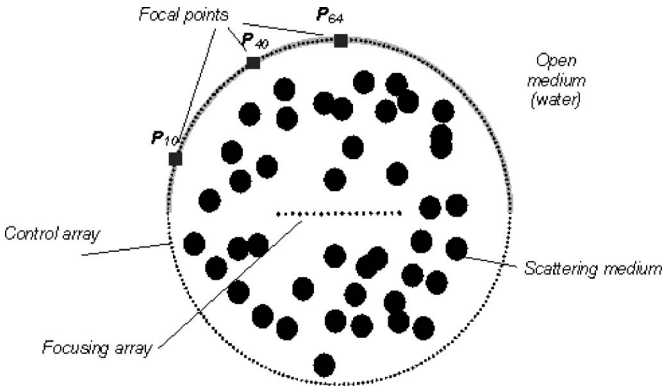


FIG. 2. Simulation setup. The focusing array lays in the middle of the scattering medium, which is surrounded by a time-reversal cavity of 256 control points. The focusing array is used to focus on three focal positions, that coincide with control points numbers 64 (on-axis of the focusing array), 40 and 10. The control points that are used to scan the lateral focusing profiles presented in Fig. 5 are the 128 control points located in the upper half plane delimited by the focusing array; they are highlighted in gray.

A. The simulated experiment setup

The simulated experiment setup is presented in Fig. 2: the focusing array is made of $M=40$ aligned monopolar transducers. The pitch of the array is about half the shortest wavelength used in the simulated experiment (0.25 mm). The central frequency of the transducers is 1.5 MHz, and their bandwidth at −6 dB is approximately 1.0–2.6 MHz. This array lays within a scattering medium of finite extension, made of 0.4-mm-diameter fluid disks randomly scattered in water. There are on average five disks per square millimeter, and the disks can overlap. The speed of sound in the disks is twice the speed of sound in water ($c_d=3$ mm/μs, $c_w=1.5$ mm/μs) and the density of the disks is three times the density of water ($d_d=3000$ kg/m^3, $d_w=1000$ kg/m^3). The disks are all located within an area delimited by a circle of radius 10 mm around the central transducer of the focusing array. A control array of 256 evenly spaced control points is set up along this circle. The pitch between the control points is thus approximately 0.25 mm, ensuring that the control array makes up a good time-reversal cavity. Beyond the area delimited by the time-reversal cavity, the medium is homogeneous with the properties (speed of sound and density) of water. No absorption is simulated.

The control array is used for the conventional, highly invasive derivation of the STIF focusing signals, and to spatially scan the pressure fields to test the focusing quality of the different methods.

B. The relation $\mathbf{H}^H\mathbf{H}\approx\mathbf{K}+\mathbf{K}^*$

Propagation from each transducer of the focusing array to each transducer of the focusing array and to each control point of the control array is simulated thanks to the finite-difference code to measure the propagators $\mathbf{K}$ and $\mathbf{H}$, respectively.

In order to numerically check the validity of Eq. (16) and have an idea of the amount of approximation suggested by the "≈" sign, one can compare one by one the elements of the matrices $\mathbf{H}(\omega)^H\mathbf{H}(\omega)$ and $\mathbf{K}(\omega)+\mathbf{K}(\omega)^*$. Figure 3 shows elements (i',i) of both propagators with $i=1$ and $i'=1,10,20,40$ (displayed in time domain). Element (i',i) of inverse Fourier transform (IFT) $\{\mathbf{K}(\omega)+\mathbf{K}(\omega)^*\}$ is obtained by sending a pulse from transducer i, recording the signal that arrives at transducer i', and adding to it the time-reversed version of it. Element (i',i) of IFT$\{\mathbf{H}(\omega)^H\mathbf{H}(\omega)\}$ is obtained by sending a pulse from transducer i, recording, time reversing, and re-emitting what arrives at the control array, and recording the signal that arrives then at transducer i'. IFT$\{\mathbf{K}(\omega)+\mathbf{K}(\omega)^*\}$ and IFT$\{\mathbf{H}(\omega)^H\mathbf{H}(\omega)\}$ superimpose very well, with an average correlation coefficient of 0.995.

The slight differences that are observed between these propagators come from the fact that a time-reversal experiment to focus on any transducer i of the focusing array, using the control array as a time-reversal cavity, does not lead to the creation of a perfect spatial Dirac function at the location of transducer i at the focusing time $t=0$. Instead, the energy is spread around the transducer that originally emitted the pulse. The spreading is determined by the diffraction theory: the focal spot is a two-dimensional sinc function centered on transducer i, whose main lobe width is equal to λ. The extent of the focusing can be indirectly visualized in Fig. 3 [it is best illustrated in Fig. 3(d)]. The dashed black curve in this figure is the temporal signal sensed at transducer 40 of the focusing array when using the control array as a time-reversal cavity to focus on transducer 1 of the focusing array. If the focusing was perfect, a perfect impulsion would be recreated at $t=0$ at the location of transducer 1. However, one can see in Fig. 3(d) that transducer 40 does sense signal at $t=0$. The original impulse is not perfectly recreated through the time-reversal experiment, even when using a perfect time-reversal cavity.

A more quantitative way to compare $\mathbf{H}(\omega)^H\mathbf{H}(\omega)$ and $\mathbf{K}(\omega)+\mathbf{K}(\omega)^*$ is to compare their singular values at all frequencies within the useful bandwidth (Fig. 4). This way of comparing both matrices makes direct sense, because the implementations of STIF and miSTIF require their respective regularized inversions, which are performed via a singular value decomposition.[2] If the eigenvalues of both propagators are equal, their regularized inverses are also equal and the STIF and miSTIF focusing signals are the same. Figure 4 shows the first, fifth, tenth and 20th eigenvalues of both propagators as a function of the frequency. The eigenvalues of $\mathbf{H}^H\mathbf{H}$ and $\mathbf{K}+\mathbf{K}^*$ superimpose well at the highest frequencies. At the lowest frequencies, the eigenvalues of $\mathbf{H}^H\mathbf{H}$ tend to drop. Indeed, the column rank of $\mathbf{H}$ (that determines the rank of $\mathbf{H}^H\mathbf{H}$) is equal to the number of independent focal spots that can be obtained on the transducers of the focusing array, using the control array to focus on it.[12] Two focal spots are said to be "independent" if their main lobes do not overlap. As the width of the focal spot achievable with the control array in a nonabsorbing medium is λ, the number L of independent focal spots that can be obtained on the focusing array is given by

$$L=\min(M,Mp/(\lambda/2)) \tag{17}$$

where M and p are the number of transducers and the pitch of the focusing array, respectively (Mp is the aperture size of the focusing array).

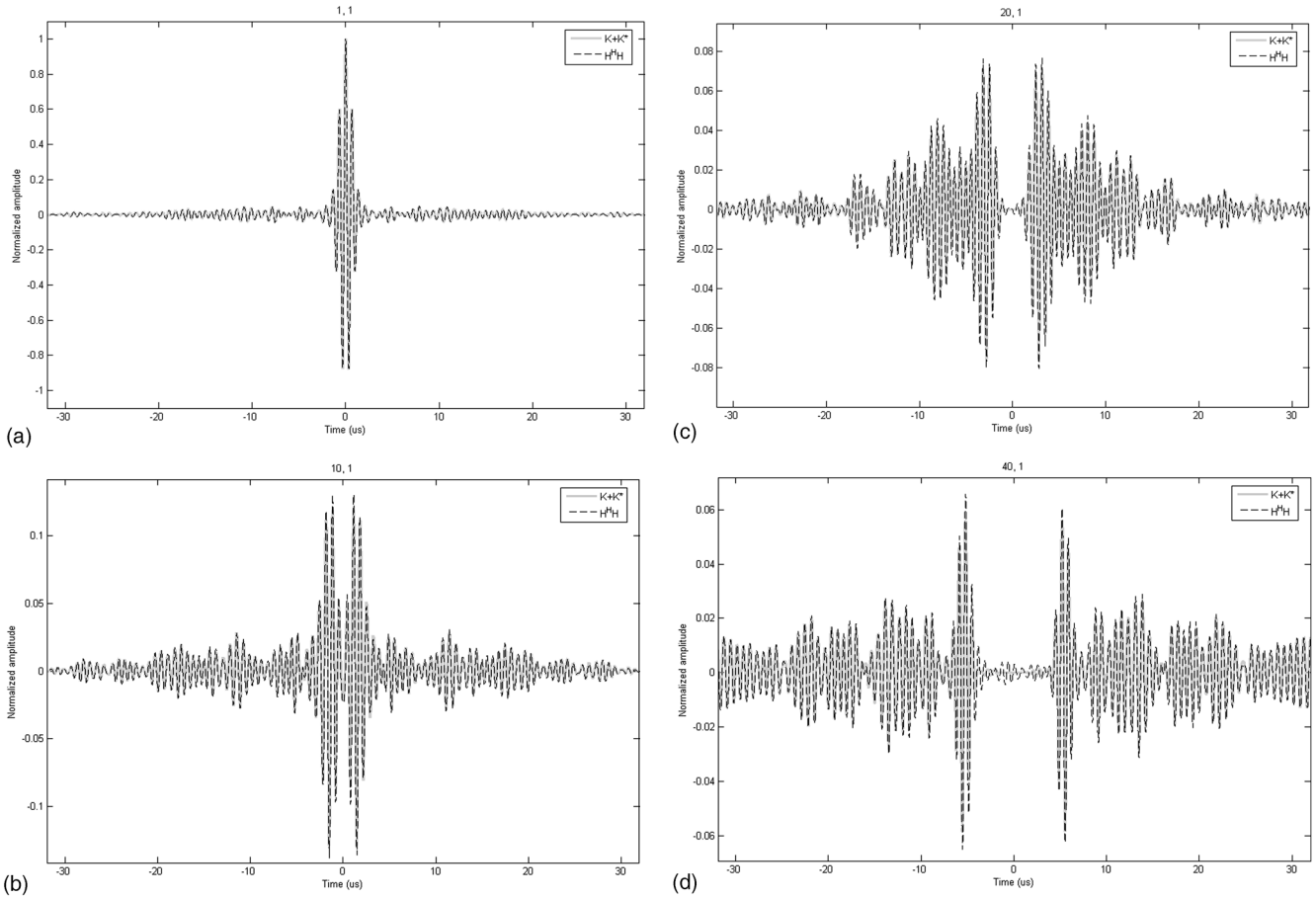


FIG. 3. (Color online) Validity of $\mathbf{H}(\omega)^H\mathbf{H}(\omega) \approx \mathbf{K}(\omega)+\mathbf{K}(\omega)^*$. The full light gray curves are the Fourier transforms of elements (a), (b), (c), and (d) of $\mathbf{K}+\mathbf{K}^*$. The Fourier transforms of the corresponding elements of $\mathbf{H}^H\mathbf{H}$ are superimposed in dashed black.

As p has been chosen to be approximately equal to half the shortest wavelength used in the simulation, the ratio $p/(\lambda/2)$ is always lower than 1. The rank of $\mathbf{H}^H\mathbf{H}$ is thus less than M at all the working frequencies. Equation (17) also shows that the rank of $\mathbf{H}^H\mathbf{H}$ decreases with the increasing wavelength (decreasing frequency). Indeed, fewer independent focal spots can be obtained on the focusing array at lower frequencies because the extent of the focusing is wider at these frequencies. This explains the drop in the highest-indexed eigenvalues of $\mathbf{H}^H\mathbf{H}$, and the fact that this drop begins sooner at lowest frequencies. The corresponding eigenvalues fall in the null space of $\mathbf{H}^H\mathbf{H}$.

On the other hand, the rank of $\mathbf{K}$ is M at all frequencies. Indeed, one can always obtain M independent focal spots on the transducers of the focusing array, using that same array to focus on itself. The focal spot on transducer i is obtained by having transducer i only emit a pulse (the focal spot and the emission vector are the same in that case in a first-order approximation: they only differ by the signal backscattered by the medium to the focusing array).

C. Focusing

The propagators $\mathbf{H}$, $\mathbf{H}^H\mathbf{H}$, and $\mathbf{K}+\mathbf{K}^*$ are used to compute the TR focusing signals [Eq. (3), one row of $\mathbf{H}$ is needed], the conventional, highly invasive STIF focusing signals [Eq. (5), the full invasive measurement of $\mathbf{H}$ is needed], and the minimally invasive STIF signals [Eq. (11), one row of $\mathbf{H}$ and the noninvasive measurement of $\mathbf{K}$ are needed]. The emitted signals are numerically propagated and the signals recorded on the 128 upper control points surrounding the focal point P are scanned (Fig. 5). In order to test the lateral focusing quality, the temporal maxima of the acoustic pressure received on these control points are displayed in a dB scale [Figs. 5(a)–5(c)]. In order to test temporal focusing quality, the envelope of the temporal signal received at the focal point is displayed in a dB scale [Figs. 5(d)–5(f)]. Three focal positions are tested: one on-axis of the focusing array (control point number 64) and two off-axis (control point numbers 40 and 10), in order to show that the method presented here can be used to focus at any location in space.

For all three focal locations, both lateral and temporal sidelobes are lower with the STIF techniques than with the TR. The STIF also leads to a better temporal resolution than the TR. The minimally invasive STIF yields a focusing pattern very similar to the conventional STIFs, showing that the propagators $\mathbf{H}^H\mathbf{H}$ and $\mathbf{K}+\mathbf{K}^*$ can be considered to be identical in practice.

The lateral sidelobes level is reduced by up to 10 dB when using the STIF techniques compared to the TR. The lateral resolution is very good for the three techniques. Indeed, the multiple scattering induces a super-resolution effect

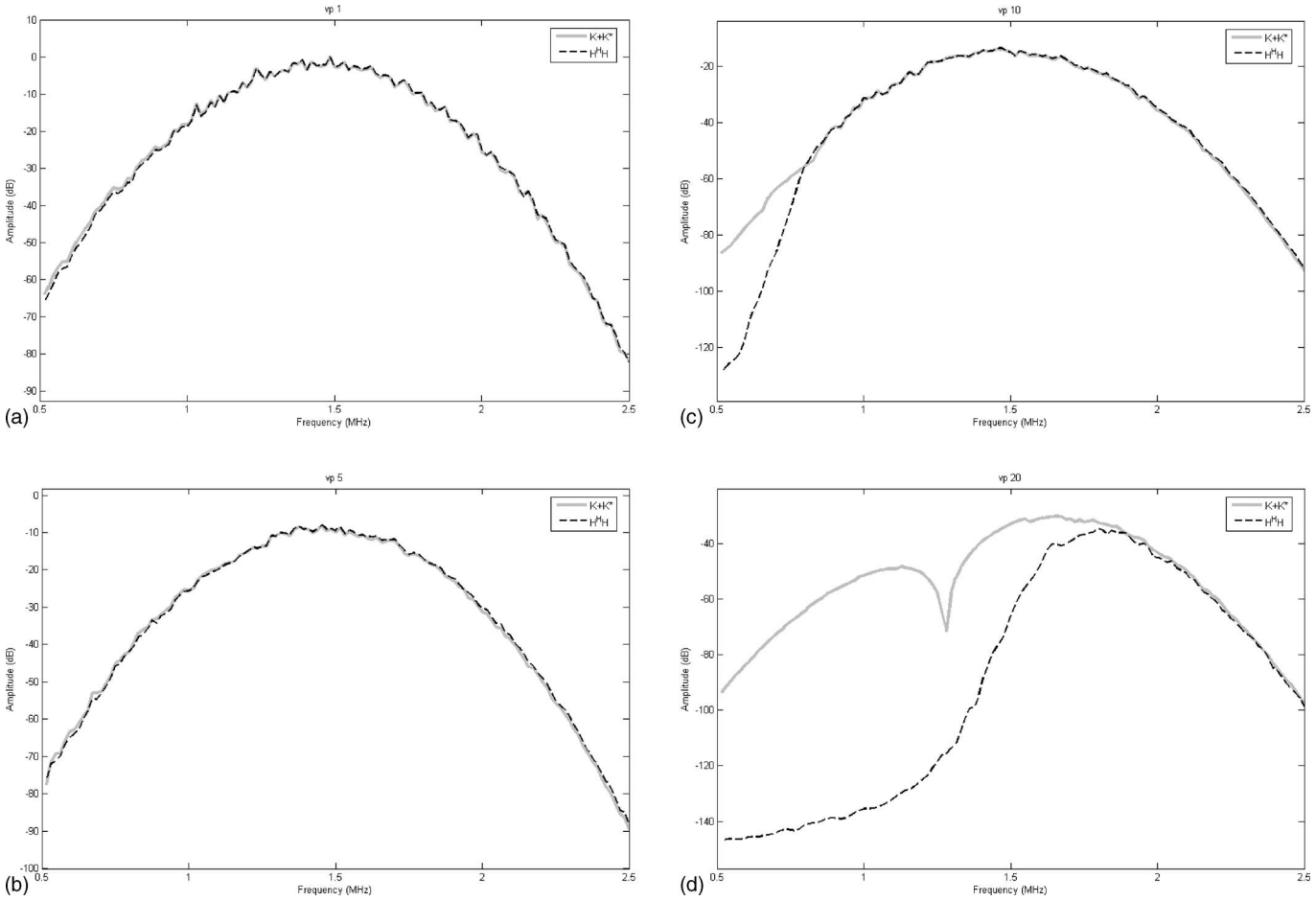


FIG. 4. Validity of $\mathbf{H}(\omega)^H\mathbf{H}(\omega) \approx \mathbf{K}(\omega)+\mathbf{K}(\omega)^*$. Each subpanel presents a plot of an eigenvalue of $\mathbf{K}(\omega)+\mathbf{K}(\omega)^*$ (full light gray curves) and $\mathbf{H}(\omega)^H\mathbf{H}(\omega)$ (dashed black curves) as a function of the frequency. The first, fifth, tenth and 20th eigenvalues are plotted in subpanels (a), (b), (c), and (d), respectively.

(a resolution sharper that the resolution that would be achieved with a similar focusing array through a homogeneous medium).[13] Notice that, surprisingly, the focusing quality is better on off-axis control points Nos. 10 and 40 than on the on-axis control point No. 64 (lower sidelobes). This result would be paradoxical in a homogeneous medium where one would expect a better focusing on the on-axis control point. Indeed, the array aperture seen by the on-axis control point is wider than the array aperture seen by the off-axis control points. However, in a multiple scattering medium, the focusing quality is no longer governed by the array aperture size but rather by an effective aperture created by the scatterers.[13] The focusing quality on a given control point thus depends chiefly on the spatial distribution of the scatterers that are closest to that control point.

The temporal sidelobes are up to 10 dB lower with the STIF techniques than with the TR. The temporal resolution is better for the STIF techniques than for the TR (the main lobe temporal width at −6 dB is 30% shorter with the STIF techniques than with the TR). Indeed, The STIF techniques compensate for the finite bandwidth of the transducers.

The nonadaptive focusing achieved when driving the focusing array with geometrically computed delays (assuming a homogeneous speed of sound in the medium) is not shown, as it is of comparatively extremely bad quality.

IV. DISCUSSION

The proof that the backscatter propagator $\mathbf{K}$ can be used to change the TR focusing signals into the STIF focusing signals is general and valid for any nonabsorbing medium.

However, the fact that the miSTIF is theoretically only effective in nonabsorbing media limits its practical interest. Indeed, inverse filter techniques are especially interesting in absorbing media because they can compensate for information losses that the TR technique is unable to recover.[12] In absorbing media, the slightly less efficient but more robust TR focusing technique might be preferred in practice. Further work should assess the robustness of Eq. (16) to losses in the propagation medium.

It is instructing to compare the behavior of the STIF and the TR focusing techniques in the simple case of a homogeneous lossless medium. In that case, and if the focusing array is sampled at $\lambda/2$, it comes that $\mathbf{H}^H\mathbf{H}=\mathbf{K}+\mathbf{K}^*=\mathbf{I}$ ($\mathbf{I}$ stands for the identity matrix) and from Eq. (7), $\mathbf{e}_{\mathrm{STIF}}=\mathbf{e}_{\mathrm{TR}}$. The time reversal and the inverse filter focusing techniques yield the same focusing vector. In general, in a weakly heterogeneous medium, and/or if the heterogeneities are far from the focusing array, the matrix $\mathbf{H}^H\mathbf{H}\approx\mathbf{K}+\mathbf{K}^*$ is dominated by its diagonal elements ($\mathbf{H}^H\mathbf{H}\approx\mathbf{K}+\mathbf{K}^*\approx\mathbf{I}$) and the STIF is not expected to yield a significantly improved focusing quality compared to the TR ($\mathbf{e}_{\mathrm{STIF}}\approx\mathbf{e}_{\mathrm{TR}}$).

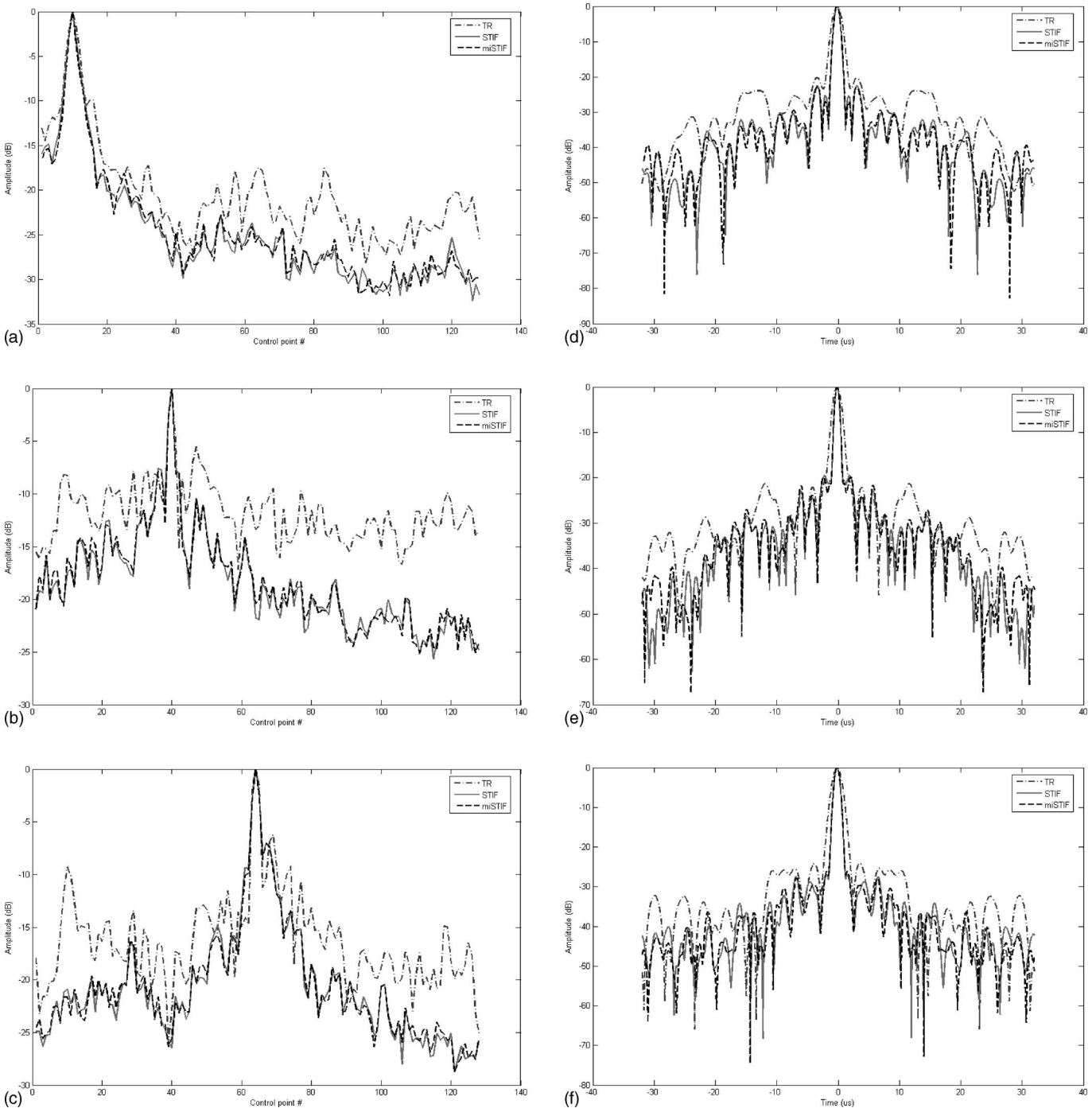


FIG. 5. Lateral and temporal focusing through the multiple scattering medium: (a), (b), (c) lateral focusing: the temporal maximum of the pressure received on the upper control points (control points from 1 to 128, highlighted in gray in Fig. 2) is plotted in dB. As the focusing array is an axis of symmetry for the experiment, the pressure scan on only the upper control points is represented. Medium gray, dash-dotted: TR focusing. Light gray, full: STIF focusing. Black, dashed: miSTIF focusing. (d), (e), (f) temporal focusing: the envelope of the temporal signal received at the focal point is plotted in dB. The focal points considered are control point number 10 (a),(d), number 40 (b),(e), number 64 (c),(f) (cf Fig. 2).

It is now necessary to emphasize here an implicit requirement that has to be met by the focusing array. This array must be able to correctly measure the backscatter propagator $\mathbf{K}$ that is needed to implement the miSTIF method. Recall that that propagator contains in any position (i',i) the signal received by transducer i' of the focusing array when transducer i of the focusing array emits a pulse. This signal is comprised by: (1) First, the signals that propagate directly from i to i'. Such signals should be recorded if transducer i' lays within the directivity pattern of transducer i *and* in the particular case where $i=i'$. (2) Second, the signals backscattered by the medium from i to i'. The two parts of the signal can clearly be seen in Fig. 3 where a first (direct) arrival of high amplitude is followed by the longer arrival corresponding to backscatter by the medium. In that case the two parts of the signal overlap because the backscattering medium is very close to the array. Both direct and backscattered signals must be entirely recorded when measuring $\mathbf{K}$, in order for

Eq. (15) to be valid. This is always true for the diagonal elements of $\mathbf{K}$ (case $i=i'$): in practice, it means that when any element of the focusing array emits, it should be able to record, or at least estimate precisely, the pressure waveform it is sending into the medium.

We tried to implement the miSTIF technique experimentally. The focusing array was a linear array of 64 transducers spaced at lambda/2, such as those used in medical imaging. Because of the "dead time" of the transducers after emission (there is a short period of time after sending any signal from the array when the array cannot record any signal), we were unable to record the direct signals and thus could not get the miSTIF to work.

V. CONCLUSIONS

A new adaptive focusing technique, the minimally invasive spatio-temporal inverse filter (miSTIF), has been presented. It allows one to focus waves through a heterogeneous medium with the quality of the conventional spatio-temporal inverse filter (contrast optimization) with an experimental setup as minimally invasive as the one used for time reversal: only one acoustic source is needed at the desired focus location. This is achieved by unveiling and using a fundamental link that exists between the spatio-temporal inverse filter and the time reversal focusing techniques.

Although this new minimally invasive inverse filter technique is appealing, some limitations do exist. It is less robust than the time reversal, and its application is theoretically limited to nonabsorbing media. Furthermore, in weakly heterogeneous media, it becomes equivalent to TR. Finally, it is difficult to implement with standard ultrasound equipment such as the transducer arrays traditionally used in medical imaging.

ACKNOWLEDGMENTS

The authors thank Michaël Tanter for providing the finite-differences code for the simulation and help in using it. The authors thank Horacio Pastawski, Oliver Dorn, and the reviewers for constructive discussions.

APPENDIX

This Appendix aims to give another demonstration of Eq. (15), with a mathematical approach.

The wave equation in frequency domain for the acoustic pressure field $p_i(\mathbf{r},\omega)$ generated by a pulsating sphere of radius a ($a \ll \lambda$) at $\mathbf{r}=\mathbf{r}_i$, pulsating at frequency ω, is[14]

$$\nabla\left(\frac{\nabla p_i}{\rho}\right)+\frac{k^2 p_i}{\rho}=-i\omega 4\pi U a^2 \delta^3(\mathbf{r}-\mathbf{r}_i) \tag{A1}$$

where $\rho(\mathbf{r})$ is the local medium density; $k=\omega/c$; and $\lambda=2\pi/k$. U is the Fourier transform of the speed of the radius of the pulsating sphere. The pulsating sphere at $\mathbf{r}=\mathbf{r}_i$ is one of the transducers of the focusing array. In Eq. (A1) and in the following equations, the spatial and frequency dependencies of physical quantities (ρ, p_i, and U) are not made explicit in order to have more compact notations.

The equation for the phase conjugate (time-reversed version) of a pressure field $p_{i'}(\mathbf{r},\omega)$ generated by an identical pulsating sphere at $\mathbf{r}=\mathbf{r}_{i'}$ is

$$\nabla\left(\frac{\nabla p_{i'}^*}{\rho}\right)+\frac{k^2 p_{i'}^*}{\rho}=+i\omega 4\pi U a^2 \delta^3(\mathbf{r}-\mathbf{r}_{i'}) \tag{A2}$$

(here we assume that the initial excitation U is real).

Multiplying Eq. (A2) by $p_{i'}/\rho$ and subtracting Eq. (A1) multiplied by p_i/ρ yields

$$\frac{p_i}{\rho}\nabla\left(\frac{\nabla p_{i'}^*}{\rho}\right)-\frac{p_{i'}^*}{\rho}\nabla\left(\frac{\nabla p_i'}{\rho}\right)=\frac{i\omega 4\pi U a^2}{\rho}[p_{i'}^*\delta^3(\mathbf{r}-\mathbf{r}_i)+p_i\delta^3(\mathbf{r}-\mathbf{r}_{i'})]. \tag{A3}$$

A nabla in the left-hand side of Eq. (A3) can be factorized

$$\nabla\left[\frac{p_i}{\rho}\left(\frac{\nabla p_{i'}^*}{\rho}\right)-\frac{p_{i'}^*}{\rho}\left(\frac{\nabla p_i}{\rho}\right)\right]=\frac{i\omega 4\pi U a^2}{\rho}[p_{i'}^*\delta^3(\mathbf{r}-\mathbf{r}_i)+p_i\delta^3(\mathbf{r}-\mathbf{r}_{i'})]. \tag{A4}$$

Integrating over a closed volume containing the two pulsating spheres located at $\mathbf{r}_i$ and $\mathbf{r}_{i'}$ and using the Green's theorem,

$$\oiint\left[\frac{p_i}{\rho}\left(\frac{\nabla p_{i'}^*}{\rho}\right)-\frac{p_{i'}^*}{\rho}\left(\frac{\nabla p_i}{\rho}\right)\right]\mathbf{n}\,dS=i\omega 4\pi U a^2\left[\frac{p_{i'}^*(\mathbf{r}_i)}{\rho(\mathbf{r}_i)}+\frac{p_i(\mathbf{r}_{i'})}{\rho(\mathbf{r}_{i'})}\right] \tag{A5}$$

where the left-hand side is a surface integral over the surface of the above-mentioned integration volume, and $\mathbf{n}$ is a unitary vector perpendicular to that surface, and pointing outwards. Let us now imagine that this surace corresponds to the delineation of the time-reversal cavity that surrounds the focusing array.

If we assume that, on the closed surface, the wave is locally planar and the wave vector is approximately perpendicular to the closed surface, one can use the following approximations:

$$\mathbf{n}\,.\,\nabla p_i \approx -ikp_i$$

$$\mathbf{n}\,.\,\nabla p_{i'}^* \approx +ikp_{i'}^* \tag{A6}$$

and Eq. (A5) can be simplified

$$2ik \oiint \frac{p_i p_{i'}^*}{\rho^2} dS = i\omega 4\pi U a^2 \left[\frac{p_{i'}^*(\mathbf{r}_i)}{\rho(\mathbf{r}_i)} + \frac{p_i(\mathbf{r}_{i'})}{\rho(\mathbf{r}_{i'})} \right]$$

$$\Leftrightarrow \oiint \frac{p_i p_{i'}^*}{\rho^2} dS = 2\pi c U a^2 \left[\frac{p_{i'}^*(\mathbf{r}_i)}{\rho(\mathbf{r}_i)} + \frac{p_i(\mathbf{r}_{i'})}{\rho(\mathbf{r}_{i'})} \right]. \quad \text{(A7)}$$

The continuous integral can be replaced by a discrete sum

$$\oiint \frac{p_i(\mathbf{r}) p_{i'}^*(\mathbf{r})}{\rho(\mathbf{r})^2} d^2\mathbf{r} \approx \sum_j \frac{p_i(\mathbf{r}_j) p_{i'}^*(\mathbf{r}_j)}{\rho(\mathbf{r}_j)^2} \delta S \quad \text{(A8)}$$

where δS is the elementary surface of a transducer of the time-reversal cavity, and the index j indexes the elements of the time-reversal cavity (the control points).

Equation (A7) can then be rewritten as

$$\sum_j \frac{p_i(\mathbf{r}_j) p_{i'}^*(\mathbf{r}_j)}{\rho(\mathbf{r}_j)^2} = \frac{2\pi c U a^2}{\delta S} \left[\frac{p_{i'}^*(\mathbf{r}_i)}{\rho(\mathbf{r}_i)} + \frac{p_i(\mathbf{r}_{i'})}{\rho(\mathbf{r}_{i'})} \right]. \quad \text{(A9)}$$

Now, let us introduce the vectors

$$\mathbf{h}_i = \left\{ \frac{p_i(\mathbf{r}_j)}{\rho(\mathbf{r}_j)} \right\}_{j=[1\ldots N]}, \quad \mathbf{h}_{i'} = \left\{ \frac{p_{i'}(\mathbf{r}_j)}{\rho(\mathbf{r}_j)} \right\}_{j=[1\ldots N]} \quad \text{(A10)}$$

and the quantity

$$\mathbf{K}_{ii'} = \frac{p_{i'}^*(\mathbf{r}_i)}{\rho(\mathbf{r}_i)}. \quad \text{(A11)}$$

Equation (A9) can be rewritten as

$$\mathbf{h}_{i'}^H \mathbf{h}_i = \frac{2\pi c U a^2}{\delta S} (\mathbf{K}_{ii'} + \mathbf{K}_{i'i}^*). \quad \text{(A12)}$$

In a reciprocal medium, we have[15]

$$\frac{p_{i'}(\mathbf{r}_i)}{\rho(\mathbf{r}_i)} = \frac{p_i(\mathbf{r}_{i'})}{\rho(\mathbf{r}_{i'})}. \quad \text{(A13)}$$

It comes then that

$$\mathbf{h}_{i'}^H \mathbf{h}_i = \frac{2\pi c U a^2}{\delta S} (\mathbf{K}_{i'i} + \mathbf{K}_{i'i}^*). \quad \text{(A14)}$$

Generalizing Eq. (A14) for any pair (i, i') of transducers of the focusing array, it comes

$$\mathbf{H}^H \mathbf{H} \propto \mathbf{K} + \mathbf{K}^* \quad \text{(A15)}$$

where $\propto$ stands for "is proportional to." Equation (A15) is Eq. (15) of the main text.

[1] M. Fink, "Time reversal of ultrasonic fields—Part I: Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(5), 555–566 (1992).
[2] 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).
[3] O. Haddadin and E. Ebbini, "Ultrasonic focusing through inhomogeneous media by application of the inverse scattering problem," J. Acoust. Soc. Am. **104**(1), 313–325 (1998).
[4] O. Lobkis and L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**(6), 3011–3017 (1991).
[5] H. Pastawski, E. Danieli, H. Calvo, and L. Foa Torres, "Towards a time-reversal mirror for quantum systems," Europhys. Lett. **77**, 40001 (2007).
[6] H. Calvo, E. Danieli, and H. Pastawski, "Time reversal mirror and perfect inverse filter in a microscopic model for sound propagation," Physica B (Amsterdam) **398**, 317 (2007).
[7] O. Dorn, "Time-reversal and the adjoint method with an application in telecommunication," arXiv:math.OC/0412379v1 (19 Dec 2004).
[8] D. Cassereau and M. Fink, "Time reversal of ultrasonic fields—Part III: Theory of the closed time-reversal cavity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(5), 579–592 (1992).
[9] Throughout this paper the transducers of the focusing array will be indexed with "i"s ($1 \leqslant i < M$) and the control points of the control array with "j"s ($1 \leqslant j < N$).
[10] S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations* (Dover, New York, 1991).
[11] A. Derode, E. Larose, M. Campillo, and M. Fink, "How to estimate the Green's function of a heterogeneous medium between two passive sensors? Application to acoustic waves," Appl. Phys. Lett. **83**(15), 3054–3056 (2003).
[12] M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**(1), 223–234 (2000).
[13] A. Derode, P. Roux, and M. Fink, "Robust acoustic time-reversal with high-order multiple scattering," Phys. Rev. Lett. **75**(23), 4206–4209 (1995).
[14] L. E. Kinsler, *Fundamentals of Acoustics* (Wiley, New York, 1982), Chap. 8, p. 163.
[15] L. D. Landau, *Fluid Mechanics* (Addison-Wesley, Reading, MA, 1959).

Distortion product otoacoustic emissions and basilar membrane vibration in the 6–9 kHz region of sensitive chinchilla cochleae

William S. Rhode[a)]
Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706

(Received 29 May 2007; revised 22 August 2007; accepted 22 August 2007)

Distortion product otoacoustic emissions (DPOAEs) and basilar membrane (BM) vibration were measured simultaneously in the 6–9 kHz region of chinchilla cochleae. BM-Input-Output functions in a two-tone paradigm behaved similarly to DPOAEs for the 2f1–f2 component, nonmonotonic growth with the intensity of the lower frequency primary and a notch in the functions around 60 dB SPL. Ripples in frequency functions occur in both BM and OAE curves as a function of the distortion frequency. Optimum f2/f1 ratios for DPOAE generation are near 1.2. The slope of phase curves indicates that for low $f2/f1(<1.1)$ the emission source is the place location while for $f2/f1>1.1$ the relative constancy of the phase function suggests that the place is the nonlinear region of f2, i.e., the wave location. Magnitudes of the DPOAEs increase rapidly above 60 dB SPL suggesting a different source or mechanism at high levels. This is supported by the observation that the high level DPOAE and BM-DP responses remain for a considerable period postmortem.
 [DOI: 10.1121/1.2785034]

PACS number(s): 43.64.Kc [BLM] Pages: 2725–2737

I. INTRODUCTION

With the discovery of sound being emitted from the ear by Kemp (1978), the use of otoacoustic emissions (OAEs) to determine cochlear status has been a major focus of numerous individuals. It is generally understood that emissions provide a measure of the status of the so-called cochlear amplifier, which reflects the active properties of the outer hair cells (OHCs). Hence, the hope is that understanding the OAE generation mechanism may lead to methods for assessing the status of the cochlea in terms of which OHCs are present and functional.

Intertwined with the early story of otoacoustic emissions is that of the discovery of a compressive nonlinearity as observed in the vibration of the basilar membrane of squirrel monkeys (Rhode, 1971). This result was later confirmed in guinea pigs and chinchilla (Sellick *et al.*, 1982; Robles *et al.*, 1986a, b). Identification of active motile properties in the outer hair cells (Brownell *et al.*, 1984) focused future studies on them as the source of the nonlinear properties of the cochlea that result in its excellent sensitivity and wide dynamic range of ~120 dB. The amplification of basilar membrane (BM) motion due to the active properties of the outer hair cells is typically ~60 dB though a single instance of ~80 dB has been reported (Rhode, 2007a). The difference between the vibration of the basilar membrane in an active cochlea and one in which the OHCs are inactive (passive cochlea) is around 60 dB (Robles and Ruggero, 2001). A correlate of a variety of insults to the cochlea is that emissions from low level stimuli are largely abolished while for levels >70 dB SPL a component remains (Whitehead *et al.*, 1992; Avan *et al.*, 2003).

While there are several kinds of emissions, spontaneous (SOAE), stimulus frequency (SFOAE), transient (TOAE), and distortion product (DPOAE), it is the latter type that have been most extensively studied. In the study of DPOAEs, two primary tones at frequencies f1 and f2 are presented that result in distortion products at frequencies $(n+1)f1-nf2$ and $(n+1)f2-nf1$, with $f2>f1$. These DPOAEs do not exist in the input signal, therefore they must be generated by nonlinearities present in the cochlea. Because OHCs are principally responsible for nonlinear compression and sensitivity of basilar membrane motion (Robles and Ruggero, 2001), the use of the noninvasive technique of recording DPOAEs to assess their functionality has been used extensively (Kemp, 2002). In particular, because DPOAEs are significantly larger in rodents than in humans, rodents have been frequently used to unravel the complexities of emissions.

There are several interpretations of the emission process, but one overriding view is that there are two sources of DPOAEs in the cochlea. The most frequently studied DP is at $2f1-f2$ as it is usually the dominant component. DPOAEs appear to be generated in the compressive or nonlinear cochlear region where f2 is the characteristic frequency. Each DP component may also propagate to its frequency-appropriate location along the cochlear partition. Each of the two locations can return some of the energy back toward the stapes and out into the ear canal, the emission. These two sources have been labeled wave- and place-fixed emissions by Kemp (2002) and distortion and reflection emissions by Shera and Guinan (1999).

Because there are considerable delays in the emissions recorded in the ear canal relative to the input sounds, delays about double the travel time to the site of the generation, it was hypothesized that the energy made its way to the stapes

[a)]Electronic mail: rhode@physiology.wisc.edu

via a backward traveling wave (Kemp, 1980). However, recent studies of DPOAEs and travel times for the distortion components in the basilar membrane have led to the proposal that the backward energy transfer is via a fast compression wave (Ren, 2004), a view that is further supported by data from SFOAEs and basilar membrane vibration delays (Siegel *et al.*, 2005).

One difference between the BM-DP and DPOAE responses is that the latter are generated from a wider spatial range in the cochlea than the single site at which the BM-DP is measured. The present study attempts to contribute to the understanding of the relation of vibration of the basilar membrane to the generation of OAEs by simultaneously measuring DPOAEs and BM vibration in response to two-tone stimuli. Input-output functions and frequency transfer functions for the two measures exhibit common features that indicate a close relation between them.

II. METHODS

Methods are essentially those detailed in Cooper and Rhode (1992). Sixteen chinchilla cochleae were studied at points between 3.5 and 4.25 mm from the basal end of the basilar membrane where the characteristic frequency (CF) is between 5.3 and 9.1 kHz. CF is defined as the frequency that yields the largest response at threshold. All procedures were approved by the Animal Care and Use Committee of the University of Wisconsin.

Each animal was anesthetized with pentobarbital using a dose rate of 75 mg/kg. Additional doses were administered to maintain the animal in a deeply areflexive state. All anesthetics were administered intraperitoneally. A tracheotomy was performed to ensure an open airway and to place the animal on a respirator if necessary, though it was never used in this set of experiments. After the ear was surgically removed, four screws were implanted in the skull and cemented in with dental cement in order to form a rigid base. A bolt was then cemented to the base to provide a stable fixation of the skull to a head holder with six degrees of freedom for the purpose of positioning the cochlea under the microscope.

Cochlear sensitivity at CF was used as the exclusive monitor of the state of the preparation. The sensitivity of the preparation at CF was measured by repeated basilar membrane vibration input-output (IO) functions throughout the experiment and in a few instances was maintained over a 12 h period. Measurements were terminated if there was more than a 6–8 dB decrease in basilar membrane sensitivity at CF.

The overlying cochlear bone in a region with CFs between 6 and 9 kHz was shaved down using a microdrill with a 0.45 mm burr until the remaining tissue and/or bone debris could be removed with a pick fabricated out of a microelectrode and/or the use of a stream of artificial perilymph to flush the debris. Gold-coated polystyrene beads 25 μm in diameter served as retroreflectors. They were placed in the perilymph and allowed to sink to the basilar membrane. They have a specific gravity of 1.05, which is near that of water (1.0) and therefore any loading of the basilar membrane by the bead should be minimal. To support that belief, Cooper (1999) showed that there are no detectable differences between the bead and no-bead measurements of BM vibration. A glass cover slip was placed over the cochlear opening with no hydromechanical seal. The cover glass served to avoid the problem of an unstable air-fluid interface.

An opening in the bony ear canal, immediately lateral to the tympanic membrane, was made so that either the plastic tube attached to an Etymotic phone (ER10C) or an acoustically calibrated probe tube to which a $\frac{1}{2}$ in. Bruel&Kjaer condenser microphone is adjoined, could be visualized as it was positioned parallel to the tympanic membrane within 1 mm of the tip of the malleus. The opening was sealed with a glass cover after a 45 μm bead was placed on the tympanic membrane at the tip of the malleus (or umbo). The bead was used as a retroreflector for the interferometer and allowed the measurement of the transfer function of the malleus. Two sound sources were used, an Etymotic ER10C phone which was also used to acoustically calibrate the delivery system, or a pair of Schure 5C phones which were calibrated using the B&K probe tube microphone. Stimuli were presented using TDT System I and purpose-built software.

A. Single-tone experiments

Basilar membrane vibration IO functions were determined using 30 ms tone bursts with 1 ms raised cosine rise and fall times, presented every 100 ms. Stimulus amplitude typically covered a 100 dB SPL range (0–100 dB SPL) in 5 dB SPL steps. Eight-to-ten responses were averaged for each stimulus condition. Frequency transfer functions were constructed from the IO functions, providing the CF of the BM location under study. CF is defined as the frequency for which the first noticeable response can be seen as the stimulus intensity of a single tone is increased.

In a sensitive cochlea, BM-IO functions are compressively nonlinear in the vicinity of CF. It is this nonlinear compression that is responsible for the generation of distortion in the resulting vibration of the BM. Nonlinear compression is hypothesized to be the result of either prestin molecules in the OHC wall or due to the nonlinear properties of the stereocilia and the transduction channels. Regardless of the actual mechanism, the nonlinear forces are coupled to the basilar membrane and as such are reflected in its vibration.

B. Two-tone experiments

In two-tone suppression experiments, the two applied tones are typically referred to as primary tones and are designated f1 and f2. f1 is referred to as the suppressor tone or frequency while f2 is referred to as the probe tone and is usually set to the CF of the cochlear region under study. Each of the two primary tones was presented through a separate phone coupled with plastic tubing to the earpiece. Stimuli were 50 ms in duration and repeated every 175 ms with the "suppressor tone" delayed 10 ms in onset, the sequence was repeated twice. 1 ms rise/fall times were used. Intensity often covered 20–80 dB SPL in either 5 or 10 dB steps. Typically, suppressor frequency ranged approximately from one-half octave below to one-half octave above the CF in steps of

300 Hz. This range was chosen because it is frequently used in the study of the cubic difference tone in otoacoustic emission measurements. The nonlinear properties of the cochlea result in energy at frequencies not present in a two-tone stimulus at frequencies $=(n+1)f1-nf2$ and $(n+1)f2-nf1$, and at the difference frequency f2–f1. The largest components correspond to $2f1-f2$ and $2f2-f1$ frequencies and are sometimes referred to as the cubic distortion products. The lower frequency component, 2f1–f2, results in the largest emission and is typically the focus of clinical studies. In this study, distortion products at 2f1–f2, 2f2–f1, 3f1–2f2, and f2–f1 were routinely analyzed as they may arise from different mechanisms and/or place of generation.

In a second two-tone paradigm, the frequency ratio, f2/f1, was specified and the two primary tones were swept across a wide range of frequencies and were presented simultaneously. The ratio could also be varied and was typically limited to: $1.05<f2/f1<1.45$. Analysis was focused on the cubic distortion product, 2f1–f2. However, the distortion products: 2f2–f1, 3f1–f2, and f2–f1 were routinely included in the data analysis, because along with 2f1–f2 they were the largest distortion components in both the mechanical and the emission responses.

C. Recording system

Mechanical responses were measured using a custom-built displacement-sensitive heterodyne laser interferometer (Cooper and Rhode, 1992). The laser was coupled to the preparation using a long working distance lens (Nikon SLWD 5X, NA 0.1). The laser was focused to an area of ~ 5 μm diameter on the reflective beads. The interferometer was not sensitive enough to measure basilar membrane vibration without the gold-coated beads. Instantaneous phase was measured using two single-cycle phase meters that worked in quadrature. The phase meter outputs were sampled at 250 kHz using a 16 bit A/D converter and the phase was unwrapped using custom software. Response amplitudes were corrected for the frequency response of the recording system. The noise floor was <5 pm/$\sqrt{}$Hz, which allows measurements down to 0.01 nm for frequencies >6 kHz.

D. DPOAE measurements

DPOAEs recordings were made with an Etymotic ER10C microphone sealed in the ear canal. An additional 20 dB gain was used to amplify the stimuli and the microphone sensitivity was set at 40 dB. Stimuli consisted of two to four repetitions, 50 ms in duration repeated every 105 ms. The microphone output was amplified 40 dB and the signal was sampled at 250 kHz sample rate. BM measurements were made simultaneously using the interferometer.

Noise levels were typically below 0 dB SPL. Signal-to-noise ratios were typically 30 dB across all frequencies. An exception is for low level primary tones with the distortion product less than 2 kHz where the low-frequency noise level increased with decreasing frequency (1/f) to as much as 10 dB SPL [example in Fig. 13(D), L1=L2: 25 to 45 dB SPL].

To ensure that the invasive opening of the cochlea did not significantly alter DPOAE measurements, DPOAEs were measured in three animals before the bulla was opened, after the bulla was opened, and after the cochlea was opened. The acoustic system was calibrated after each manipulation. Most amplitude differences between the resulting DPOAE frequency functions are less than a few decibels with the major change due to opening the bulla, which reduces the impedance that the acoustic source sees. In one instance, the magnitude of the DPOAE after the cochlea was opened was larger than that measured prior to opening the bulla by a couple of decibels. It is likely that preparing the cochlea for mechanical measurements does not significantly impact the DPOAE results as long as the cochlea remains mechanically intact (sensitive and compressive).

E. Analysis

A software phase-locked analysis of the displacement wave form was performed after application of a Hamming data window. An exact number of periods of the stimulus frequency were included in the analyzed segment. The analyzed segment excluded the rise/fall segment of the data. All analysis routines were MATLAB based and were used for all BM and DPOAE data.

III. RESULTS

A. Two-tone distortion

The two-tone distortion paradigm involves presenting tones at two primary frequencies, designated f1 and f2 with their levels designated L1 and L2. In the study of basilar membrane vibration, f2 is typically set to the CF of the cochlear location studied, while f1 is varied from around 0.5 octave below to 0.5 octave above CF. When two tones are present in a nonlinear system, the result is the production of a number of distortion products (DPs). Because DPs at 2f1–f2 and 2f2–f1 usually have the largest amplitudes, they are the most frequently studied, and are often referred to as the lower and upper cubic difference tones, respectively.

In response to the two-tone paradigm, BM-IO curves in the first row of Fig. 1 illustrate the relation between the response to primary tones, f1 (dot-dashed line) and f2 (dotted line), and resulting distortion products at 2f1–f2 (solid line) and 2f2–f1 (dashed line) as the intensity of f1 is increased while the intensity of f2 is fixed. Intensity of f2 (L2) is specified above each column while the intensity of f1 (L1) is given on the abscissa. The f2/f1 ratio was 1.21, which has been shown to often result in the largest DPOAE at the frequency 2f1–f2. As L1 is increased to $\sim$ L2, the response to f2 begins to decrease, i.e., it is suppressed (top row, Fig. 1; Rhode, 2007b). The response at 2f2–f1 increases in amplitude as L1 increases, reaches a maximum near where L1 equals L2, and decreases rapidly as L1 is increased further. The response at 2f1–f2 follows a similar pattern with one other prominent feature, a notch in the IO function near 60 dB SPL. The response at 2f1–f2 decreases rapidly as L1 becomes much greater than L2.

In the lower row, the DPOAE response functions at 2f1–f2 and 2f2–f1 have patterns similar to those of the BM

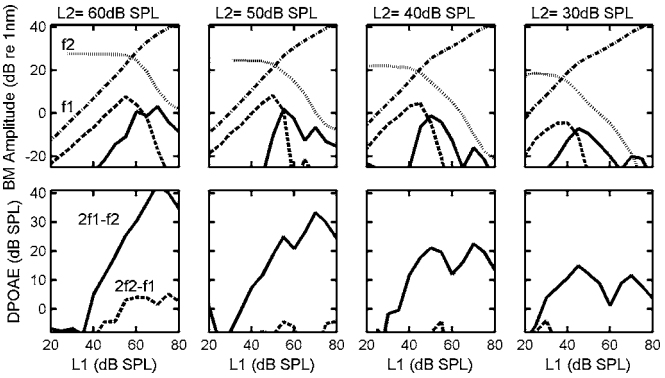


FIG. 1. Two-tone study with f2/f1=1.21 in which DPs for both BM and OAEs were recorded. Top row: Primary tones are indicated by the dotted line (f1=variable frequency primary tone) and the dashed-dotted line (f2=CF). The cubic distortion products, 2f1–f2 (solid line) and 2f2–f1 (dashed line) are shown as IO functions of the suppressor frequency with the intensity of the probe tone indicated above each column. A feature of interest is the occurrence of "notches" in the IO functions for both the BM displacement and the DPOAEs for the lower frequency cubic distortion tone in the bottom row. In the complete study, f1 ranged from 4.5 to 9 kHz with f2=6.55 kHz (16 frequencies). The intensity of f1 (L1) was varied from 80 to 20 dB SPL in −5 dB steps (13 intensities) as was L2 independently. Thus there were 16×13×13=2704 stimulus conditions. Chinchilla N92.

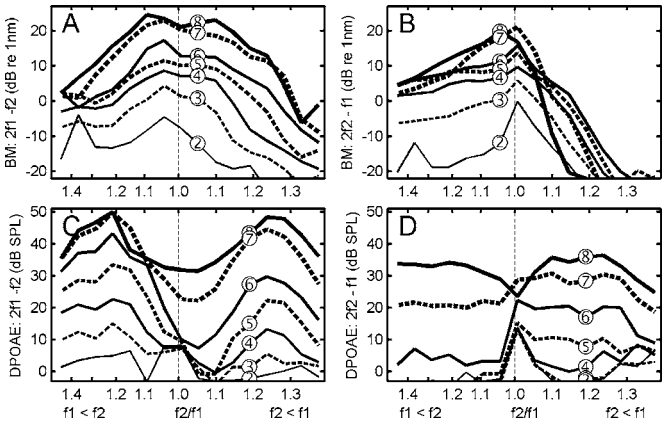


FIG. 2. The maximum values found for BM and DPOAE response when holding L1 constant and varying L2 from 20 to 80 dB SPL in 10 dB steps are presented. The L1 intensities are indicated by the symbol attached to each function (e.g., ③=3×10=30 dB SPL=L1). (A) Maximum BM displacement at 2f2–f1 vs the frequency ratio at several L1 intensities for any L2 intensity. In each instance the probe tone, f2=CF (6.35 kHz), and the second tone, f1 is swept from below f2 to above it. Note that when the f1 is >f2, the lower cubic distortion product becomes 2f2–f1. Therefore, the f2/f1 ratio is reversed, i.e., f1/f2 is used when f1 >f2 in order to keep the lower cubic tone the focus of the plot in (A). As f1 increases 2f2–f1 decreases. This is in effect two paradigms, one for sweeping f1<f2=CF and one when sweeping f1>f2=CF, that are combined. The roles of f1 and f2 are reversed, in effect f1 is fixed and f2 is swept. Intensity range was 20–80 dB for both L1 and L2 in 10 dB steps with the odd numbered intensities indicated by the dashed lines. Same assignment of line types and symbols for L2. (B) A plot of 2f1–f2 with the same rules as in (A). (C), (D) The maximum dB SPL of the distortion product otoacoustic emissions (DPOAEs=2f1−f2 and 2f2–f1 in (C) and (D), respectively, for each intensity of the probe. The recording time was 75 min and the data set is the same as used in Fig. 1. Chinchilla N92.

curves. They are nonmonotonic with a maximum around 70 dB SPL. At all but the highest intensity for L2, the DPOAE at 2f1–f2 has a notch near the same level as in the respective BM-IO curve, 60 dB SPL, but not exactly the same. This could result from the DPOAE response arising from a region of the cochlea in contrast to the BM response observed at a point in the cochlea. The DPOAE at 2f2–f1 has much smaller levels than at 2f1–f2, less than 10 dB SPL throughout, though when L2=60 dB SPL, there is a notch at L1=70 dB SPL.

The region of parameters in which these notches occur was typically limited to 1.1<f2/f1<1.4, L2<60 dB SPL and L1>L2. With f1>f2, notches in DPOAE-IO functions were sometimes seen when L2 was between 60 and 70 dB SPL. The largest notch depth was ~15 dB, though typically it was a couple of decibels. It is likely that the exact parameter combination for total cancellation of the effect of two distortion sources is rarely met with the stimulus space sampling parameters used in these experiments. It is also likely that this phenomena is only observed in sensitive cochlear preparations.

In a two-tone paradigm when f1 varied over a range of frequencies and f2 is fixed, the magnitude of DPOAEs varies with the f2/f1 ratio. In Fig. 2 both BM-DPs and DPOAEs are shown as a function of the ratio, f2/f1. This particular abscissa is used because the responses for 2f1–f2 and 2f2–f1 must tend to equality as f2/f1 tends to one and when f1 is >f2, the lower cubic distortion product is 2f2–f1. Think of this paradigm as the juxtaposition of two, two-tone paradigms: (1) when f1<f2=CF sweep f1 and (2) when f1>f2 sweep f2 with f1=CF. In case (2), there is a role reversal for the primaries as f2=CF becomes the lower primary as f1 sweeps past it and becomes the upper primary. BM maximum displacement at each value of L2 (20–80 dB SPL) is relatively symmetric about f2/f1=1 and drops off as the ratio is increased [Fig. 2(A)]. For each L2 value the individual curves all follow a similar pattern as a function of f2/f1. In contrast to the nearly symmetric pattern of maxima for 2f1–f2 [Fig. 2(A)], the response at 2f2–f1 [Fig. 2(B)] has an asymmetric pattern.

DPOAEs follow a different pattern as a function of f2/f1, attaining maxima when f2/f1 is ~1.2. The DPOAE at 2f1–f2 is nearly symmetric around f2/f1=1 [Fig. 2(C)] while the 2f2–f1 response pattern is asymmetric [Fig. 2(D)]. The 2f2–f1 component response appears to be nearly invariant as a function of f2/f1 when L2 is greater than 60 dB SPL. The DPOAE at 2f1–f2 is larger than at 2f2–f1 for all conditions with a maximum value when f2/f1=1.2. Maximum DPOAE levels at 2f1–f2 are between 20 and 30 dB lower than the primary tone levels while those at 2f2–f1 are around 10 dB less at 2f1–f2.

A second example of corresponding BM-DP and DPOAE data as a function of the f2/f1 ratio is shown in Fig. 3 when f1 was swept from much lower frequencies relative to CF than in Fig. 2. While this is not the typical case one would encounter in collecting DPOAEs, it is nevertheless instructive. BM displacement at 2f1–f2 increases with L2 in a systematic manner for f2/f1<1.4 [Fig. 3(A)]. For larger ratios it rapidly attains a maximum by 50 dB SPL and remains relatively invariant as a function of f2/f1 (1.4–1.9). It appears that this may be a result of f1 residing in the "tail" region of the response to f2 and where the BM-f1 response is linear. The BM response at 2f2–f1 [Fig. 3(B)] again has an asymmetric pattern as a function of f2/f1 and attains a fairly

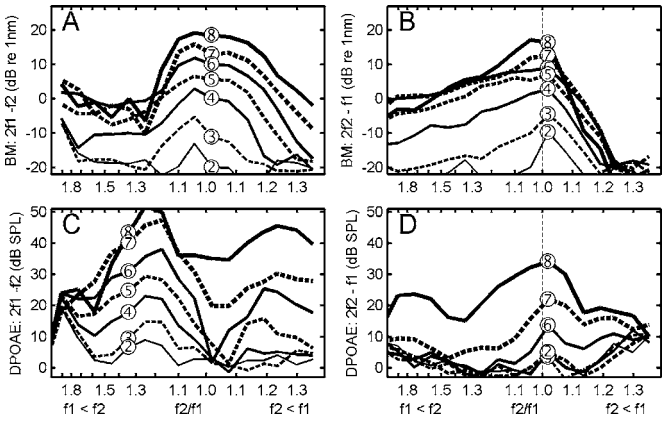


FIG. 3. Example of the maximum response amplitudes for BM and OAE DPs when f1 was swept from lower frequencies than in Fig. 2. (A) Maximum BM displacement at 2f1–f2 vs the frequency ratio at several L2 intensities indicated by the symbol attached to each function (as described in the caption of Fig. 2). (B) Maximum of 2f2–f1 as a function of L1 and the f2/f1 ratio. (C) DPOAE at 2f1–f2. (D) DPOAE of 2f2–f1. f2=CF=7.25 kHz. Chinchilla N96.

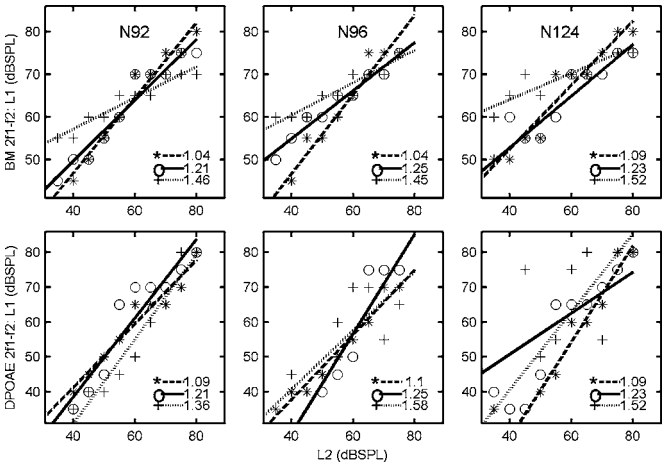


FIG. 5. Three studies of the L1 and L2 intensities that produce the maximum 2f1–f2 BM response amplitude (top row) and the maximum 2f1–f2 DPOAE displacement (bottom row) at the 2f1–f2 DP frequency for three f2/f1 ratios. The legend in each panel indicates the symbol and line style for each ratio. The lines correspond to those obtained by using the "polyfit" function in MATLAB to obtain the linear regression line fit to the data points. Chinchilla identification is given in each panel of the top row.

constant maximum when L2 > 50 dB SPL. Note that both of these BM responses grow at near linear rates for levels less than 40 dB SPL and at compressive rates above that level.

DPOAEs at 2f1–f2 have a minimum for f2/f1=1 [Fig. 3(C)]. When f1 is <f2, it increases in an orderly manner with increasing L2 with the maximum values (f2/f1=1.2) being 25–30 dB below the intensity of the primaries. DPOAEs at 2f2–f1 are considerably lower (–50 dB relative to the primaries) and are nearly undetectable when the levels are <60 dB SPL [Fig. 3(D)].

A better example of the bimodal nature of the 2f1–f2 DPOAE when f1 is swept from below to above f2 than appears in Figs. 2(C) and 3(C) is shown in Fig. 4. This contrasts strongly with the 2f1–f2 BM amplitude that peaks when f1 is near f2 (Figs. 2 and 3).

Human DPOAE studies have shown that the maximum 2f1–f2 OAE is a function of the two primary intensities and the f2/f1 ratio. The variables have been incorporated into what has been named the "scissors relation" (Kummer *et al.*, 2000; Whitehead *et al.*, 1995). A similar relation is obtained for both BM-DPs and DPOAEs in chinchillas as illustrated in Fig. 5. The raw data are somewhat noisy due to the short recording time (80 ms) and therefore the actual data is scattered around the first-order regression line fit to the data. In the upper row of Fig. 5 the BM-IO displacement functions (upper row, Fig. 5) have a pattern that is similar to the DPOAE-IO functions (lower row, Fig. 5). For small ratios where the two primaries share the same amount of compression (e.g., f2/f1=1.04), the maximum is obtained with L1 ≅L2 across the entire range of intensities. For larger f2/f1 ratios and low L2s, L1 has to be larger than L2 in order for the response to f1 to attain amplitude equal to that of f2 at the f2 location in the cochlea. Around 60 dB SPL there is a convergence of the responses where L1=L2 yields the maximum BM amplitude at 2f1–f2. As L2 is increased further, >60 dB SPL, L1 has to be less than L2 to produce the maximum displacement at 2f1–f2. This follows as a result of the greater growth of displacement at f1 with increasing intensity, i.e., less compression than at f2, which is in a more compressive region. In general, there is no one optimum relation that applies to all animals, ratios, and frequencies as has been recently found in humans (Neely *et al.*, 2005; Johnson *et al.*, 2006).

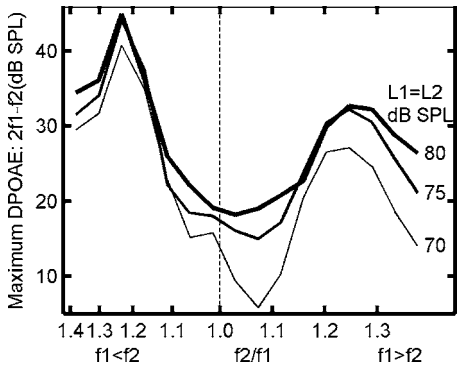


FIG. 4. Maximum of DPOAE response at 2f1–f2 for three levels of L1 that emphasizes the bimodal nature of the maximum as a function of the f2/f1 ratio as f1 is swept from a frequency below f2 to one above it. Chinchilla N92: where the simultaneous BM responses were nearly symmetric around f2/f1=1 (not shown) and were made at a different BM location than the BM location used for the measurements shown in Figs. 1 and 2. Six separate locations were studied in N92 with CFs between 6.2 and 6.75 kHz.

B. Growth of DPs with intensity

The growth of both BM-DPs and DPOAEs as a function of the level of the primary tones and the f2/f1 ratio is illustrated with contour functions in Fig. 6. BM-DPs (upper row) are nearly independent of L2 when f2/f1 is >1.25 and L1 <60 dB SPL, that is, the contour lines are nearly vertical. DPOAEs (lower row) have a similar pattern in this parameter region. In all cases the BM-DP and DPOAE responses increase with increasing L1 in a systematic manner. While the DPOAE attains its maximum response, 50 dB SPL when L1=L2=75 dB SPL for f2/f1=1.21, the BM-DP response

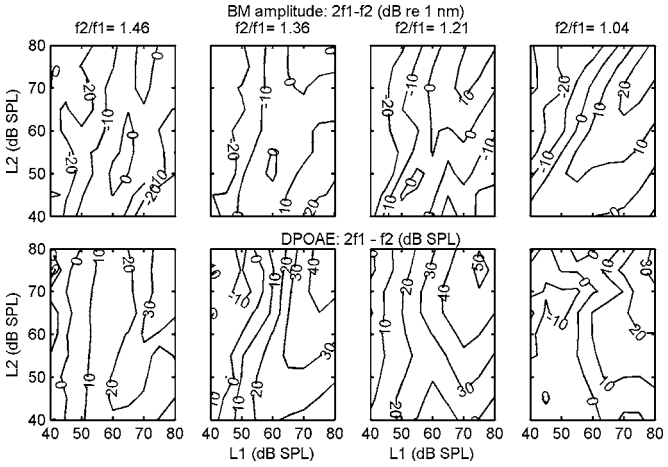


FIG. 6. Contour plots of responses at the 2f1–f2 frequency for both BM (row 1) and DPOAE (row 2) as a function of L1 and L2, which were varied in 5 dB increment between 40 and 80 dB SPL for the four f2/f1 ratios listed above each column. The BM amplitude vary between −20 and 20 dB re 1 nm (the minus signs are not always apparent though in general amplitude increases with increasing L1). Chinchilla N92.

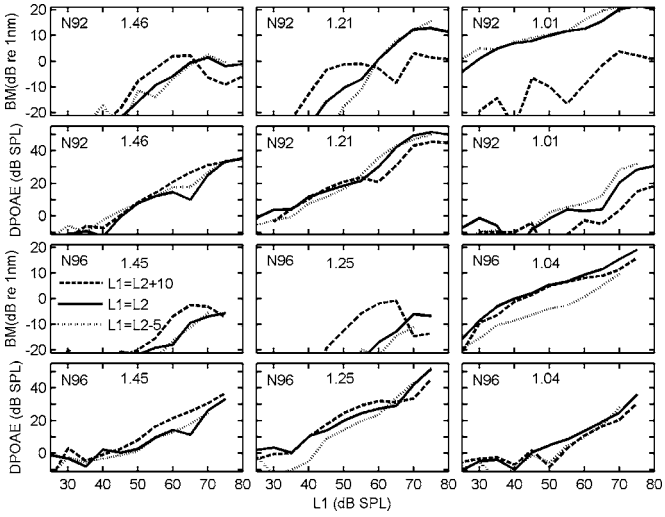


FIG. 7. Two studies of BM amplitude and DPOAE growth functions at the f2/f1 ratios indicated in each panel. Chinchilla N92 responses in the top two rows and chinchilla N96 responses in the bottom two rows. The line type code is: L1=L2+10 dB SPL (dashed line), L1=L2 (solid line), and L1=L2−5 (dotted line) as shown in the lower left panel. It is important to note that though L1 is listed as the abscissa variable, L2 is fixed relative to L1 as described and as such also varies along the abscissa.

amplitude is maximum for f2/f1=1.04 and its growth is clearly different from that of the DPOAE. There is rate-of-growth information in the contour plots as the closeness of the contours reflects the slope of IO functions as a function of L1, L2, or both. With increasing L1, the slope of the DPOAE-IO function when f2/f1=1.36 is >3 dB/dB, whereas the BM-DP IO slope is highest when f2/f1=1.04. Typically the growth rate is less than 1 dB/dB and can be negative for high levels of L1 relative to L2. The slope of the IO functions changes continuously with both stimulus level and the frequency ratio. For f2/f1=1.21 the slope is near 1.0 dB/dB up to L1=60 dB SPL, which agrees with results in humans(Dorn *et al.*, 2001; Neely *et al.*, 2003, 2005; Boege and Janssen, 2002; Gorga *et al.*, 2003).

When DPOAE-IO functions are measured, there is often a constant relation between L1 and L2. However, as indicated earlier, L1 has to be greater than L2 at low levels in order to obtain the maximum response. BM-DP and DPOAE IO responses for three L1:L2 relations and three f2/f1 ratios in two animals are illustrated in Fig. 7. The largest responses are obtained when L1=L2+10 as long as L1<60 dB SPL (N92). The latter is approximately true for N96 (lower two rows of Fig. 7) though the DPOAE remains larger for L1=L2+10 up to 75 dB SPL, the highest level used in that study. There is a deviation of the DPOAE growth with increasing levels from that of the BM-DP, in that the DPOAE continues to increase while the BM-DP saturates (f2/f1<1.2). There are instances where the DPOAE-IO slope increases dramatically to as much as 2 dB/dB at levels >70 dB SPL.

The DPOAE-IO functions have several features that deviate from a simple linear relation. There is an increase in their slopes and frequently a dip when L1 is between 60 and 70 dB SPL. This is the same L1–intensity range where notches occurred in the IO curves shown in Fig. 1. The growth rate is ~1 dB/dB for most levels below 60 dB SPL though for high f2/f1 ratios the rate decreases (N96, f2/f1=1.45). Above ~60 dB SPL the growth rate increases to 2 dB/dB, though not in all instances. While the DPOAE-IO function saturates in some instances beyond 70 dB SPL (Fig. 7, N92, f2/f1=1.21), in the bottom row there are instances of no saturation at levels up to 75 dB SPL. In another instance the DPOAE increased even at 80 dB SPL where it was 60 dB SPL and only ~20 dB less than the primary levels.

A common method of measuring DPOAEs is to sweep a frequency range while holding the f2/f1 ratio constant. A comparison of BM and OAE-DPs at several stimulus intensities with L1=L2 (intensity indicated by the symbols, e.g., ⑤=50 dB SPL) and f2/f1=1.22 is shown in Fig. 8 as a function of 2f1–f2, the frequency of the distortion product of

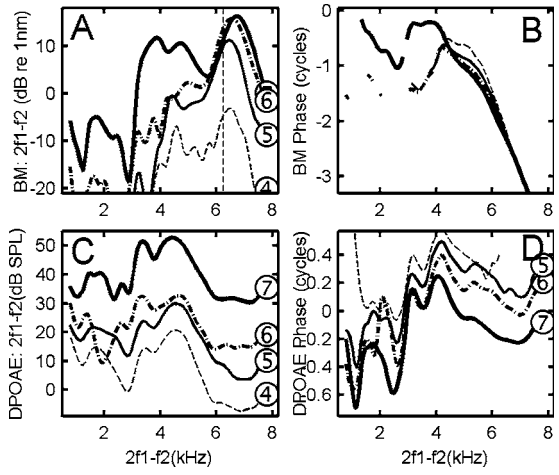


FIG. 8. Example of frequency transfer functions of BM vibration and DPOAEs for the 2f1–f2 distortion component using simultaneous recording of both responses with f2/f1=1.22. (A), (B) BM amplitude and phase functions for L1=40 (④), 50 (⑤), 60 (⑥), and 70 (⑦) dB SPL as indicated with L1=L2+5 dB. Functions were three-point filtered using the MATLAB function "filtfilt." (C), (D) Amplitude and phase, respectively, for the DP emission at 2f1−f2 kHz. N92.

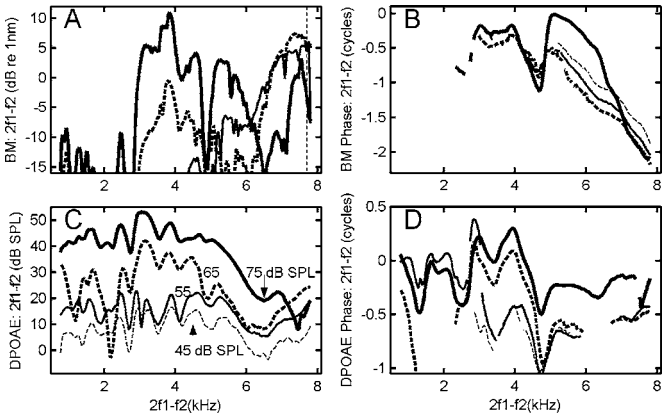


FIG. 9. Example of BM amplitude and phase responses at 2f1−f2 kHz at four intensity levels: 45–75 dB SPL in steps of 10 dB with L1=L2 and CF=7.7 kHz. Increasing intensity is indicated by increasing line thickness. (A), (B) BM frequency transfer curves for amplitude and phase, respectively. (C), (D) DPOAE (2f1–f2) frequency transfer curves for amplitude and phase respectively. N101.

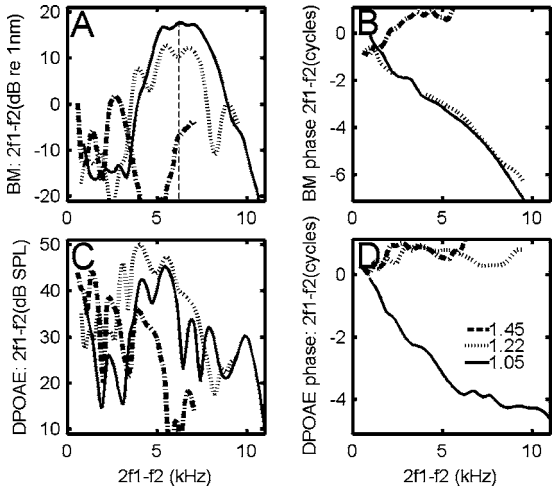


FIG. 10. (A), (B) BM frequency amplitude and phase responses at the 2f1–f2 cubic difference tone frequency for three f2/f1 ratios indicated in (D) along with the line style. (C), (D) DPOAE amplitude and phase response. L1=L2=70 dB SPL with CF=6.25 kHz. F1 was varied from 1 to 12 kHz in 100 Hz increments. N92.

greatest interest. There are a number of similarities between BM-DP and DPOAE responses, perhaps surprising given the BM is measured at a single cochlear location and OAEs arise from a region of the cochlea that varies from apex to base as f1 and f2 are swept. At 70 dB SPL, both amplitude responses [Figs. 8(A) and 8(C)] resemble each other up to ~6 kHz, which is the CF of the BM region under study. As expected in this frequency region, the BM distortion product increases and reflects cochlear tuning, that is, the filter function of the BM whereas the DPOAE is relatively flat in this region. Both BM and DPOAE functions have ripples that are similar but not identical. Their growth with level is not the same, especially in the CF region of the BM, where the BM response is nearly constant between 60 and 70 dB SPL. Both BM and OAE phase relations exhibit an increasing phase lag with increasing intensity [Figs. 8(B) and 8(D)]. The ripple below 4 kHz in the two functions is weakly similar while the BM phase has a large phase lag with increasing 2f1–f2 while the DPOAE phase varies relatively little above 5 kHz.

A second comparison of BM and OAE DPs shows there is some similarity between them in the 3–6 kHz region [Figs. 9(A) and 9(C)]. However, there is a null in the BM-DP response at ~4.5 kHz that is not present in the DPOAE response at 75 dB SPL (indicated by the thickest line). The BM-DP amplitude response (75 dB SPL) has a small amplitude, high frequency ripple superimposed on it that is absent in the DPOAE response and that was seen on occasion in other studies. The increase in BM-DP amplitude reflects the BM filter function as 2f1–f2 approaches the CF region (7.7 kHz), which is not seen in the DPOAE response. There is a decrease in BM amplitude near CF as the level is increased from 65 to 75 dB SPL that also occurs in the DPOAE. At lower levels, the DPOAE response has a ripple pattern that varies with intensity suggesting that there are multiple sources of OAEs interacting in a complex fashion. The BM response is too small at the lower levels to be able to make a valid comparison with the DPOAE response. Both phase responses show a phase lead as the level is increased from 65 to 75 dB SPL [Figs. 9(B) and 9(D)]. Both have nearly a cycle decrease as 2f1–f2 is increased from 4 to 4.5 kHz. As seen in Fig. 8(B), the BM-DP phase lags up to several cycles as 2f1–f2 approaches the CF, while DPOAE phase [Fig. 8(D)] is relatively unchanged in this region.

C. Variable f2/f1 ratios

While f2/f1=1.22 is the dominant ratio employed in OAE studies because it typically results in the largest emissions, other ratios may provide additional information about the OAE generation process. The BM-DP response at the smallest ratio of three (1.05), resulted in the largest BM distortion at 2f1–f2 over nearly the entire frequency range [L1=L2=70 dB SPL, Fig. 10(A)]. Nearly the same amplitude is obtained for a ratio of 1.22, while there is a much smaller amplitude for a ratio of 1.45. The latter reduction is partially a result of the "scissors" relation, i.e., when L2 is greater than 60 dB SPL and the f2/f1 ratio is 1.45, L1 would have to be reduced in order to attain the maximum DP amplitude because the response to f2 is more compressed than that to f1 (see Fig. 5). The phase curves for small f2/f1 ratios behave similarly, both have increasing phase lags while the f2/f1=1.45-phase function exhibits a phase lead with increasing DP frequency [Fig. 10(B)].

DPOAEs follow a similar amplitude relation as a function of f2/f1, i.e., the smallest amplitudes occur for the highest ratio except for 2f1–f2 <3 kHz [Fig. 10(C)]. The largest number of ripples in the curves occurs for the smallest ratio, 1.05, where the spacing of the ripples between 4 and 8 kHz is on the order of 1 kHz, much greater than the ~100 Hz spacing observed in humans. The maximum DPOAE occurs at 4 kHz for the 1.22 ratio with nearly 2 kHz spacing in the peaks between 4 and 6 kHz. The phase curves indicate that there is relatively little phase change as a function of the DP frequency for the two higher ratios implying that the DPOAE source location is at the wave location while a large phase

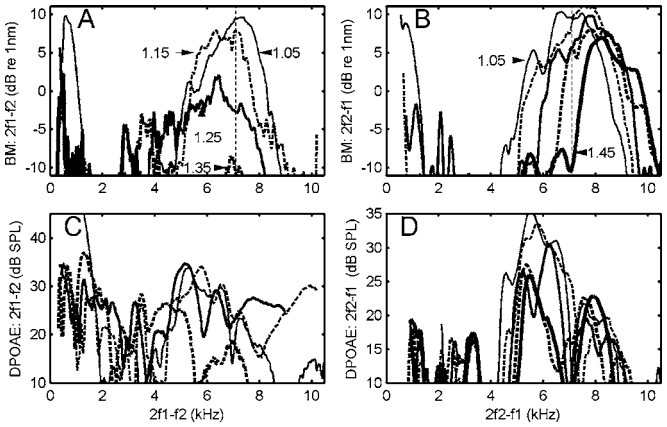


FIG. 11. An example of the lower (2f1–f2) and upper (2f2–f1) cubic difference tones for BM and DPOAEs at f2/f1 ratios that vary from 1.05 to 1.45 in 0.05 steps with L2=L1=60 dB SPL. Only every other instance of the responses for the f2/f1 ratios is shown due to the complexity of the functions. (A), (B) BM frequency amplitude responses for the DPs at 2f1–f2 and 2f2–f1 kHz, respectively. (C), (D) DPOAE frequency responses that correspond to the BM responses shown immediately above [i.e., (A), (B)]. Note the change in the maximum ordinate value between (C), (D) CF=6.7 kHz. N133.

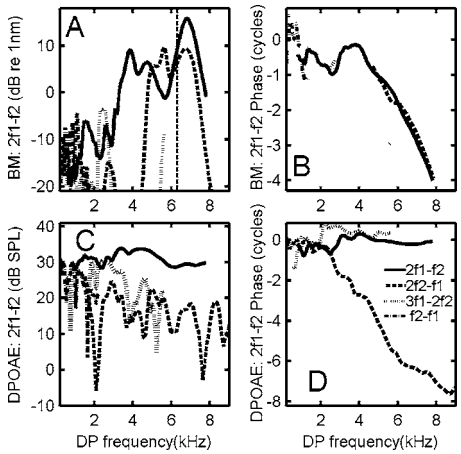


FIG. 12. Example of BM and DPOAE amplitude and phase frequency responses for four principal distortion product components as a function of their own frequency. Both BM [(A), (B)] and DPOAEs [(C), (D)] are shown as indicated. Line style assignments shown in the legend of (D) apply to all panels. In (B) the phase values are suppressed when component amplitude was <−15 dB re 1 nm. f1 was swept from 1 to 10 kHz in steps of 50 Hz. L1=L2+10=70 dB SPL with f2/f1=1.22. N92.

shift is seen for the smallest frequency ratio implying the source location is at the place location [Figs. 10(B) and 10(D)].

A second example of two-tone DP responses for differing f2/f1 ratios, where there was a smaller step of 0.05 in the ratios, has a somewhat different relation of the DPs to the DP frequency (Fig. 11). BM-DP response at 2f1–f2 has the maximum amplitude for the smaller f2/f1 ratios and is largely restricted to the frequency region around CF except for the frequency region less than 1 kHz [Fig. 11(A)]. The corresponding DPOAEs are such a complex function of both the ratio and DP frequency that only every other curve is shown [Fig. 11(C)]. There is a maximum at 1 kHz for high ratios and a second maximum around 5 kHz for lower ratios.

The BM-DP response at 2f2–f1 [Fig. 11(B)] follows a similar pattern as the lower cubic tone with respect to DP frequency. The maximum occurs above the CF frequency and shifts systematically to a higher frequency region as the f2/f1 ratio increases. This is a result of the primaries being below 2f2–f1, within the nonlinear region for this CF (CF =6.7 kHz). Between 5 and 6 kHz the maximum amplitudes are similar to those obtained by the response to 2f1–f2. The overall pattern appears to be largely determined by the compressive region or tip portion of the tuning characteristic of the BM at this location. The corresponding DPOAE at 2f2–f1 is also largely confined to the region around CF [Fig. 11(D)]. Largest amplitudes occur for the lower ratios in the vicinity of 6 kHz. At the larger ratios there is a double maximum, one at 5.5 and the other at 8 kHz. There are also smaller peaks between 1 and 3 kHz with a prominent null at 4 kHz.

D. Multiple DPs

While the DP at 2f1–f2 is the most frequently studied and used clinically, DPs at 3f1–f2, f2–f1, and 2f2–f1 are also of interest. Each has its own pattern as a function of its DP frequency. The BM-DP responses for 2f1–f2 (solid line) and 2f2–f1 (dashed line) are similar to those already shown [Fig. 12(A)]. Maximum amplitudes occur around CF. Maximum for 3f1–2f2 occurs at 2.5 kHz at which frequency the two primaries are in the nonlinear frequency region for this location. The f2–f1 distortion component is relatively small, <0.5 nm, and restricted to low frequencies. The corresponding phase functions appear to be similar for each component with little phase change below 4 kHz and increasing phase lag above it [Fig. 12(B)].

The 2f1–f2 DPOAE is the largest component with 3f1–2f2 a close second [Fig. 12(C)]. 2f2–f1 has many peaks and minima over the entire frequency range. The OAE at the f2–f1 difference frequency is relatively small [Fig. 12(C)]. These functions are considerably different than the BM responses that are more determined by the response at the f2 cochlear location to the individual components. The corresponding phase functions for 2f1–f2 and 3f1–2f2 are nearly constant as a function of their frequency [Fig. 12(D)]. In contrast, 2f2–f1 phase has a large increasing phase lag as a function its frequency.

E. DPOAEs postmortem

The origin of DPOAEs in the cochlea is of considerable interest. One manipulation that sheds some light on this matter is how DPOAEs behave as a function of time postmortem as measured from the time from respiratory arrest due to an anesthetic overdose. Changes in BM amplitude as a function of time postmortem indicate that within a few minutes the BM-DP response is nearly completely gone, i.e., <0.1 nm, reduced by over 30 dB at 6.5 kHz [Figs. 13(A) and 13(C)] except at the highest level of stimulation, 75 dB SPL. In contrast, the postmortem DPOAE is reduced to the noise level for intensities below 45 dB SPL while for the highest intensity (75 dB SPL) there is relatively little reduction across all frequencies [Figs. 13(B) and 13(D)]. Another change in the DPOAEs is that ripples in the frequency curves

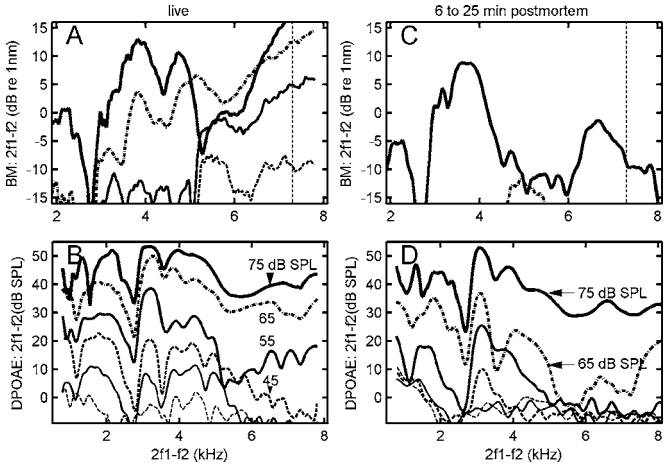


FIG. 13. Example of changes in BM vibration and OAE amplitudes upon death for a location with CF=7.3 kHz. (A) BM amplitude at 2f1−f2 kHz for intensities from 25 to 75 dB SPL while the animal is alive. Linewidth decreases with decreasing intensity and solid and dashed lines are alternated. The dashed vertical line marks CF. (B) DPOAEs at 2f1–f2 corresponding to the BM data in (A). (C) BM amplitude response at 2f1−f2 kHz after death. Measurements were made between 6 and 25 min after death. (D) DPOAE response at 2f1−f2 kHz for the same conditions as in (C). N180: cochlear gain >44 dB.

are nearly absent at intensities greater than 60 dB SPL, i.e., fine structure is more evident at low levels (e.g., Konrad-Martin *et al.*, 2001). If the rippling is a result of two source interaction then it would appear that one source is reduced faster than the other, the reflection or place source.

A second study of the effect of death on DP amplitudes, in a preparation that deteriorated over the course of recording, shows that 0.1 h after death, that DPOAEs with L1 =L2=70 dB SPL are little changed. Even after 0.5 h there is relatively little change in their magnitude [Fig. 14(A)]. At

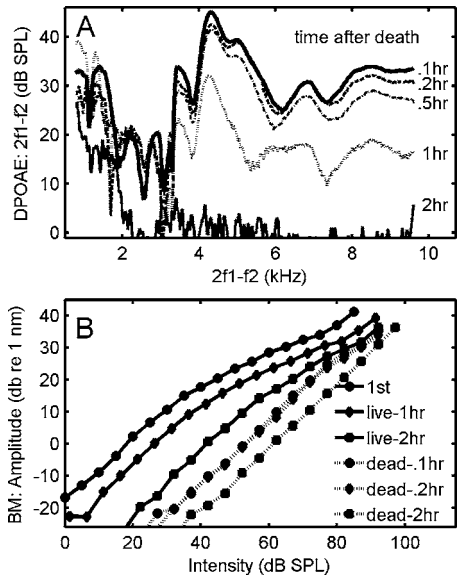


FIG. 14. (A) DPOAE (2f1–f2) at five instances in time after death. (B) The BM IO functions at three times before death and three times after death indicating that the BM becomes linear over the entire range rapidly while the DPOAE require 2 h to completely disappear. L1=L2=70 dB SPL and CF=6.7 kHz. N133: cochlear gain while alive was >35 dB.

1 h postmortem there is approximately a 10 dB reduction over the 4–9 kHz range and at 2 h after death the DPOAE has nearly disappeared.

In contrast, there was a gradual reduction in the BM-DP amplitude at CF while the animal was alive and that by 0.1 h after death an additional 10 dB of amplitude was lost [Fig. 14(B)]. Other examples, not shown, indicate that the loss of BM amplitude can occur as fast as 2 min after death with as much as a 40 dB loss. It appears that the effect of shutting down the active process attributed to the OHCs results in a nearly immediate reduction in BM amplitude to those of passive levels of vibration of the BM while there is a relatively small change of DPOAE amplitude over the same time frame. A possible correlate of this timeline has been observed in the loss of compression in the vibration of the BM and in the reduction of the endocochlear potential (Sewell, 1984). There remains a small amount of compression (0.9 dB/dB) for around 2 h for levels above 70 dB SPL (Rhode, 2007a). This remaining nonlinearity may be sufficient to explain the high level DPOAE coming from a passive cochlea though there is evidence that the high level response may be from a passive reflection of energy from cochlear mechanical irregularities (Kemp, 2002).

IV. DISCUSSION

As stated by Kemp (2002) in a review of otoacoustic emissions, "the mode of emergence of OAEs from the cochlea is not only complex, but to some extent it is an arbitrary process. … the mechanism of escape of this energy is still poorly understood." Explanations of the mechanisms by which OAEs reach the ear canal include reverse traveling waves and coherent reflection off of randomly distributed anatomical imperfections (Kemp, 1986; Shera and Guinan, 1999). The hypothesis of reverse traveling waves has been recently challenged. Ren (2004) was unable to find any BM motion indicating a reverse traveling except that of a forward traveling wave in measuring 2f1–f2 in the basal region of gerbil. The same conclusion was reached in chinchillas by Siegel *et al.* (2005). The need to explain the exact nature of OAE generation underlies the need for further studies of the relation between BM behavior and OAEs.

Findings of this study:

(1) Notches in both BM-DP and DPOAE IO functions occur at approximately, though not exactly, the same intensities and f2/f1 ratios.
(2) The f2/f1 ratio of around 1.2 appears to be optimum for obtaining greatest sensitivity and maximum amplitude of DPOAEs. In contrast, an f2/f1 ratio near 1.0 results in the maximum BM-DP. In other words, for f2/f1 near unity there is a relative minimum in the 2f1–f2 DPOAE amplitude as measured in the ear canal but a maximum in BM amplitude.
(3) The BM-DP frequency function ripple patterns exhibit a weak correlation with the corresponding DPOAE frequency response functions. This is likely a result reflecting the nature of a point measurement (BM) versus a variable location and spatial extent being the site of DPOAE generation. The DPOAE (2f1–f2) ripple fre-

quency in constant f2/f1 ratio swept primary frequency functions is much lower in chinchillas than humans.

(4) With both primary levels increasing at the same rate, the slope of the DPOAE-IO functions is often on the order of 1 dB/dB at least up to 60 dB SPL though the slope is a function of the f2/f1 ratio, the levels of the primaries, and varies between animals. For primary levels greater than 60–70 dB SPL, the growth rate can increase to as much as 2 dB/dB suggesting some additional process is invoked. For f2/f1 ratios >1.2, BM-DP IO functions saturate with increasing levels while DPOAE IO functions continue to increase (Fig. 7). However, there is considerable similarity between the rate-of-growth patterns between the two functions.

(5) For primary levels >70 dB SPL, DPOAEs remain for up to 2 h after death while BM-DPs are somewhat reduced (Fig. 13). At lower levels of the primaries, BM-DPs are nearly absent while DPOAEs are reduced 10–15 dB within minutes after death. DPOAE amplitudes are more robust for 2f1–f2 >5 kHz than for higher frequencies.

(6) Maximum DPOAE (2f1–f2) levels are ~−20 dB relative to the primary levels and remain nearly so independent of the f2/f1 ratio. In contrast maximum BM-DP amplitudes can be as little as −10 dB relative to the amplitudes of the primaries for small f2/f1 ratios (f2/f1<1.05) and decrease systematically with increasing f2/f1 to <−30 dB relative to the primaries when f2/f1>1.35.

A. DPOAE IO functions

Numerous attempts to correlate DPOAE-IO functions with auditory thresholds have resulted in partial agreement, often being limited to a portion of the frequency range and/or condition of the cochlea (e.g., Boege and Janssen, 2002; Janssen *et al.*, 1998; Avan *et al.*, 2003; Gorga *et al.*, 2003). BM displacement correlates well with AN fiber behavior and as such is expected to correlate with DPOAEs inasmuch as both are related to OHCs. It is generally thought that the BM-IO function is linear at sufficiently low levels, becomes compressive between 20 and 40 dB SPL, and returns to linearity at some high level, 90–105 dB SPL. This is not always the case as it has been shown that in highly sensitive cochleas the BM-IO at CF is mildly compressive even at 0 dB SPL (AN threshold) with increased compression above 20–30 dB SPL and remains compressive at least up to 105 dB SPL in very sensitive cochleae (Rhode, 2007a). However, either due to age, noise exposure, or insult resulting from preparing the cochlea, nearly all cochleae studied have been found to be linear below 20 dB SPL. If the cochlea were completely linear and uniform then there would be no distortion products generated. Because DPOAEs are recorded at low levels in humans there must be nonlinearities present in order for DPs to be generated, possibly in the tip links of the stereocilia (Jaramillo, 1993). Production of DPOAEs at the lowest levels would likely occur only at the location of the upper primary, f2: the wave or nonlinear reflection location. If the BM response is linear at levels <20 dB SPL there would be no distortion components to propagate to their "place." Hence, no second component would arise from the place or coherent reflection location, e.g., the 2f1–f2 location. This argues for using low stimulation levels to assess the state of the cochlea at the f2 location, though this has the penalty of requiring longer recording times. Even though DPOAEs in chinchillas are 20–30 dB greater than in humans, correlating BM-DPs with DPOAEs at low levels remains problematic because stimulus levels greater than30 dB SPL are necessary to obtain DPOAEs >−10 dB SPL. Long recording times are not possible when BM vibration is recorded due to limited duration of viability of the cochlea after it has been opened. A single two-tone study can take more than 2 h when fine sampling of frequency and/or intensity is employed. Noteworthy is the use of low stimulus levels in humans is likely not to be of great utility for clinical applications as it has several negatives such as high false-positive rates and detecting mild cochlear dysfunction might not be of clinical significance.

DPOAE-IO functions have been studied under a variety of manipulations of the cochlea that affect the OHCs. There is a persistent DPOAE at high levels in cochlea impaired by loud noise exposure (e.g., Zurek *et al.*, 1982; Dorn *et al.*, 2001), ototoxicity (e.g., Brown *et al.*, 1989) or genetic factors (Avan *et al.*, 2003). The present results show that there is a significant DPOAE after death that lasts for up to 2 h. This suggests that there is some element of the OHCs that continues to exhibit a nonlinear characteristic, likely the transduction channels. In contrast, low level DPOAEs disappear rapidly after insult to the cochlear blood supply or delivery of furosemide (Avan *et al.*, 2003). This correlates with the rapid drop in the endocochlear potential (Sewell, 1984). As shown in Fig. 13, BM-DPs for levels <30 dB SPL are absent or are below the measurement system noise level within minutes after death.

In the intact cochlea, DPOAE amplitude often grows at ~1 dB/dB rate over a range of levels <60–70 dB SPL. This is a somewhat simplistic view as the growth rate varies with the f2/f1 ratio, stimulus levels and relative levels of L1 and L2, and between animals. For example, in Fig. 7, f2/f1=1.23, the slope is ~0.3 dB/dB up to 60 dB SPL. This is in rough agreement with the slope behavior in humans where near threshold the slope is ~1 dB/dB, then decreases to ~0.5 dB/dB up to stimulus levels of 70–80 dB SPL. Above 70–80 dB SPL, the slope increases to ~2 dB/dB (Fig. 4, Dorn *et al.*, 2001). At levels greater than 60–75 dB SPL the growth rate in chinchillas increases to ~2 dB/dB though not always, as occasionally the DPOAEs saturate around 80 dB SPL (Fig. 7). Perhaps if higher stimulus levels were used there would have been a more consistent behavior. This increase in slope at higher levels may be a result of the growth of BM displacement in the "tail region" being driven at sufficient amplitudes that engage the OHCs basally located to f1 and f2 locations. That is, an extended cochlear region is vibrating at amplitudes sufficient to open and close transduction channels of OHCs.

BM-DP-IO and DPOAE-IO functions shown in Fig. 1 are similar in shape when L2 is constant and L1 is varied with f2/f1=1.2. They are nonmonotonic and have notches in the L1=60 dB SPL region. The level (L1) at which the notch occurs increases, up to 10 dB, as the f2/f1 ratio increases at least over the stimulus space covered in these experiments,

i.e., L1 and L2 < 85 dB SPL. The depth of the notches does not reflect complete cancellation due to two sources interacting. This could be due to sampling intervals in both frequency and intensity being large relative to the sampling intervals (1 Hz and 1 dB steps result in significant changes) necessary to result in deep notches as shown in single-tone IO-function notches (Rhode, 2007a). In five two-tone studies the notch depths were only a couple of decibels and limited to a range of f2/f1 ratios and intensities. Notches in human SFOAEs occur around the 60 dB SPL level (Schairer *et al.*, 2003; He and Schmiedt, 1993) and rodents (e.g., Brown, 1987). Mills (1997) used a simple two-component model to explain the notches in rodent DPOAE-IOs where one component arises from the nonlinear region (wave) and the second from the CF location for 2f1–f2 (place); when they are 180° out-of-phase and equal amplitude there is cancellation. However, note that a single saturating nonlinearity representing mechanoelectric transduction of the hair bundle can also result in notches in distortion product IO functions (Lukashkin *et al.*, 2002).

B. Frequency sweep functions with constant f2/f1 ratio

The two-tone swept frequency paradigm with a constant f2/f1 ratio results in DPOAE functions with a ripple pattern that is interpreted to result from the interference between two OAE sources (e.g., wave and place, Knight and Kemp, 2001; or nonlinear and reflection, Shera and Guinan, 1999). There is also a ripple pattern in BM-DP frequency functions that correlates somewhat with the ripple pattern in DPOAE frequency functions. However, it is a weak correlation and ripples in BM-DP functions can be very small or even absent.

In humans, the ripple frequency is around 100 Hz when f1 is in the 1 kHz region, while in chinchilla it varies from a few hundred hertz to over 1 kHz. This difference does not appear to be explained by the difference in cochlear length, 36 vs 18 mm even taking into account the upper frequency limit for humans is a few kilohertz lower than chinchilla (Fay, 1988). Low f2/f1 ratios (1.05) have a higher ripple frequency than larger ratios which could result from the two sources being closer together. The ripple pattern can be similar for several DPs (2f1–f2, 2f2–f1, 3f1–2f2) when plotted as a function of the DP frequency. Occasionally, DPOAE frequency functions for ratios of 1.2, corresponding to the ratio where the maximum OAE at 2f1–f2 is obtained, exhibit relatively sparse ripples for 2f1–f2 with more ripples for 2f2–f1. The frequency spacing between successive ripples sometimes appeared to be evenly spaced on a log axis suggesting that two OAE generation sites obey a log-scaled model of the cochlea though there were also clear violations of this relation (Zweig and Shera, 1995).

For ratios of f2/f1 < 1.1 both BM-DP and DPOAE frequency responses for 2f1–f2 exhibit a phase lag that increases fairly rapidly [Fig. 10(D)]. The rapid phase increase as a function of frequency suggests that the DPOAE response must arise from the place location (Knight and Kemp, 2001). However, ripples in the DPOAE frequency response [Fig. 10(C)] indicate the interaction of two sources for which neither source is dominant. For ratios >1.1 the DPOAE phase is relatively constant implying that the dominant component arises from the wave location based on log scaling in the cochlea (Knight and Kemp, 2001). Similar results were found here for the chinchilla though only for the highest f2/f1 ratio=1.45 [Fig. 10(D)]. The DPOAE phase response can have considerable fine structure with up to 360° variations that appear to correlate with the ripple pattern in the amplitude response (Fig. 9), usually for DP frequencies <5 kHz.

IO notches in DPOAE curves are also apparent in a set of frequency functions when data are collected in a swept frequency paradigm using multiple levels (e.g., 2.2 kHz in Fig. 9). Dips of over 40 dB occur at some level greater than the minimum level used resulting in an IO notch at the dip frequency. The largest dips occur when the primary levels are around 60 dB SPL. This suggests that in collecting these functions for the purpose of estimating cochlear status that levels lower or higher than 60 dB SPL would result in less complicated functions. Most DP-IO functions have a "shoulder" around 60 dB SPL followed by a large increase at a rate faster than linear. Whether this results from switching between "sources" or a result that above 60 dB SPL vibration in the tail region becomes significant enough for reflection to take place along a large extent of the cochlea is not resolved. Clearly the growth of DPOAEs is a complex function of 2f1–f2 and the primary levels. This feature is likely more prominent in chinchilla and may not be of significance to the human measurements.

C. DPOAEs after death

The relatively small changes in DPOAE level upon death (cf. Figs. 13 and 14) are consistent with the observation that high level DPOAEs are relatively unaffected by trauma (Whitehead *et al.*, 1996; Mills and Rubel, 1994). This could result if the high level DPOAE consists primarily of the linear reflection component of OAEs. As shown, the BM-DP decreases rapidly on a time scale that correlates directly with the time course of the EP (Sewell, 1984), leaving strictly passive mechanics. However, a couple of hours after death even the passive emissions disappear. This suggests that it is not only passive mechanics that produce the DPOAE because passive linear properties remain long after death, but likely some other short-term nonlinearity possibly associated with the cilia of the hair cells that disappears after a couple of hours. It has been shown that there is a nonlinear component in BM-IO response at high levels (growth rate of 0.9 dB/dB for levels >70 dB SPL) for a duration of up to 2 h after death (Rhode, 2007a). This high-level nonlinearity may account for the persistent DPOAE after trauma. In humans, Whitehead (1998) argued that there is no convincing evidence for discrete active/low level and passive/high level DPOAEs and cautioned about extrapolating in this regard from rodent data.

D. Two DPOAE sources and notches in IO functions

The earliest mention of notches in response to pure tones in cat auditory nerve fiber IO functions indicated they occur

at fairly high levels between 90 and 100 dB SPL (Kiang and Moxon, 1972). This was later reinforced by several studies (e.g., Kiang *et al.*, 1986; Ruggero *et al.*, 1996; Lin and Guinan, 2004). The interpretation is that there must be two out-of-phase components that under equal amplitude conditions cancel each other to varying degrees. There are ripples in BM frequency transfer curves at a frequency spacing of around 800 Hz on the high frequency side of CF that can be explained by the addition of a compression wave with a fast velocity that with equal amplitude and opposite phase cancels the traveling wave component of vibration (Rhode and Recio, 2000; Rhode, 2007a). Notches in the IO responses occur at the frequency of each minimum of the ripples at levels that vary between 50 and 90 dB SPL depending on the separation of f1 from CF (Rhode, 2007a). Whether these notches have anything to do with the DPOAE notches is debatable because the latter notches are fairly consistent at a 60 dB SPL level though it is unknown what the DPOAE response would be if primary levels of 90–100 dB SPL were used. However, the IO functions shown in Fig. 1 support a strong correlation between the two.

There is a consensus that there are two spatially separated OAE sources producing emissions that have been named wave-fixed (nonlinear) and place-fixed (reflection) waves (e.g., Kemp, 2002; Shera and Guinan, 1999). Whether they are always both present or whether one dominates over various ranges of frequency ratios and levels remains a matter of study, though it is likely that both are always present in varying proportions. It has been shown that for small f2/f1 ratios that the place component dominates (Knight and Kemp, 2001). This is demonstrated by the fact that the phase characteristic shows a rapidly increasing phase lag as the frequencies are increased while maintaining a constant f2/f1 ratio and also occurs in the BM-DP response [Fig. 10(B)]. In contrast, the DPOAE wave-fixed source phase characteristic does not vary much and rests on the log-invariant scaling of the frequency dimension. This was observed in Fig. 10(D) for f2/f1 ratios >1.1. While the BM-DP phase is nearly constant for $f2/f1=1.45$, it decreased rapidly at $f2/f1=1.22$ in contrast to the DPOAE phase remaining nearly constant at that ratio.

Two components for the DPOAEs have been demonstrated by a variety of methods including the use of a suppressor tone (f3) placed near the DP location (Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001). This tends to decrease the ripple seen in the curves showing DPOAE as a function of DP frequency. That is, the place component is suppressed by the third tone leaving only the wave location to generate the DPOAE. The ripples in BM functions are interpreted to result from the mutual interference of two processes (e.g., Recio *et al.*, 1998). A traveling wave and a fast wave exist in cochlear mechanics and account for a notch in IO functions and ripples in frequency transfer functions (Rhode, 2007a). The present data suggest that there is a correlation between the two (Fig. 1). The issue of two sources for DPOAE emissions is not directly addressed by the two waves in BM mechanics interacting at the same location where two spatially separated sites are posited to give rise to emissions. While there is a direct correlation with the BM traveling wave and the wave source DPOAE, the relation of the fast wave in cochlear mechanics to DPOAEs remains a topic for discussion and exploration with cochlear models.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Deafness and Communication Disorders, Grant No. R01 DC 1910. Thanks to C. D. Geisler, Michael Gorga, Steven Neely, and Kim Schairer for their reviews of this manuscript. Special thanks to the two anonymous reviewers for their very thorough reading and comments.

Supplementary material for Figures 2, 3, and 11 can be found at http://www.physiology.wisc.edu/rhode/suppplement/DPOAE/jasa/2007 last viewed on 9/19/2007.

Avan, P., Bonfils, P., Gilain, L., and Mom, T. (**2003**). "Physiological significance of distortion-product otoacoustic emissions at 2f1-f2 produced by high-versus low-level stimuli," J. Acoust. Soc. Am. **113**, 430–441.

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. (**1987**). "Acoustic distortion from rodent ears, A comparison of responses from rats, guinea pigs and gerbils," Hear. Res. **31**, 25–38.

Brown, A. M., McDowell, B., and Forge, A. (**1989**). "Acoustic distortion products can be used to monitor the effects of chronic gentamicin treatment," Hear. Res. **42**, 143–156.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1984**). "Evoked mechanical responses of isolated cochlear outer hair cells," Science **227**, 194–196.

Cooper, N. P. (**1999**). "Vibration of beads placed on the basilar membrane in the basal turn of the cochlea," J. Acoust. Soc. Am. **106**, L59–L64.

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., and Gorga, M. P. (**2001**). "Distortion-product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook*, Hill-Fay Associates, Winnetka, IL.

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.

He, N., and Schmiedt, R. A. (**1993**). "Fine structure of the 2f1-f2 acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

Janssen, T., Kummer, P., and Arnold, W. (**1998**). "Growth behavior of the 2f1-f2 distortion product otoacoustic emission in tinnitus," J. Acoust. Soc. Am. **103**, 3418–3430.

Jaramillo, F., Markin, V. S., and Hudspeth, A. J. (**1993**). "Auditory illusions and the single hair cell," Nature (London) **364**, 527–529.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (**2006**). "Influence of primary-level and primary-frequency ratio on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 418–428.

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.

Kemp, D. T. (**1980**). "Toward a model for the origin of cochlear echoes," Hear. Res. **2**, 533–548.

Kemp, D. T. (**1986**). "Otoacoustic emissions, traveling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kemp, D. T. (**2002**). "Exploring cochlear status with otoacoustic emission. The potential for new clinical applications," in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–47.

Kiang, N. Y. S., Liberman, M. C., Sewell, W. F., and Guinan, J. J. (**1986**). "Single unit clues to cochlear mechanisms," Hear. Res. **22**, 171–182.

Kiang, N. Y. S., and Moxon, E. C. (**1972**). "Physiological considerations in artificial stimulation of the inner ear," Ann. Otol. Rhinol. Laryngol. **81**, 1–17.

Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D. L., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). "Optimal L1–L2 primary tone level separation remains independent of test frequency in humans," Hear. Res. **146**, 47–56.

Lin, T., and Guinan, J. J., Jr. (**2004**). "Time-frequency analysis of auditory-nerve fiber and basilar membrane click responses reveal glide irregularities and non-characteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.

Lukashkin, A. N., Lukashkin, V. A., and Russell, I. J. (**2002**). "One source for distortion product otoacoustic emissions generated by low- and high-level primaries," J. Acoust. Soc. Am. **111**, 2740–2748.

Mills, D. M. (**1997**). "Interpretation of distortion product otoacoustic emission measurements. I. Two stimulus tones," J. Acoust. Soc. Am. **102**, 413–429.

Mills, D. M., and Rubel, E. W. (**1994**). "Variation of distortion product otoacoustic emissions with furosemide injections," Hear. Res. **77**, 183–199.

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.

Neely, S. A., Johnson, T. A., and Gorga, M. P. (**2005**). "Distortion-product otoacoustic emissions measured with continuously varying stimulus level," J. Acoust. Soc. Am. **117**, 1248–1259.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.

Rhode, W. S. (**2007a**). "Basilar membrane mechanics in the 6–9 kHz region of sensitive chinchilla," J. Acoust. Soc. Am. **121**, 2792–2804.

Rhode, W. S. (**2007b**). "Mutual suppression in the 6 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **121**, 2805–2818.

Rhode, W. S., and Recio, A. (**2000**). "Study of the mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986a**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364–1374.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986b**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364–1374.

Ruggero, M. A., Rich, N. C., Shivapuja, B. G., and Temchin, A. N. (**1996**). "Auditory-nerve responses to low-frequency tones: Intensity dependence," Aud. Neurosci. **2**, 159–185.

Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944–966.

Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs." J. Acoust. Soc. Am. **105**, 782–798.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," J. Acoust. Soc. Am. **72**, 131–141.

Sewell, W. F. (**1984**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hear. Res. **14**, 305–314.

Siegel, J. H., Cerka, A. J., Recio-Spinosa, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.

Whitehead, M. L. (**1998**). "Species differences of distortion-product otoacoustic emissions: 'Comment on interpretation of distortion product otoacoustic emission measurements. I. Two stimulus tones'," J. Acoust. Soc. Am. **103**, 2740–2742.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992**). "Evidence for two discrete sources of 2f1-f2 distortion-product otoacoustic emission in rabbit. II. Differential physiological vulnerability," J. Acoust. Soc. Am. **92**, 2662–2682.

Whitehead, M. L., Lonsbury-Martin, B. L., Martin, G. K., and McCoy, M. J. (**1996**). "Otoacoustic emissions: Animal models and clinical observations," in *Clinical Aspects of Hearing*, SHAR VI, edited by T. R. Van De Water, A. N. Popper, and R. R. Fay (Springer, New York), pp. 199–257.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ear. I. Effects of decreasing L_2 below L_1," J. Acoust. Soc. Am. **97**, 2346–2358.

Zurek, P. M., Clark, W. W., and Kim, D. O. (**1982**). "The behavior of acoustic distortion products in the ear canal of chinchillas with normal or damaged ears," J. Acoust. Soc. Am. **72**, 774–780.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Laser amplification with a twist: Traveling-wave propagation and gain functions from throughout the cochlea

Christopher A. Shera
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Department of Otology & Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 28 May 2007; revised 20 August 2007; accepted 21 August 2007)

Except at the handful of sites explored by the inverse method, the characteristics—indeed, the very existence—of traveling-wave amplification in the mammalian cochlea remain largely unknown. Uncertainties are especially pronounced in the apex, where mechanical and electrical measurements lack the independent controls necessary for assessing damage to the preparation. At a functional level, the form and amplification of cochlear traveling waves are described by quantities known as propagation and gain functions. A method for deriving propagation and gain functions from basilar-membrane mechanical transfer functions is presented and validated by response reconstruction. Empirical propagation and gain functions from locations throughout the cochlea are obtained in mechanically undamaged preparations by applying the method to published estimates of near-threshold basilar membrane responses derived from Wiener-kernel (chinchilla) and zwuis analysis (cat) of auditory-nerve responses to broadband stimuli. The properties of these functions, and their variation along the length of the cochlea, are described. In both species, and at all locations examined, the gain functions reveal a region of positive power gain basal to the wave peak. The results establish the existence of traveling-wave amplification throughout the cochlea, including the apex. The derived propagation and gain functions resemble those characteristic of an active optical medium but rotated by 90° in the complex plane. Rotation of the propagation and gain functions enables the mammalian cochlea to operate as a wideband, hydromechanical laser analyzer.
 [DOI: 10.1121/1.2783205]

PACS number(s): 43.64.Kc, 43.64.Bt, 43.60.Pt [BLM] Pages: 2738–2758

I. INTRODUCTION

Apart from the spontaneous emission of amplitude-stabilized sounds (Bialek and Wit, 1984; Shera, 2003a), the strongest evidence for traveling-wave amplification in the mammalian cochlea comes from measurements of basilar-membrane motion analyzed using the "inverse method" (e.g., Zweig, 1991; de Boer, 1983, 1995b; de Boer and Nuttall, 2001). Although compelling within their domain, the method's conclusions are restricted to a limited number of surgically accessible measurement locations, all in the basal turns of the cochlea. Other than at the handful of sites probed by inverse analysis, the form and characteristics—indeed, the very existence—of traveling-wave amplification remain largely unknown. Uncertainties are especially pronounced in the apex, where response measurements lack the independent controls necessary for gauging damage to the preparation (reviewed in Robles and Ruggero, 2001). Perhaps as a result, apical studies report the gamut from physiologically vulnerable nonlinearities reminiscent of the base to signs of active attenuation rather than amplification (e.g., Cooper and Rhode, 1995; Rhode and Cooper, 1996; Khanna and Hao 1999; Zinn *et al.* 2000). Even if the various findings were reliable and consistent, neither mechanical nonlinearity nor physiological lability provides evidence for actual power gain (or loss).[1] What's needed is a way of determining in uncompromised preparations the characteristics of traveling-wave amplification throughout the cochlea.

A. Propagation and gain functions

At a functional level, the form and amplification of cochlear traveling waves are described by quantities known as propagation and gain functions. By way of introduction, consider a simpler, textbook example of wave propagation: the transverse vibration of a stretched string (e.g., Georgi, 1993; Zwiebach, 2004). When the driving force is sinusoidal, the displacement, $d(x,t)$, of a uniform semi-infinite string (Fig. 1) has the wave-like form

$$d(x,t) = d_0 e^{\gamma x} \cos(\omega t - \kappa x + \phi_0), \tag{1}$$

where $\omega = 2\pi f$, and d_0 and ϕ_0 are the initial amplitude and phase. The spatial pattern of the wave on the string is determined by the propagation and gain coefficients, denoted by κ and γ, respectively; both have dimensions of inverse length. The propagation coefficient, κ, determines the spatial period (or wavelength, λ) of the wave via the equation $\lambda = 2\pi/\kappa$. The gain coefficient, γ, characterizes the effects of energy gains or losses and thereby determines whether, and how rapidly, the wave amplitude increases or decreases as it travels. For example, when the string loses vibrational energy due to viscous damping, γ is negative and the wave amplitude, controlled by the factor $e^{\gamma x}$, diminishes with increasing x. In general, both the propagation and gain coefficients depend on frequency. For example, $\kappa(f) = 2\pi f/c$, where $c = \sqrt{T/\rho}$ is the constant wave speed, here determined by the ratio of string tension, T, and mass density, ρ. For conve-

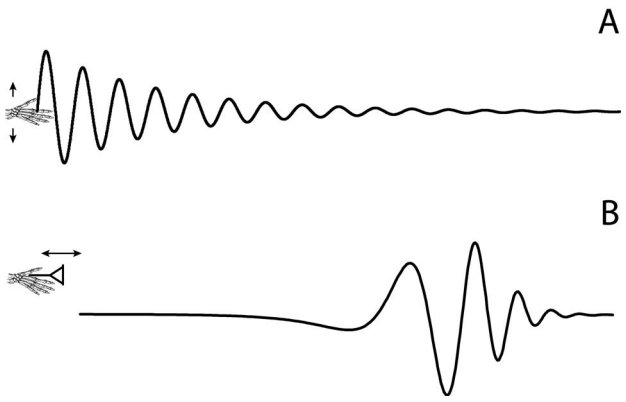


FIG. 1. Wave motion along a string and on the basilar membrane. The top panel (A) shows the transverse vibration of a uniform stretched string driven sinusoidally on the left. The propagation and gain coefficients determine the spatial pattern of the wave along the string. The change of wave amplitude reflects the sign of the string gain coefficient, γ. When the wave amplitude decreases with distance, as shown, the wave loses energy as it propagates (e.g., due to viscous damping in the surrounding medium), and $\gamma<0$. The bottom panel (B) shows the transverse vibration of the basilar membrane (BM) produced by sinusoidal motion of the stapes. Because the stiffness and other properties of the partition vary with position, the propagation and gain functions depend on x, and the sign of the gain function cannot simply be read off from the behavior of the wave envelope. For example, the declining stiffness of the partition causes BM vibration to increase to a maximum before decreasing, even when γ is everywhere negative (e.g., in a passive cochlea).

nience, κ and γ are often combined into a single complex number, k, known as the complex wavenumber and defined by $k \equiv \kappa + i\gamma$. The wave $d(x,t)$ can then be written in the more compact form

$$d(x,t) = \mathrm{Re}\{D_0 \exp i[\omega t - kx]\}, \tag{2}$$

where $D_0 \equiv d_0 e^{i\phi_0}$.

When the properties of the medium vary with position, as they do along a nonuniform string or within the cochlea, the waves manifest a somewhat more complicated relationship to the wave number. For example, the displacement of the basilar membrane (BM) produced by a pure tone has the approximate form

$$d_{\mathrm{BM}}(x,t) \cong \mathrm{Re}\left\{ D_0(f) \left[\frac{k(x,f)}{k(0,f)} \right]^{3/2} \times \exp i \left[\omega t - \int_0^x k(x',f)\,dx' \right] \right\}, \tag{3}$$

where the wave number $k(x,f)$ and its constituent propagation and gain functions $\kappa(x,f)$ and $\gamma(x,f)$ depend explicitly on position (e.g., Zweig *et al.*, 1976; de Boer, 1996). Although more intimidating in appearance than Eq. (2) for the wave on a uniform string, Eq. (3) for $d_{\mathrm{BM}}(x,t)$ is in many ways a straightforward generalization. The integral in the complex exponential simply sums the accumulating phase shift and gain as the wave propagates along. The prefactor out in front ($k^{3/2}$) arises because of energy conservation, which requires that the wave amplitude change when $k(x,f)$ changes, even in the absence of energy gains or losses (i.e., even when $\gamma=0$). In the cochlea, for example, the declining stiffness of the partition causes the wave speed and wavelength to decrease as the wave travels away from the stapes. As a result, the propagation function and prefactor increase with x, passively boosting the amplitude of vibration as the wave approaches its characteristic place. Although changes in wave amplitude are not so simply related to the sign of $\gamma(x,f)$ as they are on a uniform string, the gain function does determine the nature of power transfer to and from the wave. Regions of negative $\gamma(x,f)$ correspond to net power dissipation (e.g., via viscous losses within the organ of Corti); regions of positive $\gamma(x,f)$ correspond to power amplification.

B. Overview

Despite their significance for characterizing the form and power amplification of basilar-membrane traveling waves, the cochlear propagation and gain functions remain experimentally undetermined. Here we present a method for deriving propagation and gain functions from measurements of BM transfer functions. We apply the method to a variety of published estimates of BM motion, most notably to those obtained by (1) second-order Wiener-kernel analysis of chinchilla auditory-nerve-fiber (ANF) responses to near-threshold noise (Recio-Spinoso *et al.*, 2005) and (2) zwuis analysis of cat ANF responses to complex tones (van der Heijden and Joris, 2003, 2006). Both the Wiener-kernel (reviewed by Eggermont, 1993) and zwuis procedures estimate BM motion by extracting high-frequency timing information encoded in the neural response envelope by cochlear nonlinearities, principally the half-wave rectification that occurs at the inner hair cell synapse. When corrected for synaptic and neural transmission delays, the Wiener-kernel estimates closely resemble BM mechanical measurements made at corresponding locations and intensities (Temchin *et al.*, 2005). Because the broadband stimuli employed by the ANF-based procedures serve, in effect, to linearize the mechanical response (de Boer, 1997c; van der Heijden and Joris, 2006), the Wiener-kernel and zwuis measurements provide characterizations of cochlear tuning well suited to the linear inversion procedures employed here. Our analysis of the Wiener-kernel, zwuis, and other data determines cochlear propagation and gain functions throughout the cochlea.

II. THE INVERSION PROCEDURE

The inversion procedure described here uses measurements of the basilar-membrane traveling wave to find the cochlear propagation and gain functions.[2] At stimulus intensities in the linear regime near the threshold of hearing, the normalized mechanical response of the basilar membrane (e.g., BM velocity relative to stapes motion) depends on both the distance from the stapes, x, and the stimulus frequency, f. In the discussion that follows, we denote the normalized BM velocity by $V_{\mathrm{BM}}(x,f)$. Basilar-membrane traveling waves are iso-frequency slices of $V_{\mathrm{BM}}(x,f)$, and we denote the traveling wave at frequency f_{o} by $V_{\mathrm{BM}}(x,f_{\mathrm{o}})$, where the subscript "o" stands for "observation" and indicates that the stimulus frequency is regarded as a discrete parameter that is held constant (often at one of a relatively small number of values) while x varies continuously.

A. Finding traveling waves from transfer functions

With rare exceptions (e.g., von Békésy, 1960; Ren, 2002), BM mechanical responses are measured not as traveling waves but as transfer functions—not, that is, as $V_{BM}(x,f_o)$, a spatial response at fixed f_o, but as $V_{BM}(x_o,f)$, a frequency response at fixed x_o. To convert the transfer function $V_{BM}(x_o,f)$ at location x_o into an estimate of the traveling wave $V_{BM}(x,f_o)$ at frequency f_o, we apply the local scaling symmetry (Zweig, 1976) manifest by basilar-membrane transfer functions (Rhode, 1971; Gummer *et al.*, 1987) and neural tuning curves (e.g., Kiang and Moxon, 1974; Liberman, 1978). In its traditional form, local scaling implies that rather than depending on position and frequency independently, $V_{BM}(x,f)$ depends on the two variables x and f primarily in the dimensionless combination $\beta(x,f) \equiv f/\mathrm{CF}(x)$, where $\mathrm{CF}(x)$ is the characteristic frequency at location x (i.e., the cochlear position-frequency map). When the BM velocity scales, the function $V_{BM}[\beta(x,f)]$ describes both the transfer function and the traveling wave. In other words, frequency-domain measurements of the transfer function $V_{BM}(x_o,f)$, if plotted versus $f/\mathrm{CF}(x_o)$, provide spatial-domain measurements of the traveling wave $V_{BM}(x,f_o)$ plotted versus $f_o/\mathrm{CF}(x)$. If the scaling is local, the univariate description is most accurate in the region about the peak of the response [i.e., for x near x_o and for f near $\mathrm{CF}(x_o)$].

Figure 2 illustrates these remarks using an estimate of the BM click response at the 9-kHz place of the chinchilla cochlea obtained using second-order Wiener-kernel analysis of ANF responses to noise (Recio-Spinoso *et al.*, 2005). Panel (A) of Fig. 2 shows the Wiener-kernel estimate $v_{BM}(x_o,t)$ measured from a fiber with $\mathrm{CF}(x_o)=9$ kHz. In the time domain, scaling implies that BM click responses depend on x and t through the dimensionless combination $t\times\mathrm{CF}(x)$ representing time measured in periods of the local CF (Shera, 2001). Panel (B) shows the magnitude and phase of the transfer function $V_{BM}(x_o,f)$, found by computing the Fourier transform of the click response: $V_{BM}(x_o,f) = \mathcal{F}\{v_{BM}(x_o,t)\}$. The abscissa shows $\beta \equiv f/\mathrm{CF}(x_o)$ on a logarithmic axis. Local scaling implies that the same graph provides estimates of the envelope and phase lag of the traveling wave $V_{BM}(x,f_o)$ at frequency $f_o \cong \mathrm{CF}(x_o)$. To interpret the axis spatially, note that equal intervals along the axis represent equal intervals of space. (When the cochlear map is exponential, $\ln\beta$ is a linear function of x at fixed frequency.) Panel (C) shows a snapshot of the wave at one instant of time. Since they have units involving spikes/s that are not easily converted to those of velocity, Wiener-kernel estimates of BM velocity responses are customarily displayed in normalized form (Recio-Spinoso *et al.*, 2005).

1. Generalized local scaling in the apex

In many species, including the cat and chinchilla considered here, the cochlear position-frequency map deviates from a purely exponential form near its apical end (e.g., Liberman, 1982; Eldredge *et al.*, 1981; Greenwood, 1990). In these species, the map is better approximated by the "subtracted exponential"

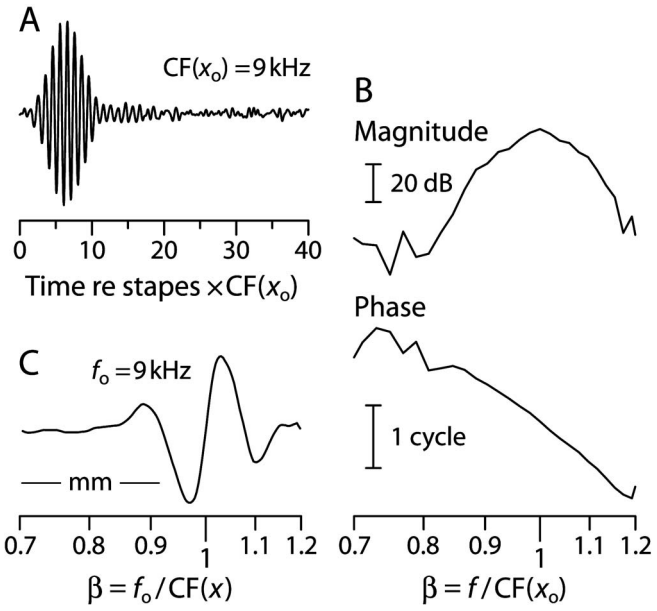


FIG. 2. (Panel A) Auditory-nerve based estimate of the BM click response, $v_{BM}(x_o,t)$, at the cochlear location x_o tuned to approximately 9 kHz in chinchilla (Recio-Spinoso *et al.*, 2005). The response estimate has been normalized by its peak value. Time, shown along the abscissa in units of the CF period, is measured relative to the approximate onset of stapes vibration by subtracting out estimates of acoustic and synaptic transmission delays amounting to a total of 1.225 ms (Temchin *et al.*, 2005). (Panel B) The magnitude and phase of the Fourier transform of $v_{BM}(x_o,t)$ provide an estimate of the BM mechanical transfer function, $V_{BM}(x_o,f)$, at the cochlear location x_o. Frequency, normalized by $\mathrm{CF}(x_o)=9$ kHz, increases along the logarithmic abscissa. (Panel C) Application of local scaling provides an estimate of the traveling wave by reinterpreting the abscissa $f/\mathrm{CF}(x)$ as a spatial axis at fixed frequency. The figure shows a snapshot of the 9-kHz wave whose envelope and phase are shown in Panel (B) The 1 mm scale bar is based on estimates of the chinchilla cochlear map (Eldredge *et al.*, 1981; Greenwood, 1990).

$$\mathrm{CF}(x) = [\mathrm{CF}(0) + \mathrm{CF}_1]e^{-x/l} - \mathrm{CF}_1, \tag{4}$$

which morphs from exponential to more linear behavior at characteristic frequencies near the transition frequency, CF_1.[3] We compensate for this apical deviation from the exponential map by working with a "generalized" local scaling variable that reduces to the traditional variable $\beta(x,f)$ in the base of the cochlea. This new scaling variable, denoted $\nu(x,f)$, is defined by

$$\nu(x,f) \equiv \frac{f+\mathrm{CF}_1}{\mathrm{CF}(x)+\mathrm{CF}_1} = \frac{\beta(x,f)+\beta_1(x)}{1+\beta_1(x)}, \tag{5}$$

where $\beta_1(x) \equiv \mathrm{CF}_1/\mathrm{CF}(x)$. In the exponential portion of the cochlear map, the two variables $\beta(x,f)$ and $\nu(x,f)$ are nearly equivalent; the distinction becomes significant only at CFs lower than about two octaves above CF_1. (In the cat, $\mathrm{CF}_1 \cong 365$ Hz; in the chinchilla, $\mathrm{CF}_1 \cong 140$ Hz.)

The form of the generalized local scaling variable $\nu(x,f)$ was chosen to preserve the linear relation between x and $\ln\beta$ mentioned above: At fixed frequency, equal intervals of $\ln\nu$ correspond to constant distances along the BM. Use of the generalized scaling variable improves the validity of the local scaling approximation in the apex. When $V_{BM}(x,f)$ is approximated by a function of $\beta(x,f)$, the resulting transfer functions necessarily have the same quality factors (Q val-

FIG. 3. Spacetime slice through a symmetric, two-dimensional box model of height H. The x axis extends longitudinally from the base ($x=0$), and the y axis is oriented perpendicular to the basilar membrane, which spans the entire width, b, of the cochlea. The snapshot shown here has caught the BM participating in a traveling wave, whose vertical displacements have been hugely exaggerated for the purposes of illustration. The inversion procedure described in the text finds the wave's propagation and gain functions by analyzing measurements of the spatial displacement pattern.

ues) at all locations. But when $V_{\mathrm{BM}}(x,f)$ is approximated by a function of $\nu(x,f)$, the corresponding transfer functions are more broadly tuned (i.e., have smaller Q values) at more apical locations.[4] This broadening of tuning in the apex is qualitatively consistent with the trends observed experimentally (e.g., Liberman, 1978; Robles and Ruggero, 2001).

B. Finding propagation and gain functions from traveling waves

To find the wave number from the traveling wave we relate the two using a model of cochlear mechanics. We avoid unnecessary complication by working with the simplest geometry that manifests both long- and short-wave behavior. Figure 3 illustrates the linear, two-dimensional box model of the cochlea with incompressible and inviscid scala fluids. (With straightforward modifications the derivation reviewed below is easily extended to the three-dimensional model.) Basilar-membrane motion is driven by the antisymmetric (or difference) component of the pressure. By averaging the difference pressure over the scalae height and applying boundary conditions at the scalae walls and BM, one can show that the averaged difference pressure, $\bar{P}(x,f)$,[5] satisfies the one-dimensional wave equation

$$(\partial_x^2 + k^2)\bar{P} = 0, \tag{6}$$

where $k(x,f)$ is the complex wave number (Shera *et al.*, 2005; Duifhuis, 1988; Talmadge *et al.*, 2001).

To find $k(x,f)$ we note that the BM velocity $V_{\mathrm{BM}}(x,f)$ is related to the second spatial derivative of the pressure through the equation

$$\partial_x^2 \bar{P} = bZ_{\mathrm{f}}V_{\mathrm{BM}}, \tag{7}$$

where b is the BM width and $Z_{\mathrm{f}}(f)$ is the effective acoustic impedance of the fluids [notation adopted from Shera *et al.* (2005)]. If the function $V_{\mathrm{BM}}(x,f)$ is known—whether by model or by measurement—Eq. (7) can be solved for $\bar{P}(x,f)$ by double integration:

$$\bar{P}(x,f) = bZ_{\mathrm{f}}(f)\int_x^L dx' \int_{x'}^L V_{\mathrm{BM}}(x'',f)dx'', \tag{8}$$

where x' and x'' are dummy integration variables. The constants of integration are chosen to satisfy the boundary conditions (in this case, $\bar{P}=\partial_x\bar{P}=0$ at the helicotrema).[6] Combining Eqs. (6)–(8) and solving for the wave number yields

$$k^2(x,f) = -V_{\mathrm{BM}}(x,f) \Big/ \int_x^L dx' \int_{x'}^L V_{\mathrm{BM}}(x'',f)dx''. \tag{9}$$

This *wave number inversion formula* indicates how to find the complex wave number from the traveling wave. Derived in an earlier publication (Shera *et al.* 2005), Eq. (9) generalizes the long-wave procedure suggested by de Boer (1995a). The propagation and gain functions are then found by taking the real and imaginary parts of the wave number:

$$\kappa(x,f) = \mathrm{Re}\, k(x,f) \text{ and } \gamma(x,f) = \mathrm{Im}\, k(x,f). \tag{10}$$

The sign of the square root needed to obtain k from k^2 is set by the presumption that the wave travels forward in the vicinity of its peak [i.e., $\kappa(\nu\cong 1)>0$].[7]

Inversion formula (9) has a number of happy features that render it well suited to our application. First and foremost, the derived wave number $k(x,f)$ depends only on the given BM response function, $V_{\mathrm{BM}}(x,f)$, and is independent of the value of H, the effective height of the model. [Derived here using a 2D model, the inversion formula generalizes to 3D if $V_{\mathrm{BM}}(x,f)$ is understood as the velocity of the BM averaged over its width.] In other words, the function $V_{\mathrm{BM}}(x,f)$ yields the same wave number irrespective of whether the assumed hydrodynamic model is everywhere long wave or short wave, or manifests a transition between the two, as in the real cochlea. In addition, the inversion procedure does not assume that the measured $V_{\mathrm{BM}}(x,f)$ contains only a forward-traveling wave; contributions from both forward- and reverse-traveling components are allowed.[8] Furthermore, the procedure involves neither algorithmic iteration nor numerical differentiation of the measured response, both of which require special care and handling (Zweig, 1991) to suppress confounding magnification of error. Finally, note that Eq. (9) for $k^2(x,f)$ is independent of the overall magnitude of $V_{\mathrm{BM}}(x,f)$; any constant scale factor cancels in the ratio. The fact that the wave number depends on the shape and fine structure of the traveling wave, but not on its absolute scale, makes possible its determination from the indirect estimates of BM motion provided by Wiener-kernel and zwuis analysis.

Figure 4 shows the propagation and gain function $\kappa(\nu)$ and $\gamma(\nu)$ obtained from Eq. (9) using the Wiener-kernel estimate of the traveling wave $V_{\mathrm{BM}}(\nu)$ shown in Fig. 2. The spatial integrals were evaluated using the differential relation $d\nu/\nu = dx/l$, where $l\cong 3.8$ mm is the space constant of the chinchilla cochlear map (Eldredge *et al.*, 1981; Greenwood, 1990). Note that the derived propagation and gain functions become unreliable outside the peak region (i.e., at locations $|\nu-1|\gtrsim 0.2$ in this example), where noise contaminates the Wiener-kernel transfer function.

C. Validation by reconstruction

To validate the inversion procedure we used the Wentzel–Kramers–Brillouin (WKB) approximation to reconstruct the traveling wave from the derived wave number (e.g., Mathews and Walker, 1964; Zweig *et al.*, 1976). Combining the WKB solution to Eq. (6) for $\bar{P}(x,f)$ with Eq. (7) for $V_{\mathrm{BM}}(x,f)$ yields the formula

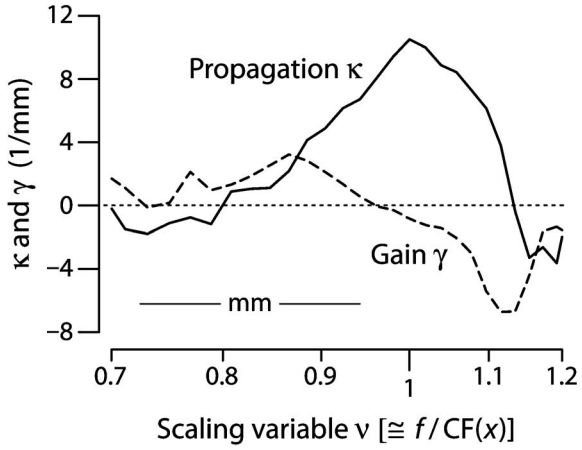


FIG. 4. Derived propagation and gain functions. Plotted vs the generalized scaling variable using solid and dashed lines, respectively, the functions $\kappa(\nu)$ and $\gamma(\nu)$ were obtained by inversion from the estimate of $V_{\mathrm{BM}}(\nu)$ shown in Fig. 2 (CF=9 kHz). Because $\mathrm{CF}_1 \ll 9$ kHz, ν is nearly equivalent to the normalized frequency $f/\mathrm{CF}(x)$. Parameters for the chinchilla cochlear map were taken from Greenwood (1990). For reference, thin dashes mark the zero line. The scale bar represents a distance of 1 mm.

$$V_{\mathrm{BM}}(x,f) \propto k^{3/2}(x,f)\exp\left[-i\int_0^x k(x',f)dx'\right], \tag{11}$$

where the proportionality factor depends on f. Readers may recognize the WKB formula from its earlier appearance in Eq. (3). Figure 5 shows that both the amplitude and the phase of the reconstructed response are in good agreement with the original. Thus, independent of their origin, the propagation and gain functions shown in Fig. 4 evidently provide a valid representation of the complex wave number of the traveling wave.

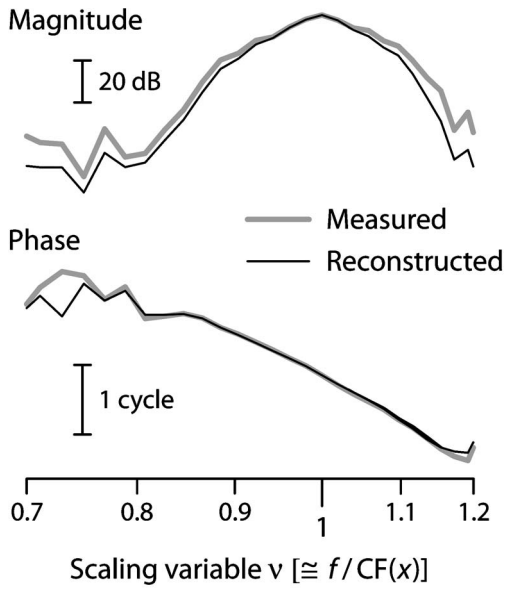


FIG. 5. Traveling wave/transfer function reconstructed from the derived wave number using the WKB approximation. The reconstructed response (thin solid line) was obtained from the wave number in Fig. 4 using the WKB formula [Eq. (11) for $V_{\mathrm{BM}}(x,f)$] and evaluating the integral using generalized scaling. An overall complex scale factor was determined by matching the data at the peak. For comparison, the original Wiener-kernel measurements are reproduced from Fig. 2 (thick gray line).

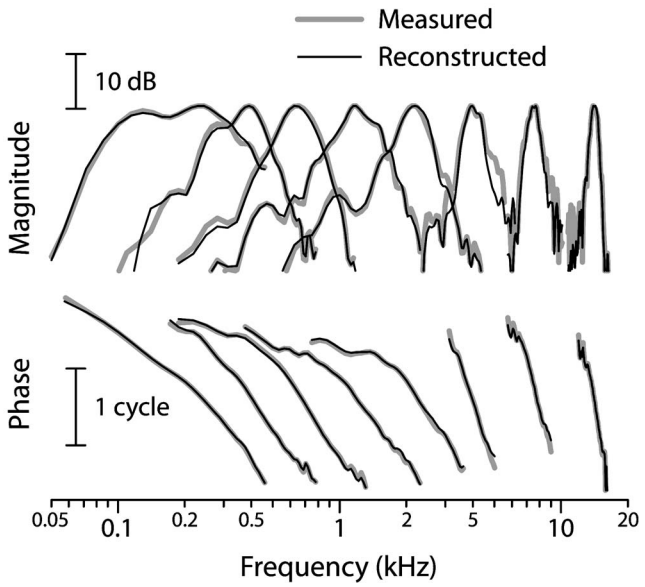


FIG. 6. Original and reconstructed Wiener-kernel transfer functions from locations throughout the chinchilla cochlea. Transfer functions are shown normalized to the same peak amplitude.

We applied the inversion and reconstruction procedure outlined above to each of the 137 near-threshold Wiener-kernel BM click responses used to assemble the recently published map of near-CF BM group delays (see Fig. 13 of Temchin *et al.* 2005). More than 60% of the Wiener kernels (86/137) yielded reconstructions judged satisfactory by visual inspection.[9] Unsuccessful reconstructions often contained anomalous spikes or other discontinuous behavior within the peak region. Similar problems can occur when inverting model responses if the boundary conditions assumed by the inversion formula are not well satisfied. Figure 6 shows typical comparisons of measured and successfully reconstructed transfer functions at CFs spanning the length of the chinchilla cochlea. The quality of the reconstructions demonstrates that the wave number inversion procedure can be successfully applied throughout the cochlea. For logical consistency, we restrict further attention to results obtained from wave numbers validated using the reconstruction procedure.

III. TRAVELING-WAVE PROPAGATION AND GAIN FUNCTIONS

A. Qualitative features

Figure 7 shows $\kappa(\nu)$ and $\gamma(\nu)$ successfully derived from nine Wiener-kernel estimates of $V_{\mathrm{BM}}(\nu)$ with CFs in the range 8–10 kHz (gray lines) together with trend lines that capture the mean behavior.[10] Although individual Wiener kernels and corresponding propagation and gain functions manifest considerable individual variability in their fine detail, all share the same qualitative form. Regarded as a function of position, the propagation functions $\kappa(\nu)$ start out small in the region basal to the wave peak ($\nu \lesssim 0.85$), increase to a maximum near the characteristic place ($\nu \cong 1$), and then decrease again beyond. (Equivalently, the wavelength starts out large near the stapes, decreases to a mini-

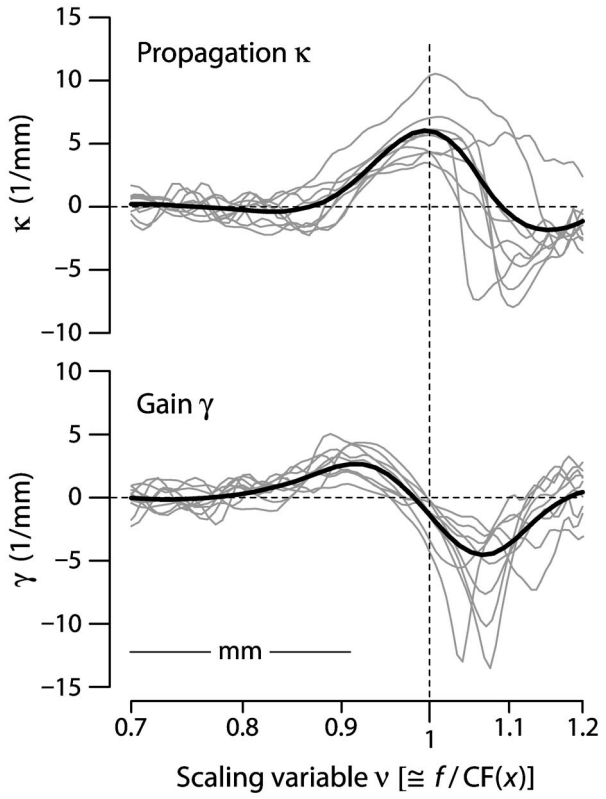


FIG. 7. Propagation and gain functions derived from all nine Wiener kernels with CFs in the range 8–10 kHz. Gray lines show individual functions $\kappa(\nu)$ and $\gamma(\nu)$; black lines show trends obtained by loess fitting (Cleveland, 1993). For reference, thin dashes mark the zero lines and the location of the wave peak ($\nu=1$). The scale bar represents a distance of 1 mm.

mum in the vicinity of the peak, and then increases in the apical "cutoff" region where the wave amplitude decreases rapidly.) At even more apical locations, κ generally passes through zero and goes slightly negative, beyond which the Wiener kernels are generally too noisy to provide reliable trends. The decrease in κ apical to the best place reflects the severe attenuation of the response in this region as the wave transitions from traveling to evanescent behavior (see Sec. IV B 3).

Like the propagation functions, the gain functions also manifest pronounced spatial dispersion (i.e., they vary considerably with x at fixed frequency). In particular, $\gamma(\nu)$ starts out close to zero, where its sign is poorly determined, and increases to a maximum located basal to the wave peak ($\nu \cong 0.9$). The gain function then decreases, passes through zero near $\nu=1$, and becomes negative in the cutoff region.

Because of the assumed local scaling, the spatial perspective adopted above can be flipped and the horizontal axis in Fig. 4 regarded instead as a frequency axis at fixed position. Viewed in this complementary way, strong spatial dispersion becomes strong frequency dispersion. The figure then shows that the propagation and gain functions at the 9 kHz place start out small and increase with the frequency of stimulation. Whereas the propagation function increases all the way to CF and then decreases at higher frequencies, the gain function peaks at a frequency below CF and then reverses sign as it passes through zero near CF.

1. Interpretation in terms of energy flow

To interpret the form of the gain function we note that energy conservation implies that at every location the power transferred by the organ of Corti to the fluid must be the negative of that expended moving the partition. Consequently, when the scalae fluids are assumed inviscid the time-averaged power per unit BM area transfered to the traveling pressure-difference wave is given by

$$w(x,f) = -\tfrac{1}{2}\,\mathrm{Re}\,P_0 V_{\mathrm{BM}}^{*}, \tag{12}$$

where the quantity $P_0(x) \equiv P(x, y=0^+)$ is the pressure difference driving the cochlear partition.[11] Equation (12) can be rewritten in terms of the BM velocity and impedance by using the relation $P_0 = bZ_{\mathrm{BM}}V_{\mathrm{BM}}$:

$$w(x,f) = -\tfrac{1}{2} b|V_{\mathrm{BM}}|^2\,\mathrm{Re}\,Z_{\mathrm{BM}}. \tag{13}$$

In the shortwave regime near the response peak, the BM impedance and wave number are related by $Z_{\mathrm{BM}} = -Z_f H/k$.[12] Hence,

$$w(x,f) = -\omega\rho_0|V_{\mathrm{BM}}|^2\,\mathrm{Im}\,\frac{1}{k} = \omega\rho_0\left|\frac{V_{\mathrm{BM}}}{k}\right|^2\gamma, \tag{14}$$

where we have used the definition $Z_f = 2i\omega\rho_0/bH$ (Shera *et al.*, 2005). Note that $\mathrm{sign}(w) = \mathrm{sign}(\gamma)$. In other words, the direction of power transfer to or from the traveling wave is determined by the sign of the gain function, γ. Negative gain functions ($\gamma<0$) correspond to net wave power absorption by the organ of Corti; positive gain functions ($\gamma>0$) correspond to net wave power amplification.

Equation (14) and local scaling enable us to interpret the empirical gain functions shown in Fig. 7 in terms of power flow. Near the 9 kHz place, the derived gain functions imply that the organ of Corti amplifies waves of corresponding frequency as they travel towards their characteristic place and squelches them beyond. Viewed from the complementary perspective of frequency, the CF place amplifies waves with frequencies less than CF and attenuates those with frequencies greater than CF.

The existence of power amplification at high CFs corroborates features of the BM impedance obtained by applying the inverse method to mechanical transfer functions from the base of the cochlea (e.g., Zweig, 1991; de Boer, 1995b; de Boer and Nuttall, 2001). In both species previously examined (squirrel monkey and guinea pig), the BM impedance was found to manifest a bowl-shaped region of negative resistance just basal to the response peak. Since regions of negative resistance correspond to regions of positive γ (see Sec. IV B 3), our findings from the basal end of the chinchilla cochlea are consistent with these previous results.

B. Generalization to other locations, species, and measurements

Employing ANF-derived estimates of cochlear tuning allows us to extend the analysis of cochlear propagation and gain throughout the cochlea. Figures 8–10 demonstrate that the qualitative features of the derived propagation and gain functions evident in Fig. 7 apply not just to chinchilla responses near the 9 kHz place but to cochlear traveling waves

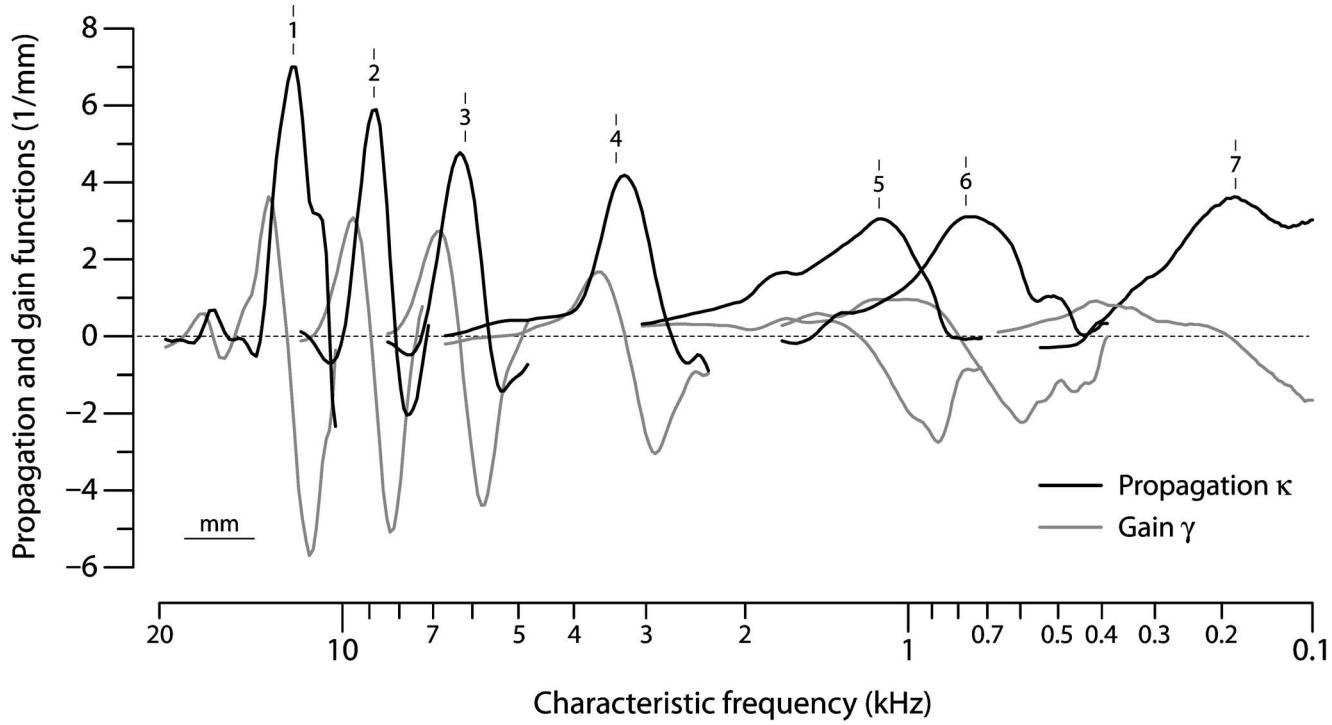


FIG. 8. Propagation and gain trends throughout the cochlea. The figure shows trend functions $\kappa(x,f_n)$ and $\gamma(x,f_n)$ at seven frequencies f_n spanning the frequency range of the chinchilla cochlea. Trend functions were computed as described for Fig. 7 by binning the 86 reconstructed Wiener kernels into seven CF groups with bin edges of $\{16,10,8,5,2,1,0.5,0.1\}$ kHz. The results are plotted on a spatial axis converted to equivalent CF using an estimate of the chinchilla cochlear map (Eldredge *et al.*, 1981; Greenwood, 1990). The nth pair of functions is plotted at the location where $\mathrm{CF}(x_n)=f_n$, where f_n is the geometric mean CF of the group. Short vertical lines, indexed by integers $(n=1,2\ldots,7)$ for future reference, identify the nominal frequency f_n of each group. The trend values at the 9-kHz site $(n=2)$ appeared earlier in Fig. 7. The scale bar represents a distance of 1 mm. For reference, thin dashes mark the zero line.

more generally. Figure 8 illustrates traveling-wave propagation and gain functions derived from Wiener-kernel responses throughout the chinchilla cochlea. The figure shows trend values of $\kappa(x,f_n)$ and $\gamma(x,f_n)$ on a spatial axis expressed in equivalent CF using the cochlear map; functions are shown at seven nominal frequencies $f_{n=1,2,\ldots,7}$ spanning the frequency range of chinchilla hearing. The functions were computed by binning the propagation and gain functions from the 86 successfully reconstructed Wiener kernels into seven CF groups, sliding each to align the CFs using the generalized scaling variable $\nu(x,f_n)$, and performing loess fits to each aligned ensemble. For example, the trend values shown above in Fig. 7 are reproduced at $\mathrm{CF}(x_2)=9$ kHz.

To facilitate comparisons across CF, Fig. 9 collapses the trend functions shown spread along the cochlea in Fig. 8 onto a common center by replotting them as a function of the generalized local scaling variable, ν. Although quantitative details vary (see below), propagation and gain functions evidently share the same qualitative form throughout the cochlea. For example, gain functions at locations from base to apex all manifest extended regions of power amplification and absorption located on opposite sides of the response peak.

We found similar results in other species by inverting tuning measurements obtained using different techniques. For example, propagation and gain functions derived from published estimates of $V_{\mathrm{BM}}(x_o,f)$ in cat obtained from auditory-nerve responses to zwuis stimuli (van der Heijden and Joris, 2003, 2006) have the same qualitative form as those derived from the chinchilla Wiener kernels. Figure 10 illustrates these remarks by showing the functions $\kappa(\nu)$ and $\gamma(\nu)$ obtained from the apex of the cat cochlea (CF <1 kHz); similar functions are obtained from more basal locations (not shown). Evidently, propagation and gain functions in the cat cochlea share the same basic form as those in the chinchilla (cf. Figs. 7 and 8). Likewise in the gerbil, a species for which we obtained similar results by applying the wave number inversion procedure both to direct mechanical measurements of the traveling wave (Ren, 2002) and to frequency-domain transfer functions (de La Rochefoucauld and Olson, 2007). In the latter case, the recovered gain functions lacked the prominent region of positive gain basal to the wave peak, consistent with the presumed passive status of the preparation.

C. Variation along the cochlea

Although the derived propagation and gain functions have the same qualitative form throughout the cochlea, quantitative details vary systematically with characteristic frequency.

1. Near-peak and maximum values

The trend functions shown in Figs. 8 and 9 suggest that peak values of the propagation function are largest in the base of the cochlea and decrease at locations further from the stapes. Figure 11 corroborates this trend with a scatterplot showing values of $\hat{\kappa}\equiv\kappa(\nu=1)$ for all 86 successfully derived

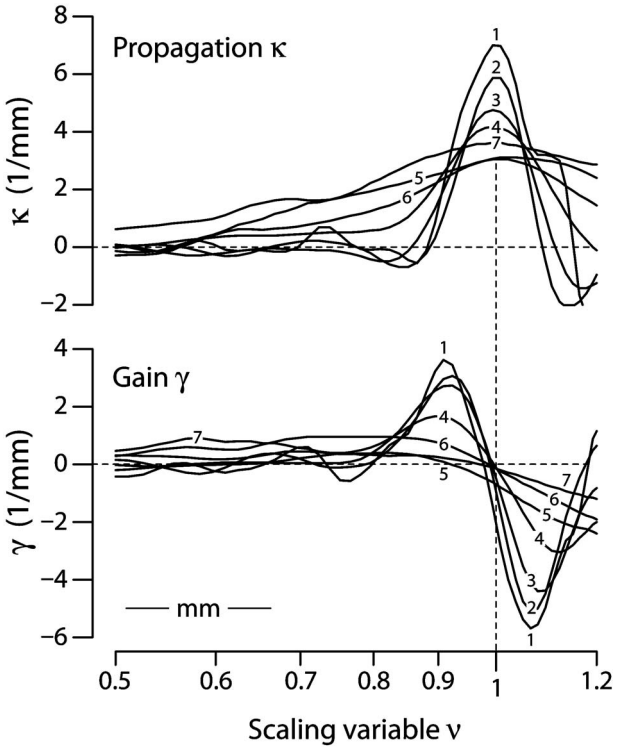


FIG. 9. Superposed propagation and gain functions from throughout the cochlea. The figure shows the trend functions $\kappa(x,f_o)$ and $\gamma(x,f_o)$ from Fig. 8 superposed by plotting them versus the generalized scaling variable, $\nu \equiv [f+\mathrm{CF}_1]/[\mathrm{CF}(x_n)+\mathrm{CF}_1]$. The integer labels ($n=1,2,\ldots,7$) correspond to those used in Fig. 8. The horizontal ν axis differs significantly from $\beta = f/\mathrm{CF}(x_n)$ only for curve $n=7$ from the extreme apex. The scale bar represents a distance of 1 mm. For reference, thin dashes mark the zero lines and the location of the wave peak ($\nu=1$).

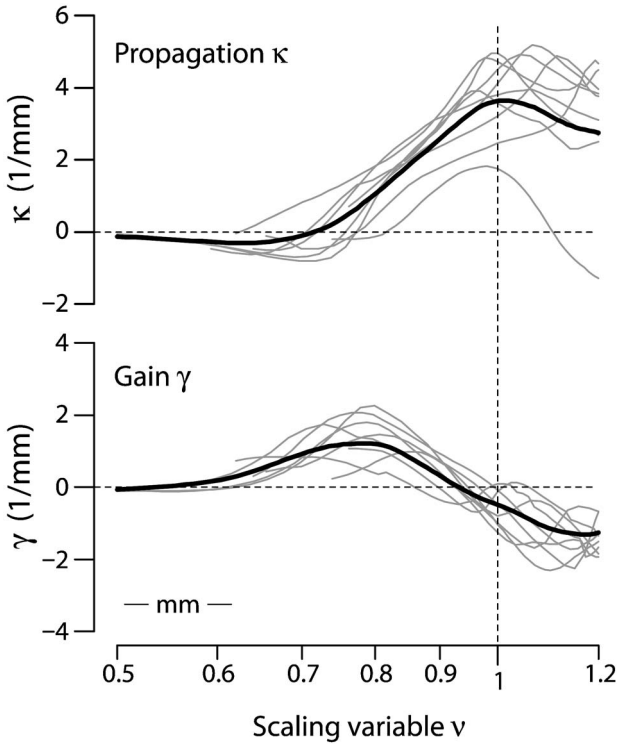


FIG. 10. Propagation and gain functions derived from estimates of $V_{\mathrm{BM}}(x_o,f)$ from the apex of the cat cochlea (CFs less than 1 kHz) obtained by van der Heijden and Joris (2006). Gray lines show individual functions $\kappa(\nu)$ and $\gamma(\nu)$; black lines show corresponding trends obtained by loess fitting. The scale bar represents a distance of 1 mm. For reference, thin dashes mark the zero lines and the location of the wave peak ($\nu=1$).

propagation functions in chinchilla. The decrease in $\hat{\kappa}$ implies that the wavelength at the peak ($\hat{\lambda} \equiv 2\pi/\hat{\kappa}$) increases by a factor of 3–4 from base to apex, from a minimum near the stapes of about 0.5–0.6 mm (i.e., 50–60 rows of hair cells) to a maximum of about 1.6–2 mm (160–200 rows) in the apex. These values of the near-peak wavelength correspond well with estimates obtained from mechanical phase data measured at nearby sites along the BM, at least at the handful of locations where such measurements have proved possible (Robles and Ruggero, 2001, Table IV). In addition to corroborating the overall trend apparent in Figs. 8 and 9, the scatterplot in Fig. 11 provides an indication of the substantial variability in $\hat{\kappa}$ associated with individual Wiener kernels.

Since the data are relatively sparse below 1 kHz, apical trends should be viewed with caution. (Because the cochlear map is more linear in the apex, the apical end also occupies a somewhat disproportionate space on the logarithmic axes of the graph.) Contributing to the uncertainty are doubts about the validity of generalized local scaling and the reliability of the cochlear map, both of which are not as well characterized in the apex. The dotted lines in Fig. 11 provide a measure of these uncertainties by showing how overall trends in $\hat{\kappa}$ change when the propagation and gain functions are derived by using traditional rather than generalized scaling (i.e., by using β rather than ν as the independent variable). The use of traditional scaling converts a shallow apical upturn in $\hat{\kappa}$ into a monotonic decline throughout the cochlea.

Notwithstanding these uncertainties, the apical upturn predicted by generalized scaling appears consistent with the behavior of the near-peak wavelength obtained from spatial phase patterns reconstructed from neural measurements in cat (van der Heijden and Joris, 2006).

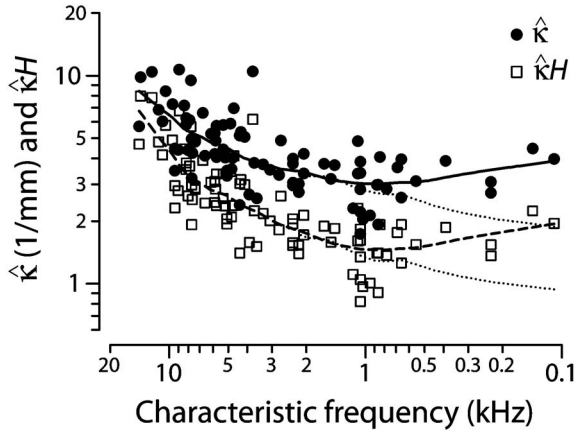


FIG. 11. Scatterplots of $\hat{\kappa}$ and $\hat{\kappa}H$ from throughout the chinchilla cochlea. The black dots give values of $\hat{\kappa}$, defined as the value of the propagation function at the wave peak ($\nu=1$), vs characteristic frequency. Values are shown for all 86 successfully reconstructed Wiener kernels. The open squares give values of $\hat{\kappa}H$, where H is the effective height of the scalae, defined as the radius of the equivalent circle (area equal to the combined areas of the scala vestibuli and tympani) as computed from measured scalae dimensions (Salt, 2001). Values $\hat{\kappa}H > 1$ imply that the hydrodynamics at the peak is short wave. The solid and dashed lines are loess trend lines (Cleveland, 1993) superposed to guide the eye. The dotted lines show trends computed from propagation functions derived using traditional rather than generalized local scaling.

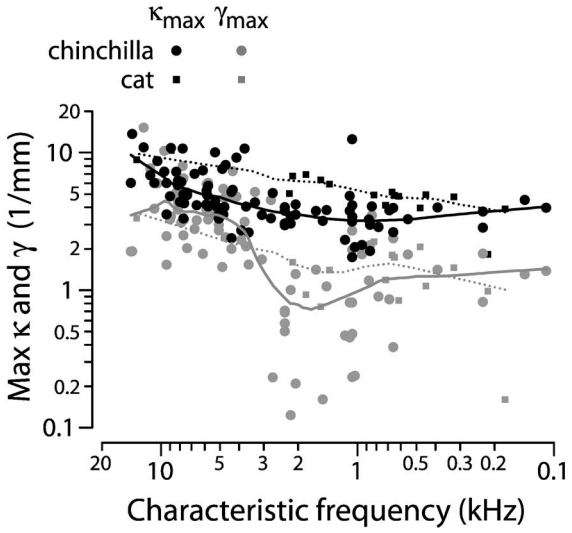


FIG. 12. Scatterplot of the maximum values of κ and γ versus CF. Black symbols show values of κ_{max} in chinchilla (circles) and cat (squares); gray symbols show corresponding values of γ_{max}. Loess trends for chinchilla and cat are shown with solid and dotted lines, respectively.

Figure 12 shows the variation of the maximum values of the propagation and gain functions. The points represent values of $\kappa_{max} \equiv \max(\kappa)$ and $\gamma_{max} \equiv \max(\gamma)$ versus CF for both chinchilla and cat. Since maximum values of κ occur at locations near the wave peak (see Fig. 13), values of κ_{max} resemble those of $\hat{\kappa}$ from Fig. 11. Trends for the maximum gain, γ_{max}, generally parallel those for κ_{max} in both the base and the apex. Note, however, that the ratio of the two makes a relatively abrupt transition at CFs near 3 kHz. (An apical-basal transition in the ratio $\gamma_{max}/\kappa_{max}$ is less apparent in the cat, although the data here are rather more sparse.) Because of this transition, the ratio $\gamma_{max}/\kappa_{max}$ is significantly larger in the base than in the apex.

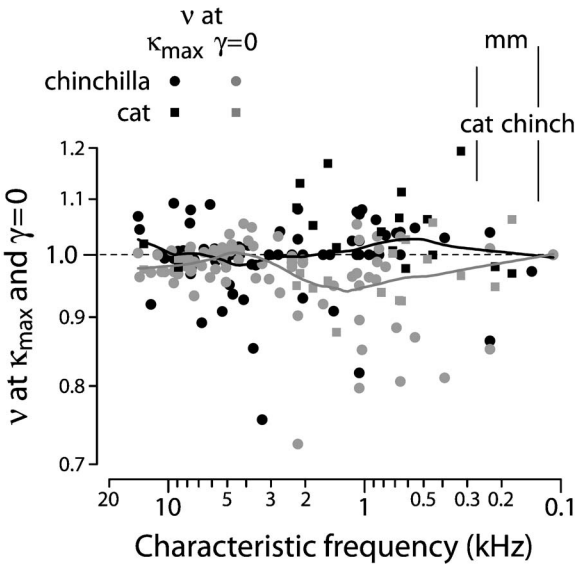


FIG. 13. Scatterplot of the maxima of $\kappa(\nu)$ and zeros of $\gamma(\nu)$ vs CF. Black symbols represent values of $\nu_{\kappa_{max}}$ in chinchilla (circles) and cat (squares); gray symbols show values of $\nu_{\gamma=0}$. Solid lines show corresponding trends computed from the data pooled across species. For reference, a dashed line marks the wave peak ($\nu=1$). Scale bars of 1 mm are shown for use when interpreting the ordinate spatially.

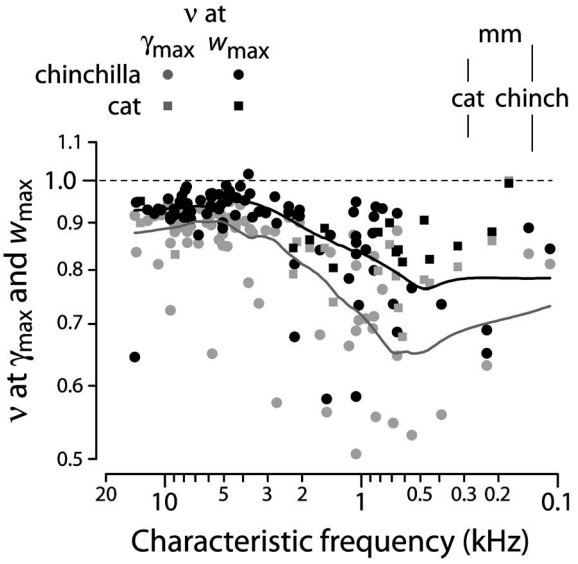


FIG. 14. Scatterplot of the maxima of $\gamma(\nu)$ and $w(\nu)$ vs CF. Gray symbols represent values of $\nu_{\gamma_{max}}$ in chinchilla (circles) and cat (squares); black symbols show values of $\nu_{w_{max}}$. Solid lines show corresponding trends computed from the data pooled across species. For reference, a dashed line marks the wave peak ($\nu=1$). Scale bars of 1 mm are shown for use when interpreting the ordinate spatially.

2. Locations of functional landmarks

Apical-basal differences are also evident in Figs. 13 and 14, which show scatterplots of the locations of functional landmarks of the propagation and gain functions. Figure 13 plots values of the scaling variable ν at locations where the propagation function $\kappa(\nu)$ reaches its maximum and where $\gamma(\nu)$ passes through 0 (i.e., where the power gain reverses sign). Both of these locations (denoted $\nu_{\kappa_{max}}$ and $\nu_{\gamma=0}$, respectively) reside close to the wave peak ($\nu=1$) throughout the cochlea. Although $\nu_{\kappa_{max}}$ and $\nu_{\gamma=0}$ appear nearly coincident in the base, the trend lines suggest that they separate slightly in the apex. In particular, the sign reversal in the gain function appears to occur somewhat basal to the peak ($\nu_{\gamma=0} < 1$) at CFs below about 3 kHz.

Figure 14 shows the locations of other features relevant to wave amplification. The gray symbols and trend line locate maximal values of the gain function γ, which occur everywhere basal to the wave peak. Whereas at high CFs the locations $\nu_{\gamma_{max}}$ cluster about 0.9, at low CFs maximum values of γ occur further from the wave peak, at values averaging about 0.7. This decrease in the value of $\nu_{\gamma_{max}}$ reflects the apical broadening of κ and γ evident in Figs. 8 and 9. Once again, the transition between apical and basal behavior occurs at CFs near 3 kHz. More functionally relevant than γ itself is the actual power transfer to the wave. The black symbols locate maximal values of $w(\nu)$, the per area power transfer given by Eq. (14). The trend indicates that maximal power transfer occurs closer to the wave peak than the maxima of γ. The apical-basal transition in the spatial extent of the gain region remains clearly apparent.

Although the derived gain functions provide estimates of the locations where the power gain peaks (namely, $\nu_{w_{max}}$ or $\nu_{\gamma_{max}}$ from Fig. 14) and estimates of where the gain region ends (namely, $\nu_{\gamma=0}$ from Fig. 13), the gain functions do not

provide especially reliable estimates of where the gain region actually begins. Although γ must first become positive somewhere basal to $\nu_{\gamma_{max}}$, and likely outside the peak region of the traveling wave, the dynamic range of the Wiener-kernel estimates of $V_{BM}(x_o,f)$ is too small to pinpoint the location. The higher dynamic range of the zwuis estimates renders them somewhat more informative in this regard, although the inversion procedure itself appears least accurate in the tail region of the response, where κ and γ are both small. These limitations notwithstanding, the gain functions shown in Fig. 10 indicate that in the apex the gain region extends for at least several millimeters basal to the peak.

Note that the data in Fig. 14 and elsewhere appear more variable below the apical-basal transition frequency than they do above. At least in part, the increased scatter reflects the larger spatial uncertainties associated with the broader propagation and gain functions characteristic of the apex (i.e., since broader functions are inherently less localized, their peak positions are more susceptible to noise).

3. Phase shifts and integrated power

The general decline in peak values of κ_{max} and γ_{max} observed during the progression from base to apex is accompanied by an increase in the width of the propagation and gain functions. This broadening is reflected in the basalward spread of the gain-function maximum at lower CFs ($\nu_{\gamma_{max}}$ in Fig. 14). In addition, reference to the Wiener-kernel transfer functions shown in Fig. 6 shows that the phase change across the peak is roughly constant in the base of the cochlea but increases in the apex. Approximate constancy in the base requires that the height and width of the peak in $\kappa(\nu)$ vary inversely in such a way that the area under the propagation function (i.e., the total phase change) remains roughly constant. At CFs below about 3 kHz, the propagation function broadens more rapidly than its height decreases (compare curves 4 and 5 in Fig. 9), and the phase shift across the (broader) peak region is therefore larger.

The gain functions show a similar pattern of variation along the cochlea, broadening while decreasing in peak amplitude in the base and then transitioning to a different behavior in the apex. To explore how these systematic changes in γ and κ affect the net power supplied to the wave, we integrate the power transfer per unit area, $w(x,f)$, along the cochlea, from the base to the point at which the power transfer reverses sign near the wave peak. More specifically, we compute[13]

$$\mathcal{P}_{\uparrow} = b\int_0^{x_0} w(x,f)dx, \tag{15}$$

where b is the effective BM width[14] and $x_0 \equiv x|_{\gamma=0}$ locates the downward zero-crossing of the gain function γ. The arrow on $\mathcal{P}_{\uparrow}$ indicates the direction of positive power transfer when the wave is pictured riding along on "top" of the organ of Corti—thus, $\mathcal{P}_{\uparrow}>0$ implies power transfer from the organ of Corti down below to the wave up above. Before plotting, we divide the computed power by $\omega^3|\hat{D}_{BM}|^2=\omega|\hat{V}_{BM}|^2$, where $|\hat{D}_{BM}|=|\hat{V}_{BM}|/\omega$ is peak BM displacement. Normalizing in this way removes changes in net power due simply to overall wave amplitude and frequency. We are asking, in effect, how the net power gain needed to produce the measured response would vary at different cochlear locations if the waves all had the same peak displacement and frequency. Although removing the dependence on peak displacement is mandated by the ANF-based data sets, which do not determine the absolute magnitude of V_{BM}, the normalization used here has the advantage of isolating only those contributions to the power gain that depend on the propagation and gain functions. (If desired, the effects of changes in wave frequency can be recovered by introducing an overall 18 dB/octave tilt.)

Figure 15 shows a scatterplot of the relative normalized power supplied to the wave ($\mathcal{P}_{\uparrow}/\omega^3|\hat{D}_{BM}|^2$) versus CF. The units along the ordinate are effectively dimensionless; they were chosen so that the trend approaches one at the highest CFs. The figure shows that the integrated contributions of κ and γ to the power gain of the cochlear amplifier are actually largest in the *apical* half of the cochlea and decrease systematically toward the base. The apical expansion of the spatial extent of the gain region evidently more than compensates for concomitant declines in γ_{max}.

D. Interdependence of κ and γ

Figure 16 demonstrates that the characteristic features of the propagation and gain functions described above are not independent of one another. In particular, the figure shows that $\kappa(\nu)$ and $\gamma(\nu)$ obey the so-called "Kramers–Krönig" dispersion relations (e.g., Kramers, 1927; Krönig, 1926; Zweig, 1976; Koshigoe and Tubis, 1982). Kramers–Krönig relations are nonlocal, integral equations that connect the real and imaginary parts of indices of refraction, scattering amplitudes, and other causal functions (e.g., Bode, 1945).[15] The dispersion relations for $\kappa(\nu)$ and $\gamma(\nu)$ have the form[16]

$$\kappa(\nu) = -\frac{2}{\pi}P\int_0^{\infty}\frac{\mu\gamma(\mu)}{\mu^2-\nu^2}d\mu, \tag{16}$$

and

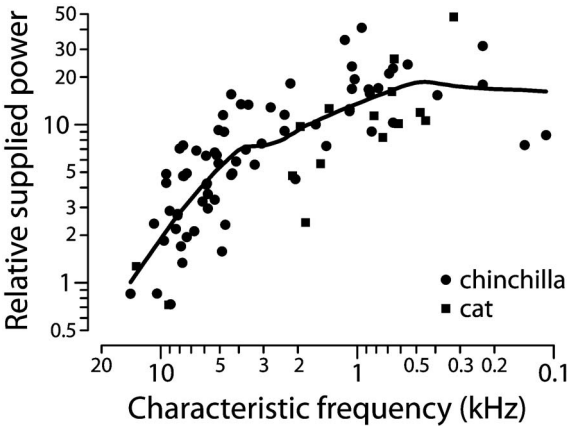


FIG. 15. Scatterplot of the relative power supplied to the wave ($\mathcal{P}_{\uparrow}/\omega^3|\hat{D}_{BM}|^2$) vs CF in chinchilla (circles) and cat (squares). Net power gains were computed using Eq. (15) and are shown in arbitrary units normalized so that the trend (solid line) is unity at the highest CF. The trend was computed from the pooled data.

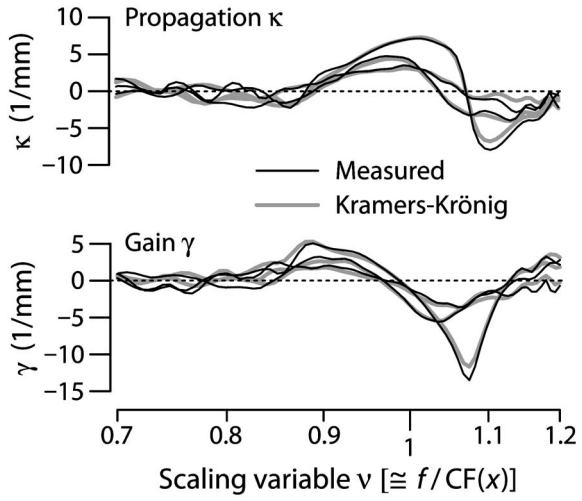


FIG. 16. Cochlear propagation and gain functions obey Kramers–Krönig dispersion relations. The figure shows empirical propagation and gains functions $\kappa(\nu)$ and $\gamma(\nu)$ derived from three example Wiener kernels with CFs near 9 kHz (thin black lines) along with their Kramers–Krönig counterparts computed one from the other (thick gray lines). The Kramers–Krönig $\kappa(\nu)$ in the top panel was computed from the empirical $\gamma(\nu)$ in the bottom panel using Eq. (16). The Kramers–Krönig $\gamma(\nu)$ in the bottom panel was computed from the empirical $\kappa(\nu)$ in the top panel using Eq. (17).

$$\gamma(\nu) = +\frac{2}{\pi} P \int_0^\infty \frac{\nu\kappa(\mu)}{\mu^2 - \nu^2} d\mu, \tag{17}$$

where the P before the integral denotes its Cauchy principal value (e.g., Mathews and Walker, 1964).[17] The figure shows derived propagation and gain functions for three example Wiener kernels along with corresponding functions obtained, one from the other, by evaluating Eqs. (16) and (17). The agreement is generally excellent in the peak region, where the propagation and gain functions are well determined. Matches of comparable or better quality, as assessed by computing the fractional rms error, were obtained for about 75% of the $\kappa(\nu)$ and $\gamma(\nu)$ pairs derived from the Wiener-kernel and zwuis responses (no systematic patterns were evident among the remainder). Evidently, the propagation and gain functions are mutually constrained by the Kramers–Krönig relations: The empirical form of $\kappa(\nu)$ determines that of $\gamma(\nu)$, and vice versa.

IV. DISCUSSION

A. Synopsis

Propagation and gain functions characterize the form and amplification of cochlear traveling waves. The properties of these functions, and their variation along the length of the cochlea, are central to an understanding of cochlear mechanics. We have presented a method for deriving propagation and gain functions from measurements of BM mechanical transfer functions. By applying the method to indirect estimates of near-threshold BM velocity obtained from: (1) Wiener-kernel analysis of chinchilla auditory-nerve responses to noise (Recio-Spinoso *et al.*, 2005) and (2) zwuis analysis of cat auditory-nerve responses to complex tones (van der Heijden and Joris, 2003, 2006), we derived propagation and gain functions throughout the cochlea in preparations uncompromised by surgical access to the scalae. In both species, and at all locations examined, the gain functions reveal a region of positive power gain basal to the wave peak. These results establish the existence of traveling-wave amplification throughout the cochlea, including the apex, where definitive conclusions have otherwise proved elusive.

Although qualitatively similar in form throughout the cochlea, the propagation and gain functions do manifest important quantitative variations with characteristic frequency. The catalog of basal to apical variations includes a systematic decline in the peak values of both κ and γ, as well as a concomitant increase in the spatial extent of the region of positive power gain. Unexpectedly, the integrated power gain attributable to the form of κ and γ appears largest in the apex rather than the base. The matching trends found in the behavior of the propagation and gain functions follow from the demonstration that the two functions are mutually constrained by the Kramers–Krönig dispersion relations.

Several properties of κ and γ undergo a relatively abrupt basal-apical transition at CFs of 3–4 kHz. (The data are more numerous and convincing in chinchilla, but a similar transition at similar CFs can also be discerned in the cat.) Interestingly, the CF location of the transition corresponds approximately with the frequency at which cat ANF tuning curves begin changing from the classic tip/tail form characteristic of high-CF fibers to the more complex, multilobed shapes found in the apex (Liberman, 1978; Liberman and Kiang, 1978). This CF also roughly corresponds to the location of a prominent bend in the frequency dependence of stimulus-frequency-emission delay in cat, guinea pig, and chinchilla (e.g., Shera and Guinan, 2003; Siegel *et al.*, 2005; Shera *et al.*, 2007). The origin and functional significance of these apparently correlated transitions between basal and apical-like behavior deserve further study.

B. Methodological and interpretive issues

1. Generality of the wave number

The model of cochlear hydrodynamics that underlies the wave number inversion procedure applies for short-, long-, and intermediate-wavelength traveling waves (Shera *et al.*, 2005). This generality plays an important role in the analysis, for although the long-wave approximation appears valid in the basal-most "tail" region of the cochlear response (e.g., Nedzelnitsky, 1980), the approximation breaks down near the peak, where the short-wave model is more appropriate. Validity of the long-wave approximation requires that the wavelength, λ, be large compared to the dimensions of the scalae. A rough guideline specifies that $2\pi H/\lambda < 1$, where H is the height (or radius) of the cochlear duct (e.g., Lighthill, 1981; de Boer, 1996). Figure 11 shows values of $\kappa H \equiv 2\pi H/\lambda$ for the chinchilla evaluated at the peak of the traveling wave using the derived propagation functions. The scatterplot demonstrates that except perhaps at the extreme apex of the cochlea, the hydrodynamics near the peak are everywhere short wave (see also van der Heijden and Joris, 2006).

Although the model supports both long and short waves, and waves of all lengths in between, the wave number inver-

sion formula derived in Sec. II B [Eq. (9)] implies that the propagation and gain functions are independent of the effective height (or width) of the scalae. In fact, they depend only on the given BM response function, $V_{BM}(x,f)$. In other words, any given mechanical response $V_{BM}(x,f)$ has unique propagation and gain functions irrespective of whether the hydrodynamics are long wave, short wave, or manifest a transition between the two, as in the real cochlea.

2. Consistency with inverse solutions for the BM impedance

The uniqueness of the propagation and gain functions that characterize a given BM mechanical response differs significantly from the situation surrounding the BM impedance and inversion procedures that can be used to find it. In contrast to the derived propagation and gain functions, the BM impedance necessary to produce a given $V_{BM}(x,f)$ depends strongly on the assumed nature of the hydrodynamics (Zweig, 1991; de Boer, 1995a; de Boer and Nuttall, 1999; Shera *et al.*, 2005). For example, although the BM impedances obtained using inversion procedures in long- and short-wave models have qualitative similarities (e.g., regions of negative resistance when employed on data from uncompromised preparations), they differ in important quantitative details, such as the locations of their "resonant" frequencies (zeroes of the reactance) and the strengths of the active forces.

To make these quantitative differences more explicit, we note that in short-wave models the BM impedance depends on the propagation and gain function through the relation

$$Z_{BM} = -(\gamma + i\kappa)|Z_f H/k^2| \quad \text{(short wave)}, \tag{18}$$

where Z_f is the effective acoustic impedance of the fluids [Eq. (7)]. In long-wave models, by contrast, the BM impedance has the value

$$Z_{BM} = -(2\gamma\kappa + i[\kappa^2 - \gamma^2])|Z_f/k^4| \quad \text{(long wave)}. \tag{19}$$

As these equations indicate, the real parts of both short- and long-wave BM impedances share a region of negative resistance where the gain function γ is positive (i.e., basal to the characteristic place) but have rather different magnitudes. As a result, the strength of the active forces necessary to produce a given BM velocity response differs considerably between the two models. The reactive components of the two BM impedances are even more dissimilar: Whereas the short-wave reactance remains stiffness dominated throughout the peak region (where $\kappa > 0$), the long-wave reactance has a zero (resonant frequency) at the point where $\kappa = -\gamma$, a condition satisfied at a location just apical to the response peak. These conclusions, deduced from Eqs. (18) and (19) and the qualitative forms of κ and γ described above, are consistent with inverse solutions for the BM impedance performed in 1D and 3D cochlear models (e.g., Zweig, 1991; de Boer, 1995a; de Boer and Nuttall, 1999).

3. Behavior beyond the best place

After decreasing from a maximum near the best place, the derived propagation functions often pass through zero and become negative at more apical locations (cf. Fig. 7). Simply interpreted, negative values of κ suggest that waves are traveling backward toward the stapes. However, the decrease in κ and the appearance of negative values in this region actually reflect the transition from propagating to evanescent (nonpropagating) wave behavior characteristic of the "cutoff" region beyond the wave peak. Wave propagation requires the dynamical interplay of the kinetic energy of fluid motion and the potential energy of partition stiffness. In the cutoff region, however, the reactive component of the BM impedance is controlled not by the partition stiffness, but by its mass (e.g., de La Rochefoucauld and Olson, 2007). Note, for example, that negative values of κ yield a positive (i.e., mass-dominated) BM reactance when substituted into the short-wave expression for Z_{BM} [Eq. (18)]. As a result, wave propagation cannot occur in this region, and the straightforward interpretation of the sign of κ no longer applies.

When combined with local scaling, the decline in the propagation function apical to the best place predicts a corresponding decrease in the wave group delay. The group delay τ_{grp} is defined by

$$\tau_{grp} \equiv -\frac{1}{2\pi}\frac{\partial\varphi}{\partial f} \cong -\frac{1}{2\pi}\frac{\partial}{\partial f}\int_0^x \kappa(x',f)dx', \tag{20}$$

where $\varphi(x,f)$ is the wave phase, and the approximation neglects small contributions from the $k^{3/2}$ prefactor in Eq. (11). Local scaling implies that the spatial integral and its frequency derivative can both be reexpressed in terms of ν. In particular, $dx = (\partial x/\partial\nu)d\nu$ and $\partial/\partial f = (\partial\nu/\partial f)\partial/\partial\nu$, where the necessary partials can be computed from Eq. (5). Evaluating Eq. (20) then yields

$$\tau_{grp}(\nu,f) \cong \frac{l}{2\pi(f + CF_1)}\kappa(\nu), \tag{21}$$

where l and CF_1 are parameters of the cochlear map [Eq. (4)]. At fixed frequency, Eq. (21) predicts that the wave group delay manifests the same qualitative behavior as $\kappa(\nu)$. Specifically, the analysis predicts that measured group delays increase to a maximum near the best place and decrease beyond. This paradoxical (because seemingly acausal) decrease in the group delay at more distant locations results from the transition to evanescent behavior.[18]

4. Local scaling and other approximations

Although the inversion procedure does not itself rely on local scaling (Sec. II B), the transformation applied here to convert measured frequency-domain transfer functions (e.g., Wiener kernels) into the traveling waves needed as raw material for the inversion does assume this approximate symmetry (Sec. II A). Although significant deviations from scaling become apparent when cochlear responses are compared across one or two octaves of CF (Fig. 6; see also Shera and Guinan, 2003), the conversion used here requires only that the symmetry apply over distances corresponding to the width of the transfer function peak. (The dynamic range of the Wiener kernels is generally insufficient to provide useful information outside the peak region. Fortunately for our purposes, most of the amplification and dispersion occurs near

the peak.) Except perhaps at the extreme apex of the cochlea, where transfer functions can span several octaves, the local scaling approximation appears well satisfied.

Direct measurements of the traveling wave performed using scanning interferometry (Ren, 2002) corroborate the validity of local scaling in the base of the cochlea. To wit, Ren's gerbil traveling waves and transfer functions, although measured in different preparations, nearly overlie one another when plotted using the scaling variables β or ν. As an additional check on the validity of local scaling, we applied the inversion procedure directly to Ren's traveling waves, for which no scaling assumptions are needed, and obtained propagation and gain functions with the same qualitative features that emerge from the analysis of the Wiener-kernel and zwuis transfer functions in the chinchilla and cat. In addition, the inversion procedure yields values of κ that agree well with those obtained by direct spatial computation of the real part of the wave number in gerbil (de La Rochefoucauld and Olson, 2007).

In the apex, van der Heijden and Joris (2006) confirm the violation of traditional scaling expected from changes in the shapes of cat neural tuning curves (e.g., Kiang and Moxon, 1974; Liberman, 1978). Deviations from traditional scaling can be substantially reduced, although not eliminated, by means of the generalized local scaling introduced here. Although quantitative details, such as the values of $\hat{\kappa}$ shown in Fig. 11, depend on to the choice of scaling variable (ν versus β), qualitative features of the propagation and gain functions, such as the existence of power gain basal to the wave peak, are quite robust. Although we expect residual violations of generalized scaling to introduce systematic errors, it is reassuring to note that the values of κ we obtained in cat are consistent with estimates of the wavelength obtained from "panoramic" (i.e., spatial) profiles of cochlear phase constructed from the same data by van der Heijden and Joris (2006).[19] Thus, the consistency of our results with independent estimates of the wavelength from both the base and the apex suggests that our findings are unlikely to be artifacts introduced by the local scaling approximation.

Since we obtained similar results when analyzing direct mechanical measurements (e.g., Ren, 2002; de La Rochefoucauld and Olson, 2007), neither are our findings a consequence of employing Wiener-kernel or zwuis-based estimates of BM velocity responses. The Wiener-kernel and similar measurements characterize the traveling wave as seen from the auditory nerve. In principle, they therefore include contributions from internal motions of the organ of Corti that may be visible to the inner hair cell but appear less prominent in the motion of the BM (e.g., Guinan *et al.*, 2005; Nowotny and Gummer, 2006; Karavitaki and Mountain, 2007). Although subtle systematic differences will doubtless be discovered, Wiener-kernel responses corrected for synaptic and neural transmission delays closely resemble BM mechanical measurements made at corresponding locations and intensities (Temchin *et al.*, 2005).

5. When things fall apart

The major limitation of the wave number inversion procedure is that it doesn't always work when applied to real data. In a significant minority of cases the inversion produces a reconstructed $V_{\mathrm{BM}}(x,f)$ with sharp spikes or discontinuities not present in the original. [Like the little girl with the curl (Longfellow, 1922), when the reconstructions are good, they are very good indeed, but when they are bad, they are horrid.] Anomalies and other problems were present even in regions where the estimated BM response appeared well determined; they were especially common in the "cutoff" region apical to the best place (or at frequencies above CF). Tests of the inversion procedure using model-generated responses for which the actual wave number is known suggest that poor reconstructions can result from errors in the assumed boundary conditions and/or from contamination by higher-order, nonpropagating (evanescent) modes. They may also result from inaccuracies in the Wiener-kernel or zwuis estimates of BM motion. Interestingly in this regard, the overall failure rate of about 30%–40% masks a significant dependence on CF: the probability of obtaining an unsuccessful reconstruction was roughly twice as large for CFs above 3 kHz than it was below (0.45 versus 0.21). However, similar reconstruction problems were also encountered when analyzing BM mechanical measurements obtained using laser vibrometry. The invention of methods to mitigate or eliminate these problems would improve the reliability, and presumably also the accuracy, of the procedure. The trends and conclusions reported here are based on cases validated by successful reconstruction. This restriction was imposed more for logical consistency than for practical necessity; including errant wave numbers in the analysis increased the amount of scatter but did not alter the overall trends (e.g., in Figs. 11–15).

6. Estimating the power gain of the cochlear amplifier

In principle, the propagation and gain functions can be used to find the power gain of the cochlear amplifier.[20] Energy conservation requires that the total power entering the cochlea at the stapes ($\mathcal{P}_{\Rightarrow}$) equal the power absorbed within the cochlea ($\mathcal{P}_{\Downarrow}$) plus any power reemitted ($\mathcal{P}_{\Leftarrow}$). In other words,

$$\mathcal{P}_{\Rightarrow} = \mathcal{P}_{\Downarrow} + \mathcal{P}_{\Leftarrow} \cong \mathcal{P}_{\Downarrow}, \tag{22}$$

where the approximation is valid when the reemitted power (e.g., stimulus-frequency emission) is small compared to the stimulus.[21] As discussed in Sec. III C 3, the absorbed power can be computed from the integral

$$\mathcal{P}_{\Downarrow} = -b\int_0^L w(x,f)dx, \tag{23}$$

where $-w(x,f)$ is the power per unit BM area dissipated within the organ of Corti. By splitting the integral into two parts, Eq. (23) can be written in the form

$$\mathcal{P}_{\Downarrow} = -\mathcal{P}_{\uparrow} + \mathcal{P}_{\downarrow}, \tag{24}$$

where

$$\mathcal{P}_{\uparrow} \equiv b\int_0^{x_0} w\,dx \text{ and } \mathcal{P}_{\downarrow} \equiv -b\int_{x_0}^{L} w dx. \tag{25}$$

Introduced in Eq. (15), $\mathcal{P}_{\uparrow}$ represents the net power transfer

to the wave in the region basal to x_0; $\mathcal{P}_\downarrow$ is the net power absorbed *from* the wave in the region apical to x_0. Figure 13 shows that x_0 is nearly coincident with or slightly basal to the wave peak.

A physically relevant measure of the net power gain of the cochlear amplifier can now be obtained by computing the ratio

$$G \equiv \frac{\mathcal{P}_\Rightarrow + \mathcal{P}_\uparrow}{\mathcal{P}_\Rightarrow} = 1 + \frac{\mathcal{P}_\uparrow}{\mathcal{P}_\Rightarrow}. \tag{26}$$

Equivalent to the metric adopted by de Boer and Nuttall (2001), the power gain G is just the total forward power at x_0 normalized by the power input at the stapes. Since the total power is the input power plus the net power supplied to the wave, G is greater than one when the organ of Corti provides net power gain ($\mathcal{P}_\uparrow > 0$) and less than one when there is net power loss ($\mathcal{P}_\uparrow < 0$), as in a passive preparation. Note, however, that because G measures only the *net* power gain, it underestimates the *total* power supplied to the wave by the organ of Corti, some of which necessarily goes toward compensating for passive losses. Unless the total power supplied exceeds these losses, the net gain G will remain less than one.[22]

Equation (26) can be simplified by combining Eqs. (22) and (24) to yield $\mathcal{P}_\Rightarrow \cong \mathcal{P}_\downarrow - \mathcal{P}_\uparrow$. Consequently,

$$G \cong \frac{1}{1 - \mathcal{P}_\uparrow/\mathcal{P}_\downarrow}, \tag{27}$$

a quantity that can be computed from the derived propagation and gain functions. Note that the gain G is independent of any factors common to both $\mathcal{P}_\uparrow$ and $\mathcal{P}_\downarrow$ that divide out in their ratio (e.g., the BM width and the unknown overall wave amplitude, $\hat{V}_{\mathrm{bm}}$). Unfortunately, Eq. (27) also implies that accurate estimates of G are likely unattainable from experimental data in individual animals, at least when the gain is large. Large values of G occur when the ratio $r \equiv \mathcal{P}_\uparrow/\mathcal{P}_\downarrow$ is close to one, where $G(r)$ is singular. Since they depend on the small difference between two larger numbers, large values of G are therefore especially susceptible to error. Indeed, since $\Delta G = G^2 \Delta r$, uncertainties Δr are boosted by the factor G^2 and increase faster than the gain itself. Thus, as the power gain becomes large, errors in computing r appear magnified, sometimes enormously, in G.

Figure 17 shows empirical values $r \equiv \mathcal{P}_\uparrow/\mathcal{P}_\downarrow$ representing ratios of net power supplied to/delivered by the wave. Although r declines somewhat at the highest CFs, the values generally cluster near $r=1$, indicating substantial power gains throughout the cochlea. In particular, r appears large even in the apex (indeed, perhaps especially in the apex). Note that values $r \geqslant 1$ are forbidden by energy conservation if $\mathcal{P}_\Leftarrow = 0$. Empirical ratios r at values greater than 1 must thus reflect violations of the no-emission approximation[23] and/or errors due to the finite precision of the data and their analysis. Collapsing the estimates across CF yields a median r of 0.95 and a mean of 0.94 ± 0.07, with uncertainties representing approximate 95% confidence intervals. The confidence intervals for r correspond to net power gains spanning the range $G \in [7, \infty]$, with a mean of approximately 17. (When expressed in dB, computed as $10 \log G$, the corresponding gain range is $[8, \infty]$ dB, with a mean near 12 dB.) Of course, the semi-infinite uncertainty on the estimate of G provides an important caveat. Were it to prove representative of actual cochlear power gains, the value $r=0.94$ would imply that at least 94% of the wave power reaching its characteristic place originates *within* the cochlea.

Both our rough estimate of the mean power gain and its substantial uncertainty (as well as the large apparent variability in gain between individuals) are consistent with results reported by de Boer and Nuttall (2001). Based on an inverse analysis of data from 20 good preparations, they report net gains in guinea pig ranging from 0.2 to 17.7 dB computed using an equivalent metric. Our analysis shows that much if not all of the variability in estimates of the net power gain arises through the error magnification inherent in Eq. (27) for G at large gains. Although experimental uncertainties and their magnification preclude precise quantification of the power gain from the data employed here, our results do demonstrate significant gain throughout the cochlea.

C. Interpretation as a hydromechanical laser amplifier

The empirical cochlear propagation and gain functions obtained here resemble those that characterize electromagnetic wave propagation in an active optical gain medium. Indeed, our results corroborate previous suggestions (Zweig, 1991; Russell and Kössl, 1999; Kemp, 2002; Shera, 2003a) by demonstrating that the cochlea functions as a biological analog of a laser amplifier.

An optical laser amplifier consists of a "gain medium" that supports electromagnetic wave propagation, perhaps a gas or crystal, and a power supply or pumping process that puts the atoms of the medium into an excited state [see Fig. 18(A)]. When the atoms relax they spontaneously emit incoherent light—light, that is, with a mix of directions, phases, and polarizations. But when light from an external source illuminates the medium, it stimulates the atoms to radiate in phase with the input, producing a like-upon-like coalescence

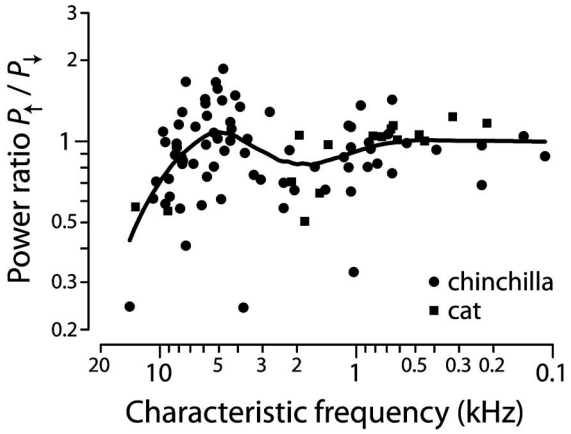


FIG. 17. Scatterplot of $r \equiv \mathcal{P}_\uparrow/\mathcal{P}_\downarrow$ vs CF in chinchilla (circles) and cat (squares). Computed using Eq. (25), r is the ratio of the net powers supplied to/delivered by the wave in regions basal/apical to the point (x_o) where the power flow reverses near the peak of the wave. The trend (solid line) was computed from the pooled data and indicates that r is fairly close to one, implying the existence of substantial power gains throughout the cochlea.

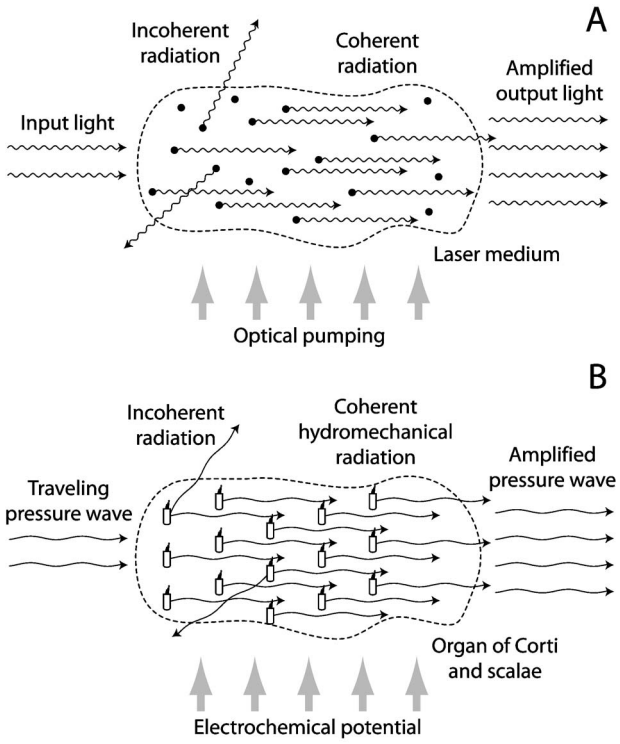


FIG. 18. Analogy between laser and cochlear amplifiers. The top panel (A) shows the essentials of a laser amplifier. When excited by an optical pumping process, atoms embedded in a gain medium that supports electromagnetic wave propagation spontaneously emit incoherent light. Light of the same frequency applied at the input stimulates the atoms to radiate in phase, amplifying the applied beam. The bottom panel (B) shows corresponding features of the cochlear amplifier. When stimulated via bundle displacements induced by the traveling pressure-difference wave, hair cells create forces that couple to their mechanical environment, producing hydromechanical radiation that combines coherently with and amplifies the incident pressure wave. Although each cell may radiate symmetrically in both directions, the backward radiation combines incoherently and tends to cancel out (Shera, 2003b; Shera and Guinan, 2007); only the forward radiation is shown here. Brownian motion also drives the individual hair bundles, resulting in incoherent hydromechanical radiation whose amplitude and phase differ from cell to cell. (For the purposes of illustration, this incoherent radiation is shown propagating obliquely.) The amplification process is powered by the electrochemical potentials that drive ionic currents through the hair-cell transduction channels.

of external and stimulated radiation. The result is an amplified output beam: The power in the signal beam is boosted by coherent amplification. The process of stimulated emission and amplification occur when the frequency of the incident light closely matches the frequency of the radiation emitted by the atoms of the gain medium, a frequency determined by atomic energy levels, in particular by the quantized energy difference between the excited and relaxed states (for an excellent review see Siegman, 1986).

The propagation and gain functions derived here indicate that when stimulated by a passing wave, the cochlear gain medium "emits" hydromechanical radiation that combines coherently with and boosts the amplitude of the incident wave ($\gamma > 0$) in a region basal to the characteristic place (see also Zweig, 1991). Although many details of the biophysical workings of the gain medium remain uncertain, the cochlear analogues of major functional elements of the optical laser are clear [see Fig. 18(B)]. In the cochlea, traveling-wave amplification is hydromechanical rather than electromagnetic and quantum mechanical. The disturbances that propagate in the medium are cochlear traveling pressure-difference waves rather than light. The gain medium responsible for coherent wave amplification consists not of excited atoms but of the cellular force generators and concomitant mechanical, hydrodynamic, and electrical processes known collectively as the "cochlear amplifier" (Neely, 1983). The power supply, or battery, is the electrochemical gradient that drives ionic currents through the hair cells, changing their receptor potentials and inducing conformational changes in transmembrane proteins. These conformational changes exert forces that are transmitted through the organ of Corti via its elaborate cytoarchitectural scaffolding. Ultimately (and herein much of the mystery lies), these forces create pressure gradients in the surrounding fluids that, according to the present analysis, combine coherently with the stimulus wave to boost the amplitude of the response.

1. An analytic twist

Unlike a conventional laser amplifier, the cochlea has functional requirements more fundamental than providing power gain. First, the cochlea must *analyze* the signal it amplifies. Although the beam directed to an optical amplifier simply passes through the medium for a boost on its way to other things, the sound input to the cochlea must be separated into frequency components and directed onto different populations of sensory cells embedded within the gain medium itself. Immediately thereafter, the wave excitation must be rapidly extinguished to minimize interference with the analysis of ongoing sounds. Second, cochlear operation must be *wideband*: Although a conventional optical amplifier operates only within the extremely narrow frequency range determined by the atomic transition, the cochlea must cope with any sound within the frequency range of hearing. Accommodating these basic requirements for wideband analysis and amplification requires a crucial modification to the propagation and gain functions found in a conventional laser medium.

In an idealized optical gain medium the wave number k_{optical} has the form

$$k_{\text{optical}} = \kappa_0 + k_{\text{laser}}, \tag{28}$$

where $\kappa_0 = \omega/c$ and c is the speed of light in the host medium. (Ohmic losses are usually small and have been neglected for simplicity.) The interesting action occurs in k_{laser}, which characterizes the atomic contribution to the wave number due to amplifying interactions within the medium. The real and imaginary parts of k_{laser} represent the laser propagation and gain functions:

$$k_{\text{laser}} = \kappa_{\text{laser}} + i\gamma_{\text{laser}}. \tag{29}$$

The top panel of Fig. 19 illustrates the qualitative behavior of κ_{laser} and γ_{laser} at frequencies in the immediate neighborhood of the atomic transition frequency, f_{atomic}. The laser gain function is positive ($\gamma_{\text{laser}} > 0$) but vanishingly small everywhere except in a narrow band near the frequency f_{atomic}, about which it manifests a symmetric Lorentzian resonance characteristic. By contrast, the laser propagation function is

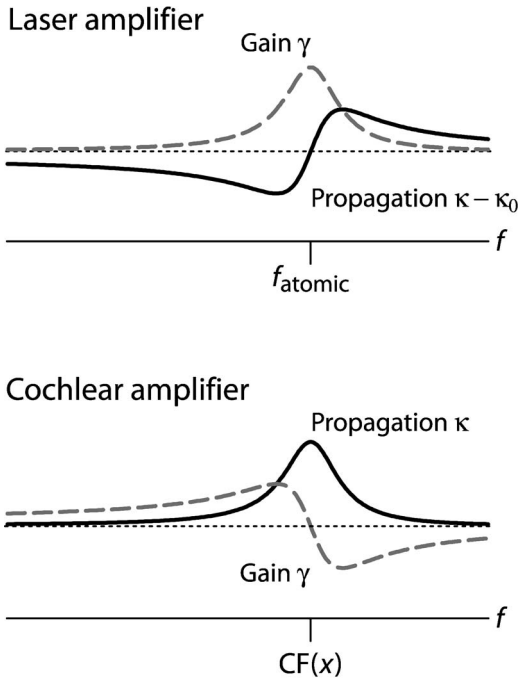


FIG. 19. Propagation and gain functions for laser and cochlear amplifiers versus frequency. The top panel shows stylized propagation and gain functions for an optical laser amplifier in the frequency region about the atomic transition frequency, f_{atomic}. Only the resonant, atomic contribution to the laser propagation function is shown; the constant wave number κ_0 corresponding to background propagation in the host medium has been subtracted off. [Adapted from Figs. 2.8 and 7.3 of Siegman (1986).] The bottom panel shows stylized versions of the empirical cochlear propagation and gain functions derived here (see Figs. 8 and 9). The cochlear wave number appears isomorphic to the laser wave number rotated 90° clockwise in the complex plane. In both panels, dotted lines mark the zero along the ordinate. Note that the laser functions are substantially narrower than those in the cochlea. Had they been drawn on the same logarithmic axis used for the cochlea, the laser functions would have bandwidths smaller than the dots that compose the zero lines. In the cochlea, local scaling can be used to convert the frequency axis into a spatial one. The clockwise twist shown here is crucial to the operation of the cochlea, which must also analyze the signals it amplifies. Useful analysis requires both level-dependent amplification (to match the variance of the incoming signal to the dynamic range of the detectors) and attenuation (to clear the stage for the analysis of future sounds). The result is a region with $\gamma>0$ followed by one with $\gamma<0$.

antisymmetric about f_{atomic}. At frequencies below the atomic transition frequency ($f<f_{\text{atomic}}$), the propagation function is negative ($\kappa_{\text{laser}}<0$), and the effective propagation coefficient ($\kappa_0+\kappa_{\text{laser}}$) is therefore smaller than κ_0. At these frequencies, amplifying interactions with the gain medium therefore increase the wavelength and velocity of the light wave through the medium; at frequencies above the transition, the same interactions decrease them. The differing behaviors of κ_{laser} and γ_{laser} are mutually interdetermined through the Kramers–Krönig dispersion relations.

The overall "shapes" of the cochlear propagation and gain functions derived here resemble those of an optical medium, but with the crucial twist that the roles played by κ_{laser} and γ_{laser} are interchanged. More precisely, they are rotated in the complex plane in a way that preserves the Kramers–Krönig relations. In particular,

$$k_{\text{cochlea}} \sim -ik_{\text{laser}}, \tag{30}$$

where the tilde signifies the overall isomorphism and the subscript "cochlea" has been added for clarity. (Any cochlear analogue of the wave number κ_0 appears small.) The $-i$ multiplying the laser wave number rotates its components clockwise through one quarter cycle, effectively switching the roles of propagation and gain functions:

$$\kappa_{\text{cochlea}} \sim \gamma_{\text{laser}} \tag{31}$$

and

$$\gamma_{\text{cochlea}} \sim -\kappa_{\text{laser}}. \tag{32}$$

For comparison with their laser counterparts illustrated in the top panel, the bottom panel of Fig. 19 shows stylized versions of the cochlear propagation and gain functions derived from experiment (cf. Figs. 8 and 9). In the cochlea, the role of the atomic transition frequency is played by CF(x), which varies with location within the medium. Although the relative scales were chosen to emphasize the isomorphism of shape, the fractional bandwidths of the curves about their center frequencies [e.g., the width of γ_{laser} about f_{atomic} and of κ_{cochlea} about CF(x)] are much smaller (narrower) for the optical laser than they are in the cochlea. In optical lasers, relative bandwidths are often considerably less than fractions of a percent;[24] in the cochlea, they are at least 10% in the base and can be significantly greater in the apex (see Fig. 9).

Because of the effective rotation of the wave number, the cochlear gain medium—unlike the laser medium—is strongly dispersive. Whereas light amplified by a laser undergoes negligible change in wavelength or velocity as it propagates through the medium [$\kappa_{\text{laser}}/\kappa_0 \ll 1$ with $\kappa_{\text{laser}}(f_{\text{atomic}}) \cong 0$], the cochlear propagation function resembles the laser gain function and depends strongly on frequency. Because of local scaling, frequency dispersion is equivalent to spatial dispersion: The wavelength and velocity of the cochlear traveling wave therefore depend strongly on position, and the wave slows considerably as it approaches its characteristic place. The cochlear gain function mirrors the antisymmetric laser propagation function. Unlike an optical laser, where maximum gain occurs at the frequency f_{atomic}, the maximal value of γ_{cochlea} occurs at frequencies below CF(x). At frequencies above CF(x) the gain function reverses sign, and waves are strongly squelched. Again, local scaling converts this frequency variation into a spatial pattern in the cochlea—traveling waves are amplified as they approach their characteristic place and strongly attenuated beyond. Thus, in contrast to an optical laser, where amplification occurs uniformly throughout the medium, amplification in the cochlea is spatially localized.

Thus, the strong dispersion and concomitant absorption introduced by rotating the laser wave number and varying CF(x) with position allow the cochlear gain medium to perform both wideband amplification and analysis. Although each section of the cochlea amplifies a limited range of frequencies—according to the cochlear gain functions shown in Fig. 9 this range extends in the chinchilla from roughly one third of an octave below CF in the base of the cochlea to as much as an octave below CF in the apex—together they amplify any frequency for which vibration of the stapes launches a traveling wave. In principle, wideband amplification—but not analysis—could be achieved using conventional laser gain functions; for example, by construct-

ing the gain medium from a mixture of atoms with a wide array of transition frequencies. But useful analysis requires both strong dispersion, to separate signals into their component frequencies, and regions of *negative* power gain, so that amplified signal energy, once transduced, will be reabsorbed within the medium rather than be allowed to linger, compromising the analysis of future sounds.

2. Emergence of laser oscillations

Our results demonstrating laser-like *amplification* within the mammalian cochlea corroborate earlier findings of laser-like *oscillations* emitted from the ear (Kemp, 1979; Shera, 2003a). An optical laser amplifier can be transformed into a laser oscillator—a device that *generates* light rather than simply amplifying it—by embedding the gain medium in a resonant cavity; for example, by adding mirrors at the boundaries to reflect light back through the medium, creating a feedback loop that recycles the output back to the input. If the size of the cavity is tuned just right, so that the round-trip distance encompasses an integral number of wavelengths of the emitted light, then standing waves can build up inside the cavity. The amplitude of these standing waves is determined by the round-trip power gains and losses. If one of the mirrors is made partially transparent, the system will emit a beam of coherent light—a laser beam.

The cochlea, too, contains mirrors. In the cochlea, the "resonant cavity" spans the region between the stapes and the peak of the traveling wave. At either end of this region, cochlear traveling waves are partially reflected back into the cavity. At the apical end of the cavity, forward-traveling waves are partially reflected (or reemitted) through the process that generates stimulus-frequency otoacoustic emissions [e.g., coherent reflection off mechanical "imperfections" in the gain medium near the peak of the traveling wave (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.*, 2005)]; at the stapes, backward-traveling waves reflect due to the impedance mismatch with the middle ear (Shera and Zweig, 1991; Puria, 2003).[25]

On each pass through the cavity, cochlear traveling waves are amplified by their interaction with the gain medium. At certain special frequencies—those for which the round-trip phase change is an integral number of cycles—multiple internal reflection creates standing waves. If the round-trip gain matches the round-trip losses (e.g., due to damping and acoustic radiation into the environment), stable oscillations can result that appear in the ear canal as spontaneous otoacoustic emissions (SOAEs). Unlike optical oscillators, the cochlea can emit at multiple, nonharmonically related frequencies. In an optical laser, the cavity size is fixed and tuned to a single frequency and its harmonics. Wave propagation in the cochlea, however, is highly dispersive and the location of the wave peak—and hence the location of the partially reflecting mirror—depends on frequency. Consequently, the round-trip phase condition is satisfied at many frequencies and the cochlea can—and does—produce multiple SOAEs simultaneously.

Just as in an optical laser, oscillation amplitudes are self-stabilizing. Since the cochlear gain medium is limited in the energy it can produce, amplification is nonlinear (amplitude dependent), decreasing as the wave amplitude grows. Standing-wave amplitudes are therefore stable against perturbations: If some random fluctuation increases the wave amplitude slightly, the round-trip gain decreases a little and the wave amplitude falls back down. Conversely, if the wave amplitude decreases, the total gain increases, pulling the oscillation amplitude back up. Amplitude stabilization of this sort produces acoustic SOAE signals with statistical properties identical to those that characterize the coherent electromagnetic radiation generated by optical lasers (e.g., Golay, 1961; Siegman, 1986; Shera, 2003a).

3. Meaning of the dispersion relations

In optical lasers and other physical systems, adherence to Kramers–Krönig dispersion relations is generally a consequence of causality (e.g., Toll, 1956). Causality requires that a system not respond before it is driven. Although causality is thought to constrain all physical interactions, its application to the propagation and gain functions derived here arises not as the universal and inevitable consequence of fundamental physics but reflects the detailed and contingent dynamics of the cochlea. According to Eqs. (6) and (7), k^2 can be written in the form

$$k^2 = -bZ_f V_{BM}/\bar{P}. \tag{33}$$

In this equation, k^2 plays a role analogous to an admittance, characterizing the system's response (V_{BM}) to an impressed force ($\bar{P}$). Thus, the approximate empirical adherence to the Kramers–Krönig relations demonstrated in Fig. 16 reflects the approximate causality of the BM response to $\bar{P}$.

The approximate causality implied by the Kramers–Krönig relations is dynamically significant because $\bar{P}$ is not necessarily the only force acting upon the BM. (In fact, because $\bar{P}$ is the pressure difference averaged in the "vertical" direction above the BM, the pressure $\bar{P}$ is not physically driving the BM at all.) In principle, the motion of the BM could—and at some level must—be driven by forces communicated from nearby locations by means other than scalae-fluid pressure. Possible alternate routes include "feed-forward" mechanisms such as the longitudinal tilt of the outer hair cells (e.g., Kolston *et al.*, 1989; Steele *et al.*, 1993; Geisler and Sang, 1995), fluid flow in the tunnel of Corti (Karavitaki and Mountain, 2007), or subtectorial space (Nowotny and Gummer, 2006), mechanical and electrical coupling within the organ of Corti (reviewed in Santos-Sacchi, 2000) or along the tectorial membrane (Ghaffari *et al.*, 2007), as well as neural interactions among the outer hair cells (e.g., Thiers *et al.*, 2002). When timed appropriately, "nonclassical" forces such as these can give rise to both noncausal point-impedance functions (de Boer, 1997a) and noncausal values of k^2. The fact, however, that the empirical adherence to the Kramers–Krönig relations appears as good as it does indicates that any noncausal contributions to BM propagation and amplification (i.e., forces fed forward relative to $\bar{P}$) must be relatively small, at least in their effect on the wave peak.

4. Interrelationship of propagation and gain

The propagation and gain functions determined here describe qualitatively different phenomena in the cochlea. Whereas $\kappa(x,f)$ characterizes traveling-wave dispersion, $\gamma(x,f)$ describes power amplification and absorption. Direct measurement of one would require equipment and analysis ill suited to measurement of the other. Despite these *qualitative* differences, cochlear propagation and gain functions appear *quantitatively* related: The Kramers–Krönig relations satisfied by $\kappa(x,f)$ and $\gamma(x,f)$ [Eqs. (16) and (17)] imply that the form of one determines the form of the other. Thus, the properties of cochlear amplification (e.g., gains and spatial distributions) cannot be changed without simultaneously changing the characteristics of wave propagation (e.g., wavelengths, velocities, and delays), and vice versa (see also Zweig, 1976; de Boer, 1997b). In the cochlea, propagation and gain are inseparably complected aspects of the same process: In a deep way, made precise by Eqs. (16) and (17), the wavelength of the traveling wave *is* the amplifier.

5. Biophysics of the gain medium

By stripping away uncertainty about the biophysical underpinnings to reveal the basic functional operation of the cochlea—a wideband, hydromechanical laser analyzer—the present analysis emphasizes a point often obscured by molecular detail: Cochlear traveling-wave amplification is distributed and involves coupling among elements over a relatively broad region of the cochlea. According to the results from Fig. 8, this region is at least 1 mm in the base and several times that in the apex (1 mm comprises approximately $3\times100=300$ outer hair cells). In the base of the cochlea, this lower bound on the spatial extent of power amplification is consistent with indirect estimates in guinea pig derived either from acoustic-lesion studies (Cody, 1992) or from manifestations of mechanical nonlinearity (Russell and Nilsen, 1997).

Although it is sometimes implied that outer hair cells or their components—their soma, their bundles, even individual proteins—are by themselves responsible for traveling-wave amplification, the cochlear amplifier is an emergent property of collections of cells coupled together by mechanical, electrical, and hydrodynamic interactions. As a result, there can be no simple relation between the forces produced by isolated hair cells, or by the proteins expressed within them, and the gain or other functional characteristics of the cochlear amplifier. Indeed, identical hair cells placed in different hydromechanical environments (e.g., in scalae of different heights) can produce dramatically different BM response functions (Shera *et al.*, 2005). Implicit in most models of cochlear mechanics, these observations warrant emphasis here only to remind ourselves that the obvious is easily overlooked (e.g., Poe, 1844).

As reviewed above, the propagation and gain functions depend only on BM velocity; they are robust to changes in the effective dimensionality of cochlear hydrodynamics. In a similar way, they place few specific constraints on the biophysical mechanisms involved in realizing the cochlear amplifier. Just as one can construct functioning optical lasers around a variety of gain media, so from the perspective of the cochlear propagation and gain functions, many details of organ of Corti micro- and nanomechanics remain shielded from view. A variety of different impedance and coupling combinations can combine to produce similar, even identical, response functions, and thereby yield similar propagation and gain functions.

The problem of understanding the biophysics of the gain medium must therefore be tackled simultaneously both from above and from below. Because the amplifier involves cooperation and coupling among many components, it will never be understood solely by dissecting individual cells or proteins. And because the emergent, hierarchical organization masks much of the underlying biophysics, the cellular workings of the amplifier cannot be uniquely determined from the mechanical response. Although this ultimate "inverse problem" remains ill posed, inverse methods such as those outlined here (see also Zweig, 1991; de Boer, 1995a) may nonetheless offer invaluable guidance during the two-pronged pursuit. They provide, in effect, invertible functional transformations of the data $[V_{\mathrm{BM}} \rightleftarrows (\kappa,\gamma)]$ that bring the mechanistic constraints those data impose more fully and directly into view.

ACKNOWLEDGMENTS

We thank Nigel Cooper, Marcel van der Heijden, Philip Joris, Ombeline de La Rochefoucauld, Elizabeth Olson, Alberto Recio-Spinoso, Tianying Ren, Mario Ruggero, and Andrei Temchin for generously sharing their data. We also thank Christopher Bergevin, Egbert de Boer, Paul Fahey, John Guinan, Ombeline de La Rochefoucauld, Jeffery Lichtenhan, Robert Withnell, and the two anonymous reviewers for valuable comments on the manuscript. This work was supported by grant R01 DC03687 from the NIDCD, National Institutes of Health.

[1]Both the physiological vulnerability and nonlinearity of mammalian cochlear mechanics could be achieved without cycle-by-cycle power amplification (negative damping) if outer hair cells (OHCs) serve primarily to modify the reactive (rather than the resistive) component of the BM impedance (cf. Kolston *et al.*, 1990). An example would be if the stiffness of the partition were to depend on the dc component of OHC receptor potentials (e.g., Allen, 2001).

[2]The inversion procedure earns its name because it solves an "inverse problem"—it obtains the values of model parameters (in this case, the propagation and gain functions) directly from experimental data.

[3]The subtracted exponential map [Eq. (4)] is equivalent to Greenwood's (1990) equation

$$\mathrm{CF}(x) = A[10^{a(1-x/L)} - k].$$

The parameters are related by $\mathrm{CF}(0)=A(10^a-k)$, $\mathrm{CF}_1=Ak$, and $l=L/(a\ln 10)$. (Greenwood's parameter k is a dimensionless constant and should not be confused with the wave number.) Approximate parameter values for the species {chinchilla, cat, gerbil} considered here are $\mathrm{CF}(0)=\{20.4,57,62.8\}$ kHz, $\mathrm{CF}_1=\{0.14,0.365,0.25\}$ kHz, and $l=\{3.8,5.17,2.19\}$ mm (Liberman, 1982; Greenwood, 1990; Müller, 1996).

[4]At fixed location, Eq. (5) implies that $\Delta\beta=[1+\beta_1(x)]\Delta\nu$. If $\Delta\nu$ represents the (fixed) width of the function $V_{\mathrm{BM}}[\nu(x_o),f)]$ when plotted on a ν axis, then the quality factor $[Q=\mathrm{CF}(x_o)/\Delta f]$ of the transfer function is given by $1/Q=\Delta\beta=[1+\beta_1(x)]\Delta\nu$. As a result, Q decreases and the transfer function broadens as the value of $\beta_1(x_o)$ increases toward the apex.

[5]In the two-dimensional model, the averaged pressure $\bar{P}(x,f)$ is defined by

$$\bar{P}(x,f) \equiv \frac{1}{H}\int_0^H P(x,y,f)dy,$$

where H is the scalae height. The difference pressure is defined by $P(x,y) \equiv p(x,y) - p(x,-y)$, where $p(x,y)$ is the scala pressure (i.e., the scala-vestibuli pressure for $y>0$ and the scala-tympani pressure for $y<0$). The pressure difference is antisymmetric about the partition: $P(x,-y) = -P(x,y)$.

[6]To relate these apical boundary conditions to the velocity of the stapes (oval window), we evaluate $\partial_x \bar{P}$ at $x=0$ using Eq. (8). The result is

$$\partial_x \bar{P}|_{x=0} = -bZ_f \int_0^L V_{BM}(x,f)dx.$$

The linearized Euler equation implies that

$$\partial_x \bar{P}|_{x=0} = -Z_f U_{ow},$$

where U_{ow} is the volume velocity of the oval window (Shera *et al.*, 2005). Together, these two equations imply that the oval-window volume velocity equals the total (integrated) volume velocity of the BM:

$$U_{ow}(f) = b\int_0^L V_{BM}(x,f)dx.$$

When evaluated using experimental data, the integral on the right-hand side might be termed the "virtual stapes volume velocity" (cf. de Boer and Nuttall, 1999).

[7]Problems with branch cuts can be minimized by defining the square root as $\sqrt{z(x)} = \sqrt{|z(x)|}e^{i\theta(x)/2}$, where $\theta(x)$ is the *unwrapped* phase of $z(x)$.

[8]An example using plane waves with constant amplitude and wavelength may be helpful. Consider the (standing) wave $P(x) = \cos[k(x-L)]$, which can be represented as a superposition of forward- and reverse-traveling components:

$$P(x) = \tfrac{1}{2}[e^{-ik(x-L)} + e^{+ik(x-L)}].$$

If the function $V(x)$ is related to $P(x)$ by the differential equation $V(x) = \partial_x^2 P(x)$, then $V(x) = -k^2 P(x)$ and the wave number can be found using the formula $k^2 = -V(x)/P(x)$. When given only $V(x)$, one can find $P(x)$ by double integration:

$$P(x) = 1 + \int_x^L dx' \int_{x'}^L V(x'')dx'',$$

where the constants and limits of integration have been chosen to satisfy the boundary conditions [in this example, $P(L)=1$ and $\partial_x P(L)=0$]. The formula $k^2 = -V/P$ for the wavenumber then yields

$$k^2 = -V(x) \Bigg/ \left[1 + \int_x^L dx' \int_{x'}^L V(x'')dx''\right],$$

a result easily verified by direct substitution of the value $V(x) = -k^2\cos[k(x-L)]$.

[9]About half (27/51) of the remaining Wiener kernels gave fair reconstructions; the rest were very bad.

[10]Performing causal fits (e.g., Zweig, 1991) to the ensemble of functions yielded results almost identical to those obtained using loess (see Sec. III D).

[11]Although the result is more general, Eq. (12) is easily derived for the two-dimensional box model. The total time-averaged power $W(x,f)$ passing the point x towards the helicotrema is given by

$$W(x,f) = \tfrac{1}{2}b\,\mathrm{Re}\int Pu^* dy,$$

where $u(x,y)$ is the x component of the fluid particle velocity; b is the scala width; and the integral is over the scala height. The time-average power per unit BM area transfered to the traveling pressure-difference wave from the organ of Corti is given by $1/b$ times the spatial derivative of $W(x,f)$:

$$w(x,f) = \frac{1}{b}\partial_x W(x,f) = \tfrac{1}{2}\,\mathrm{Re}\int u^*\partial_x P + P\partial_x u^* dy.$$

The first term in the integrand can be eliminated using the linearized Euler equation $[\partial_x P = -2i\omega\rho_0 u]$. Substitution yields $-2i\omega\rho_0|u|^2$, which is purely imaginary and does not contribute to the real part of the integral that defines the power transfer. The second term can be simplified by evaluating $\partial_x u$ using the Euler equation and applying Laplace's equation $[(\partial_x^2 + \partial_y^2)P = 0]$. The result is

$$w(x,f) = -\frac{1}{4\omega\rho_0}\,\mathrm{Im}\int P\partial_y^2 P^* dy.$$

The integration can be done by parts. Putting in the limits of integration yields

$$\int P\partial_y^2 P^* dy = P\partial_y P^*\Big|_0^H - \int |\partial_y P|^2 dy.$$

The second term on the right-hand side is real and drops out when taking the imaginary part indicated above. The contribution from the boundary at $y=H$ vanishes because the vertical component of the fluid velocity, and hence $\partial_y P$, must be zero at the ceiling. At the cochlear partition the particle velocity must equal the BM velocity; hence $\partial_y P|_{y=0+} = 2i\omega\rho_0 V_{BM}$. Putting it all together yields Eq. (12):

$$w(x,f) = -\tfrac{1}{2}\,\mathrm{Re}\,P_0 V_{BM}^*,$$

where $P_0(x) \equiv P(x,y=0^+)$ is the pressure across the cochlear partition.

[12]More generally, the impedance and wave number are related by $Z_{BM} = -\alpha Z_f/k^2$ [Shera *et al.*, 2005, Eq. (20)], where $\alpha \equiv P_0/\bar{P}$. Equation (13) therefore becomes

$$w(x,f) = -\omega\rho_0 H|V_{BM}|^2\,\mathrm{Im}\frac{\alpha}{(kH)^2},$$

where we have used the definition $Z_f = 2i\omega\rho_0/bH$. In the shortwave regime ($|kH|>1$) near the response peak, $\alpha \to kH$, and the power transfer simplifies to the value given in Eq. (14).

[13]To compute the values $\mathcal{P}_\uparrow$ shown in Fig. 15 (and the ratios $\mathcal{P}_\uparrow/\mathcal{P}_\downarrow$ in Fig. 17), we used the general formula for $w(x,f)$ given in Note 12, approximating α using the WKB approximation: $\alpha_{WKB} = kH/\tanh kH$ (Shera *et al.*, 2005). Values of the scalae height H were obtained from the measurements reported by Salt (2001) for chinchilla and Wysocki (2001) for cat.

[14]When computing $\mathcal{P}_\uparrow$ the BM with was assumed constant, consistent with the assumed model geometry. Although the BM width varies slowly along the cochlea, it remains approximately constant over the region of the wave peak, where the dominant contributions to the net power arise. Had changes in BM width been included in the calculation, the values of $\mathcal{P}_\uparrow$ shown in Fig. 15 would increase by roughly another factor of 3 from base to apex.

[15]The Kramers–Krönig dispersion relations are also known as Hilbert transforms, depending on context.

[16]The overall signs of the Kramers–Krönig relations given in Eqs. (16) and (17) differ from those in many references (e.g., Mathews and Walker, 1964; Lucarini *et al.*, 2005). The signs depend on the sign convention chosen for the time-domain Fourier transform, which, confusingly, has not been standardized across fields. Physicists use one convention, engineers the other. For historical reasons related to telephony, auditory physics has adopted the engineering convention.

[17]To obtain Eqs. (16) and (17) the more general Kramers–Krönig equations were simplified by assuming that $k(\nu) = k^*(-\nu)$, where $*$ represents complex conjugation. The existence of the integrals also requires that $\kappa(\nu)$ and $\gamma(\nu)$ vanish as ν approaches infinity.

[18]Note further that the group delay can only rigorously be interpreted as the physical delay of a wave packet (or "group") when the corresponding amplitude function is constant over the band of interest (e.g., Papoulis, 1962). This condition fails in the cutoff region, where the wave amplitude changes rapidly.

[19]van der Heijden and Joris (2006) estimate the wavelength of the traveling wave by fitting a regression line to the spatial phase profile in the "slow propagation region" near the characteristic place. The values they obtain agree well with values of $2\pi/\kappa$ averaged over the peak region.

[20]Although we focus here on the *power gain* of the amplifier, other measures

of gain have been employed in the literature (reviewed in Robles and Ruggero, 2001; de Boer and Nuttall, 2001).

[21]Chinchilla stimulus-frequency emissions (SFOAEs) evoked by low-intensity pure tones have mean levels about 20 dB below the stimulus (Siegel *et al.*, 2005). However, because of the substantial variability across frequency and between subjects, SFOAE and ANF measurements in the same animals are needed to quantify the validity of the no-emission approximation ($\mathcal{P}_{\Leftarrow} \cong 0$) in individual ears. The task is further complicated by variability in forward and reverse middle-ear transmission (Songer and Rosowksi, 2007). Since we lack the measurements needed to fully resolve the issue, we simply note that Eq. (27) overestimates the power gain in ears that produce significant stimulus reemission.

[22]The ratio G/G_{passive} provides a measure of power gain that partially accounts for the effects of passive losses. (G_{passive} is the value of G measured in the same preparation after the active elements have been disabled.) The accounting is imperfect because at fixed stimulus level (i.e., fixed $\mathcal{P}_{\Rightarrow}$) the passive losses themselves depend on BM velocity and therefore vary with G.

[23]Note that the dip in the power ratios evident near 2 kHz in Fig. 17 corresponds to the frequency region where chinchilla SFOAEs, and presumably their emitted power, are generally the smallest (Siegel *et al.*, 2005). This observation is consistent with the fact that emitted power ($\mathcal{P}_{\Leftarrow} > 0$) increases the value of r.

[24]In fact, the bandwidths are so narrow that they are called "linewidths."

[25]Although compelling evidence suggests that the backward-traveling waves involved in the production of SOAEs are slow, pressure-difference waves (Nuttall *et al.*, 2004), both slow and fast (i.e., compressional) waves "reflect" (i.e., produce forward-traveling slow waves) at the stapes (e.g., Shera *et al.*, 2007). Thus, the production of intracochlear standing waves—and the subsequent emergence of laser-like oscillations from the ear—does not depend, in principle, on the mechanisms or speed of reverse propagation.

Allen, J. B. (**2001**). "Nonlinear cochlear signal processing," *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Singular, San Diego, CA), pp. 393–442.

Bialek, W., and Wit, H. P. (**1984**). "Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear," Phys. Lett. **104A**, 173–178.

Bode, H. (**1945**). *Network Analysis and Feedback Amplifier Design* (Van Nostrand Reinhold, Princeton, NJ).

Cleveland, W. S. (**1993**). *Visualizing Data* (Hobart Press, Summit, NJ).

Cody, A. R. (**1992**). "Acoustic lesions in the mammalian cochlea: Implications for the spatial distribution of the 'active process'," Hear. Res. **62**, 166–172.

Cooper, N. P., and Rhode, W. S. (**1995**). "Nonlinear mechanics at the apex of the guinea-pig cochlea," Hear. Res. **82**, 225–243.

de Boer, E. (**1983**). "No sharpening? A challenge for cochlear mechanics," J. Acoust. Soc. Am. **73**, 567–579.

de Boer, E. (**1995a**). "The 'inverse problem' solved for a three-dimensional model of the cochlea. I. Analysis," J. Acoust. Soc. Am. **98**, 896–903.

de Boer, E. (**1995b**). "The 'inverse problem' solved for a three-dimensional model of the cochlea. II. Application to experimental data sets," J. Acoust. Soc. Am. **98**, 904–910.

de Boer, E. (**1996**). "Mechanics of the cochlea: Modeling efforts," *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 258–317.

de Boer, E. (**1997a**). "Classical and non-classical models of the cochlea," J. Acoust. Soc. Am. **101**, 2148–2150.

de Boer, E. (**1997b**). "Cochlear models and minimum phase," J. Acoust. Soc. Am. **102**, 3810–3813.

de Boer, E. (**1997c**). "Connecting frequency selectivity and nonlineary for models of the cochlea," Aud. Neurosci. **3**, 377–388.

de Boer, E., and Nuttall, A. L. (**1999**). "The 'inverse problem' solved for a three-dimensional model of the cochlea. III. Brushing-up the solution method," J. Acoust. Soc. Am. **105**, 3410–3420.

de Boer, E., and Nuttall, A. L. (**2001**). "Power gain of the cochlear amplifier," *Physiological and Psychological Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 1–7.

de La Rochefoucauld, O., and Olson, E. S. (**2007**). "The role of organ of Corti mass in passive cochlear tuning," Biophys. J. in press.

Duifhuis, H. (**1988**). "Cochlear macromechanics," *Auditory Function: Neurological Bases for Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 189–212.

Eggermont, J. J. (**1993**). "Wiener and Volterra analyses applied to the auditory system," Hear. Res. **66**, 177–201.

Eldredge, D. H., Miller, J. D., and Bohne, B. A. (**1981**). "A frequency-position map for the chinchilla cochlea," J. Acoust. Soc. Am. **69**, 1091–1095.

Geisler, C. D., and Sang, C. (**1995**). "A cochlear model using feed-forward outer-hair-cell forces," Hear. Res. **86**, 132–146.

Georgi, H. (**1993**). *The Physics of Waves* (Prentice-Hall, Englewood Cliffs, NJ).

Ghaffari, R., Aranyosi, A. J., and Freeman, D. M. (**2007**). "Longitudinally propagating traveling waves of the mammalian tectorial membrane," Proc. Natl. Acad. Sci. U.S.A. in press.

Golay, M. J. E. (**1961**). "Note on coherence vs narrow-bandedness in regenerative oscillators, masers, lasers, etc," Proc. IRE **49**, 958–959.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Guinan, J. J., Lin, T., and Cheng, H. (**2005**). "Medial-olivocochlear-efferent inhibition of the first peak of auditory-nerve responses: Evidence for a new motion within the cochlea," J. Acoust. Soc. Am. **118**, 2421–2433.

Gummer, A. W., Smolders, J. W. T., and Klinke, R. (**1987**). "Basilar membrane motion in the pigeon measured with the Mössbauer technique," Hear. Res. **29**, 63–92.

Karavitaki, K. D., and Mountain, D. C. (**2007**). "Evidence for outer hair cell driven oscillatory fluid flow in the tunnel of Corti," Biophys. J. **92**, 3284–3293.

Kemp, D. T. (**1979**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Kemp, D. T. (**2002**). "Exploring cochlear status with otoacoustic emissions: The potential for new clinical applications," *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–47.

Khanna, S. M., and Hao, L. F. (**1999**). "Reticular lamina vibrations in the apical turn of a living guinea pig cochlea," Hear. Res. **132**, 15–33.

Kiang, N. Y. S., and Moxon, E. C. (**1974**). "Tails of tuning curves of auditory-nerve fibers," J. Acoust. Soc. Am. **55**, 620–630.

Kolston, P. J., Viergever, M. A., de Boer, E., and Diependaal, R. J. (**1989**). "Realistic mechanical tuning in a micromechanical cochlear model," J. Acoust. Soc. Am. **86**, 133–140.

Kolston, P. J., Viergever, M. A., de Boer, E., and Smoorenburg, G. F. (**1990**). "What type of force does the cochlear amplifier produce?" J. Acoust. Soc. Am. **88**, 1794–1801.

Koshigoe, S., and Tubis, A. (**1982**). "Implications of causality, time-translation invariance, linearity, and minimum-phase behavior for basilar-membrane response functions," J. Acoust. Soc. Am. **71**, 1194–1200.

Kramers, H. A. (**1927**). "La diffusion de la lumiére par les atomes," *Atti del Congresso Internazionale dei Fisici* (Zanichelli, Bologna, Italy), Vol. **2**, pp. 545–547.

Krönig, R. (**1926**). "On the theory of dispersion of x-rays," J. Opt. Soc. Am. **12**, 547–557.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.

Liberman, M. C. (**1982**). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.

Liberman, M. C., and Kiang, N. Y. S. (**1978**). "Acoustic trauma in cats: Cochlear pathology and auditory-nerve activity," Acta Oto-Laryngol., Suppl. **358**, 1–63.

Lighthill, J. (**1981**). "Energy flow in the cochlea," J. Fluid Mech. **106**, 149–213.

Longfellow, E. W. (**1922**). *Random Memories* (Houghton Mifflin, Boston, MA).

Lucarini, V., Peiponen, K.-E., Saarinen, J. J., and Vartiainen, E. M. (**2005**). *Kramers-Kronig Relations in Optical Materials Research* (Springer, Berlin, Germany).

Mathews, J., and Walker, R. (**1964**). *Mathematical Methods of Physics* (Benjamin, New York).

Müller, M. (**1996**). "The cochlear place-frequency map for adult and developing Mongolian gerbil," Hear. Res. **94**, 148–156.

Nedzelnitsky, V. (**1980**). "Sound pressures in the basal turn of the cat cochlea," J. Acoust. Soc. Am. **68**, 1676–1689.

Neely, S. T. (**1983**). "The cochlear amplifier," *Mechanics of Hearing*, edited

by E. Boer and M. A. Viergever (Martinus Nijhoff, The Hague), pp. 111–118.
Nowotny, M., and Gummer, A. W. (**2006**). "Nanomechanics of the subtectorial space caused by electromechanics of cochlear outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **103**, 2120–2125.
Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). "Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig," J. Assoc. Res. Otolaryngol. **5**, 337–348.
Papoulis, A. (**1962**). *The Fourier Integral and its Applications* (McGraw-Hill, New York).
Poe, E. A. (**1844**). "The purloined letter," *The Gift: A Christmas, New Year's, and Birthday Present, 1845*, edited by E. L. Carey and A. Hart (Carey and Hart, Philadelphia, PA), pp. 41–61.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.
Recio-Spinoso, A., Temchin, A. N., van Dijk, P., Fan, Y.-H., and Ruggero, M. A. (**2005**). "Wiener-kernel analysis of responses to noise of chinchilla," J. Neurophysiol. **93**, 3615–3634.
Ren, T. (**2002**). "Longitudinal pattern of basilar membrane vibration in the sensitive cochlea," Proc. Natl. Acad. Sci. U.S.A. **99**, 17101–17106.
Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*," Aud. Neurosci. **3**, 101–121.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Russell, I. J., and Kössl, M. (**1999**). "Micromechanical responses to tones in the auditory fovea of the greater mustached bat's cochlea," J. Neurophysiol. **82**, 676–686.
Russell, I. J., and Nilsen, K. E. (**1997**). "The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.
Salt, A. N. (**2001**). "Cochlear fluids simulator v1.6h," ⟨http://oto.wustl.edu/cochlea/model.htm⟩. Last viewed 9/26/07.
Santos-Sacchi, J. (**2000**). "Cell coupling in Corti's organ," Brain Res. Rev. **32**, 167–171.
Shera, C. A. (**2001**). "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion," J. Acoust. Soc. Am. **109**, 2023–2034.
Shera, C. A. (**2003a**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.
Shera, C. A. (**2003b**). "Wave interference in the generation of reflection- and distortion-source emissions," *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 439–453.
Shera, C. A., and Guinan, J. J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., and Guinan, J. J. (**2007**). "Mechanisms of mammalian otoacoustic emission," *Active Processes and Otoacoustic Emissions in Hearing*, edited by G. A. Manley, R. R. Fay, and A. N. Popper (Springer, New York), in press.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2007**). "Otoacoustic estimates of cochlear tuning: Validation in the chinchilla," Assoc. Res. Otolaryngol. Abs. **30**, 519.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.
Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. F., and Guinan, J. J. (**2007**). "Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.
Shera, C. A., and Zweig, G. (**1991**). "Reflection of retrograde waves within the cochlea and at the stapes," J. Acoust. Soc. Am. **89**, 1290–1305.
Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.
Siegman, A. E. (**1986**). *Lasers* (University Science Books, Sausalito, CA), Errata available at ⟨http://www.stanford.edu/~siegman⟩. Last viewed 9/26/07.
Songer, J. E., and Rosowksi, J. J. (**2007**). "Transmission matrix analysis of the chinchilla middle ear," J. Acoust. Soc. Am. **122**, 932–942.
Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**1993**). "Electromechanical models of the outer hair cell," *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 207–214.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., and Tong, C. (**2001**). "Cochlear wave reflection due to roughness in 2-D and 3-D cochlear models," Assoc. Res. Otolaryngol. Abs. **24**, 44.
Temchin, A. N., Recio-Spinoso, A., van Dijk, P., and Ruggero, M. A. (**2005**). "Wiener kernels of chinchilla auditory-nerve fibers: Verification using responses to tones, clicks, and noise and comparison with basilar-membrane vibrations," J. Neurophysiol. **93**, 3635–3648.
Thiers, F. A., Burgess, B. J., and Nadol, J. B. (**2002**). "Reciprocal innervation of outer hair cells in a human infant," J. Assoc. Res. Otolaryngol. **3**, 269–278.
Toll, J. S. (**1956**). "Causality and the dispersion relation: Logical foundations," Phys. Rev. **104**, 1760–1770.
van der Heijden, M., and Joris, P. X. (**2003**). "Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies," J. Neurosci. **23**, 9194–9198.
van der Heijden, M., and Joris, P. X. (**2006**). "Panoramic measurements of the apex of the cochlea," J. Neurosci. **26**, 11462–11473.
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Wysocki, J. (**2001**). "Dimensions of the vestibular and tympanic scalae of the cochlea in selected mammals," Hear. Res. **161**, 1–9.
Zinn, C., Maier, H., Zenner, H. P., and Gummer, A. W. (**2000**). "Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in vivo guinea-pig cochlea," Hear. Res. **142**, 159–183.
Zweig, G. (**1976**). "Basilar membrane motion," *Cold Spring Harbor Symposia on Quantitative Biology*, (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. **XL**, pp. 619–633.
Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.
Zweig, G., Lipes, R., and Pierce, J. R. (**1976**). "The cochlear compromise," J. Acoust. Soc. Am. **59**, 975–982.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.
Zwiebach, B. (**2004**). *A First Course in String Theory* (Cambridge University Press, Cambridge, UK).

A state space model for cochlear mechanics

Stephen J. Elliott,[a] Emery M. Ku, and Ben Lineton
Institute of Sound and Vibration Research, University of Southampton, Southampton, Hampshire SO17 1BJ, United Kingdom

(Received 14 March 2007; revised 14 August 2007; accepted 14 August 2007)

The stability of a linear model of the active cochlea is difficult to determine from its calculated frequency response alone. A state space model of the cochlea is presented, which includes a discretized set of general micromechanical elements coupled via the cochlear fluid. The stability of this time domain model can be easily determined in the linear case, and the same framework used to simulate the time domain response of nonlinear models. Examples of stable and unstable behavior are illustrated using the active micromechanical model of Neely and Kim. The stability of this active cochlea is extremely sensitive to abrupt spatial inhomogeneities, while smoother inhomogeneities are less likely to cause instability. The model is a convenient tool for investigating the presence of instabilities due to random spatial inhomogeneities. The number of unstable poles is found to rise sharply with the relative amplitude of the inhomogeneities up to a few percent, but to be significantly reduced if the spatial variation is smoothed. In a saturating nonlinear model, such instabilities generate limit cycles that are thought to produce spontaneous otoacoustic emissions. An illustrative time domain simulation is presented, which shows how an unstable model evolves into a limit cycle, distributed along the cochlea. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2783125]

PACS number(s): 43.64.Kc, 43.64.Jb, 43.40.Vn, 43.64.Bt [BLM] Pages: 2759–2771

I. INTRODUCTION

It is important to develop mathematical models of cochlear mechanics to test our understanding of the physical processes involved. Modeling accomplishes this by providing predictions of complex system responses, based on assumptions regarding the underlying physical processes, which can be compared with direct measurements. Linear models of the cochlea, such as those presented by Zwislocki (1950), Allen (1977), and Neely and Kim (1986), as reviewed by de Boer (1996) for example, can provide predictions of the distribution of motion along the cochlea at a given frequency, or of its frequency response at a given position. These models include the macromechanical behavior of the fluid coupling along the length of the cochlea, as well as the micromechanical behavior of the individual parts of the cochlear partition. Such models are the starting point for more realistic nonlinear models. In order to produce numerical results, the cochlear partition, which has mechanical parameters that vary continuously along its length, is often approximated by a discrete set of elements. This allows a finite dimensional set of equations to be solved one frequency at a time (Neely, 1981). The number of elements is generally quite large, typically about 500, so that a single model generates many hundreds of individual frequency response functions.

One disadvantage of such a frequency domain formulation is that it can be difficult to determine whether the system being modeled is stable or not. With an active micromechanical element to represent the cochlear amplifier, energy can be supplied to the system at any point along the cochlea and so the stability of the overall coupled model is not guaranteed.

The stability of cochlear models is important for several reasons. For an unstable system, a frequency domain model will still yield results. However, these results must be treated with some care, since the assumption of a fixed amplitude output for a sinusoidal input has been violated. Also, it may not always be obvious that the model is unstable simply from inspection of the predicted frequency response, possibly leading to the misinterpretation of these results. This is particularly true of cochlear models containing distributed inhomogeneities (e.g., Zweig and Shera, 1995), where the predicted responses involve both anterograde and retrograde traveling waves. In the current paper, this problem is overcome by the development of a state space model that allows a rigorous approach to the study of model stability.

A second benefit of this model is that it facilitates the study of spontaneous otoacoustic emissions (SOAEs). For example, with this approach, it is easy to determine unambiguously whether or not a given cochlear model will lead to the generation of SOAEs; this in turn enables the examination of the conditions that lead to SOAE generation.

Two aspects of the stability of cochlear models that have been previously discussed in the literature are the stability of a single isolated point on the basilar membrane, and the stability of the entire cochlear model after all points on the basilar membrane have been coupled together via fluid interactions. Kanis and de Boer (1993, p. 3201) suggest that a single isolated point on the basilar membrane will be unstable if the resistance (i.e., the real part of the driving point impedance) is negative at any frequency. However, while a negative resistance is a necessary condition for instability, it is not sufficient. As pointed out by Zweig (1991, p. 1246), a

[a]Electronic mail: sje@isvr.soton.ac.uk

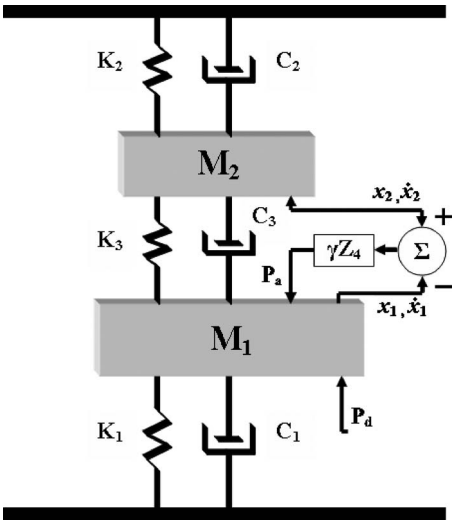


FIG. 1. Equivalent mechanical system for Neely and Kim's active micromechanical model.

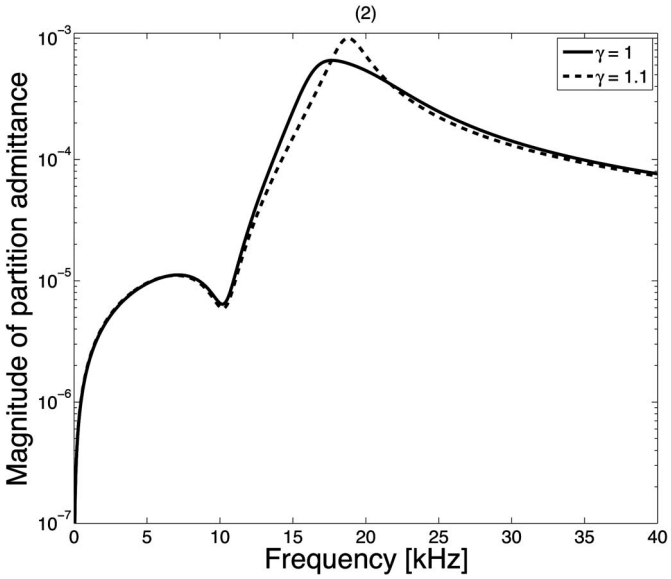


FIG. 2. Magnitude of the partition admittance of an isolated micromechanical element for feedback gains of $\gamma=1$ and 1.1.

system may have a negative resistance and still remain stable, provided the real parts of all its poles are negative.

While it is relatively straightforward to determine the stability of a single uncoupled point on the basilar membrane, this task is much more difficult when dealing with the full cochlear model. As for the single point, the suggestion by Kanis and de Boer (1993), and van Hengel (comment in Shera and Zweig, 1993, p. 62) that instability will arise whenever the resistance of the driving point is negative, is not necessarily true. Koshigoe and Tubis (1983) suggested using the Hilbert transform to check that causality (and hence stability) was implied by the resulting frequency responses of the model. While this is a valid method of determining instability, it has the disadvantage that responses over a wide range of frequencies must be obtained, and that (at least in theory) the frequency response of every elemental point in the model must be checked.

As an example of the frequency domain analysis of an unstable system, we consider the calculated response of a single, isolated, active micromechanical element as described by Neely and Kim (1986). Their micromechanical element is a two degree of freedom system, as illustrated in Fig. 1, in which an active pressure, p_a, acts on the lower mass, which is proportional to the relative displacement of the two masses. The admittance of this isolated element of the cochlear partition can be written as

$$Y_p(j\omega) = (b/g)[Z_1(j\omega) + Z_2(j\omega)(Z_3(j\omega) - \gamma Z_4(j\omega))/(Z_2(j\omega) + Z_3(j\omega)]^{-1}, \quad (1)$$

in which

$$Z_1(j\omega) = \frac{K_1}{j\omega} + C_1 + j\omega M_1, \quad Z_2(j\omega) = \frac{K_2}{j\omega} + C_2 + j\omega M_2,$$

$$Z_3(j\omega) = \frac{K_3}{j\omega} + C_3$$

and

$$Z_4(j\omega) = \frac{K_4}{j\omega} + C_4,$$

where b is the ratio of the maximum to average basilar membrane (BM) displacement, g is the lever gain, the masses and stiffnesses are all defined in Fig. 1, and γ is the gain of the feedback loop used to model the cochlear amplifier.

Equation (1) corresponds to the frequency response of the velocity of M_1 with respect to the pressure acting upon it, and its magnitude is plotted in Fig. 2 for the parameters listed in Table I at $x=5$ mm, for feedback gains of $\gamma=1$ and $\gamma=1.1$. Although the calculated frequency response for $\gamma=1.1$ is smoothly varying and apparently plausible, it has no physical significance in isolation, since the system is un-

TABLE I. Parameters of the micromechanical model of Neely and Kim (1986) with corrections for original typographical errors and converted into SI units.

Quantity	Formula (SI)
$K_1(x)$	$1.1^*10^{10}e^{-400x}$ N*m^{-3}
$C_1(x)$	$200+15\,000e^{-200x}$ N*s*m^{-3}
M_1	3^*10^{-2} kg*m^{-2}
$K_2(x)$	$7^*10^7e^{-440x}$ N*m^{-3}
$C_2(x)$	$100e^{-220x}$ N*s*m^{-3}
M_2	5^*10^{-3} kg*m^{-2}
$K_3(x)$	$1^*10^8e^{-400x}$ N*m^{-3}
$C_3(x)$	$100e^{-80x}$ N*s*m^{-3}
$K_4(x)$	$6.15^*10^9e^{-400x}$ N*m^{-3}
$C_4(x)$	$100e^{-80x}$ N*s*m^{-3}
gamma	1
g	1
b	0.4
L	0.025 m
K_S	2.1^*10^6 N*m^{-3}
C_S	4000 N*s*m^{-3}
M_S	45^*10^{-2} kg*m^{-2}
ρ	1000 kg*m^{-3}

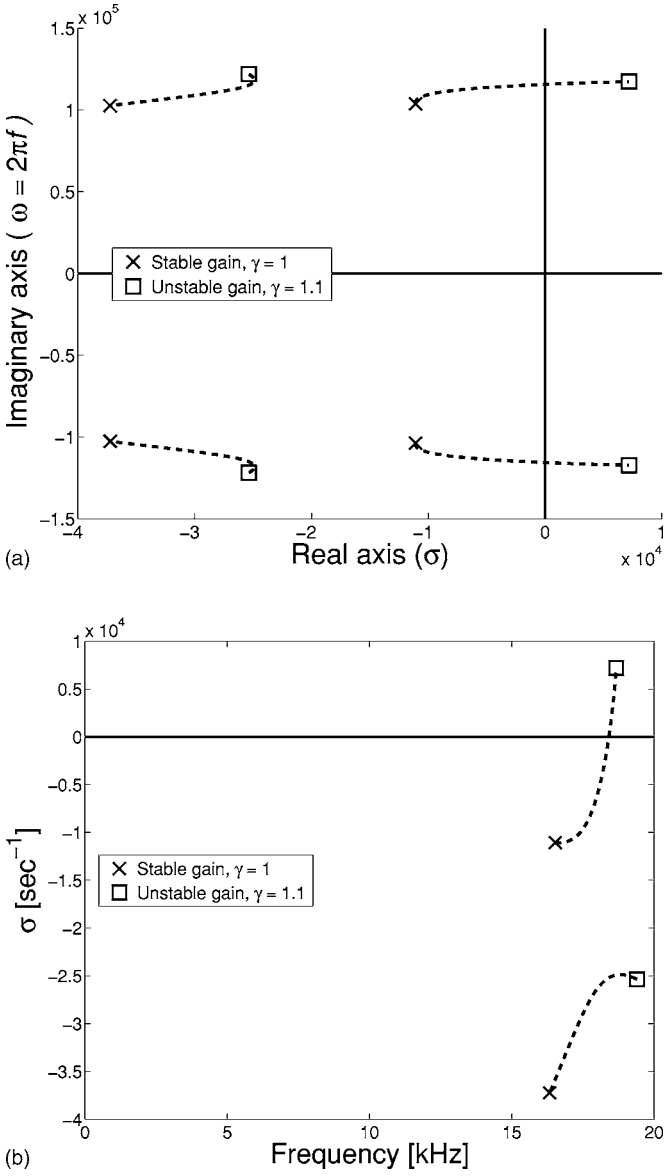


FIG. 3. Positions of the poles of the transfer function for the micromechanical element with feedback gains of $\gamma=1$ and $\gamma=1.1$, plotted: (a) using the conventional control representation with positive and negative frequencies plotted vertically; and (b) using the representation used for the cochlea model described here, with frequency plotted horizontally.

stable for this value of feedback gain and so the response to any sinusoidal input would be exponentially diverging, and ultimately infinite.

The stability of this single input, single output system can be determined by writing its transfer function as

$$Y_p(s) = (b/g)\{Z_1(s) + Z_2(s)[Z_3(s) - \gamma Z_4(s)]/ [Z_2(s) + Z_3(s)]\}^{-1}, \quad (2)$$

and calculating the values of the complex Laplace transform variable, s, for which $Y_p(s)$ tends to infinity, i.e., the poles of the system. Since the system has two degrees of freedom, it has two pairs of poles, which are plotted in the complex s plane for the system with feedback gains of $\gamma=1$ and $\gamma=1.1$ in Fig. 3. The positions of these poles are plotted in two ways in Figs. 3(a) and 3(b). Figure 3(a) shows the conventional representation used by the control community, with the real part of s (σ) along the horizontal axis and both the negative and positive components of the imaginary part of s ($\omega=2\pi f$) along the horizontal axis. The poles appear in complex conjugate pairs, so no additional information is gained by plotting those with a negative imaginary part. To help emphasise the physical significance of the poles, their imaginary parts can be divided by 2π, to give units of Hz, and the frequency plotted along the horizontal axis, as in a conventional frequency response, which is shown in Fig. 3(b).

One such pair of poles would give rise to a transient response proportional to $e^{\sigma t}\sin(2\pi f t)$. The real part of the pole position, σ, which is plotted on the vertical axis in Fig. 3(b), can thus be physically interpreted as the divergence rate of the transient. If σ is negative, the transient converges and the system is stable, but if σ becomes positive, the transient diverges and the system is unstable.

The two poles with positive real parts that appear when $\gamma=1.1$ thus clearly indicate that this system is unstable (Franklin *et al.*, 2005). Only the stability of the isolated micromechanical element is being considered in this example. When such an array of elements is coupled together in the cochlea, it is not generally possible to express the overall transfer function of the complete system analytically and hence determine the position of the poles. Although the phase of the calculated frequency response can give an indication of unstable behavior for this single input single output system, as can the causality of the inverse Fourier transform of the frequency response, it is difficult to formalize such tests for discretized cochlear models producing many hundreds of individual frequency responses. This is of particular concern given that numerical effects, such as truncation in the frequency domain, will also lead to the presence of small noncausal components.

The original motivation for developing a state space model of cochlear dynamics is that its stability can be unambiguously determined by examining the eigenvalues of a single matrix. These eigenvalues correspond to the system's poles, as described by Franklin *et al.* (2005), for example, and can be accurately calculated even for large matrices because of the wide availability of efficient numerical algorithms in software such as MATLAB. In such a state space model the dynamics of the underlying states of the system are expressed in the time domain as a set of coupled first-order differential equations (Furuta *et al.* 1988). Any instability of the system, including that of a hidden mode, is guaranteed to be represented in such a model (Skogestad and Postlethwaite, 1996). Several reliable software packages are now available specifically for solving ordinary differential equations cast in this form and for the direct time domain simulation of state space models. The interpretation of these models is also well established in the control engineering community. State space models have been proposed previously to describe the dynamics of the cochlea by Monderer and Lazar (1988), Stibler *et al.* (1998), and Lindgren and Li (2003). In each of these cases, however, the cochlea model was passive and so the stability of the model was not an issue.

A general framework for the state space description of a discretized cochlea model is set out in Sec. II. In Sec. III, the

features of this model are illustrated using Neely and Kim's (1986) formulation of the cochlea's micromechanical behavior. It is shown that when this model is stable, the frequency responses predicted using the state space model are identical to those calculated using the conventional frequency domain formulation for the parameters given by Neely and Kim.

In Sec. IV, the stability of this model is examined in more detail, particularly for variations of the cochlear amplifier's feedback gain along the length of the cochlea. It is shown that the stability of such a coupled cochlea is extremely sensitive to even small abrupt changes in gain, while the stability of the model is maintained for larger changes in feedback gain that occur more smoothly along the length of the cochlea.

The effect of random spatial inhomogeneities along the length of the cochlea is examined in Sec. V, using simulations of random gain perturbations having different magnitudes and length scales. It is shown that the number of unstable poles rises with the magnitude of the random perturbation, but perturbations of much less than 1% can easily produce many unstable poles. In a completely linear system these unstable poles would cause exponentially divergent oscillations. If the feedback forces in the micromechanical model saturate at a particular level, however, the amplitude of these oscillations will stabilize and the response reaches a limit cycle.

This nonlinear behavior is explored in Sec. VI, in which an example time domain simulation is presented, based on the state space model. The response of the cochlear model is computed when compressive nonlinearities are incorporated into each micromechanical feedback loop. A simulation of such a nonlinear cochlear model, which is unstable when linear, is presented showing that the nonlinearity does indeed limit the amplitude of the unstable behavior, resulting in a limit cycle distributed along the length of the cochlea.

II. GENERAL STATE SPACE FORMULATION

In this section, a state space formulation is developed for model of a discretized cochlea as formulated by Neely (1981), and as elaborated upon by Neely and Kim (1986). Adopting the long wavelength assumption, the starting point is the differential equation describing one-dimensional wave propagation along the cochlea (as discussed, for example, by de Boer, 1996)

$$\frac{\partial^2 p(t)}{\partial x^2} - \frac{2\rho}{H}\ddot{w}(t) = 0, \tag{3}$$

where $p(t)$ is the wave form of the pressure difference across the cochlear partition; $\ddot{w}(t)$ is the radially averaged transverse acceleration of the cochlear partition, both of which are also functions of x, although the dependence is suppressed for notational convenience; ρ is the density of the cochlear fluids; and H is the height of the canal above and below the cochlear partition, which is assumed to be constant. The boundary condition for the wave equation at the basal end can be written as

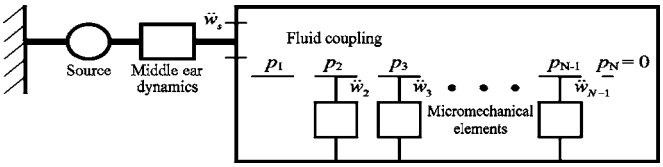


FIG. 4. The discrete model of the cochlea including the micromechanical models of the cochlear partition for elements 2 to $N-1$, the model of the middle ear dynamics at element 1 and of the helicotrema at element N.

$$\left.\frac{\partial p(t)}{\partial x}\right|_{x=0} = 2\rho\ddot{w}_S(t), \tag{4}$$

where $\ddot{w}_S(t)$ is the acceleration of the stapes footplate. This has two linearly superposing components: the unloaded part $\ddot{w}_{SO}(t)$, which is the acceleration due to an external excitation, normally the pressure in the ear canal, and $\ddot{w}_{SR}(t)$, which is the acceleration due to the loading by the internal pressure response in the cochlea at $x=0$. Thus, Eq. (4) may be written as

$$\left.\frac{\partial p(t)}{\partial x}\right|_{x=0} - 2\rho\ddot{w}_{SR}(t) = 2\rho\ddot{w}_{SO}(t). \tag{5}$$

At the apical end of the cochlea, a pressure release boundary condition at the helicotrema can be written as

$$p(t)|_{x=L} = 0. \tag{6}$$

Using finite difference approximations for the spatial derivatives in Eqs. (3), (5), and (6) as originally proposed by Neely (1981), in which the cochlear length L is divided into N sections of length Δ, the wave equation, Eq. (3), can be approximated by

$$\frac{p_{n-1}(t) - 2p_n(t) + p_{n+1}(t)}{\Delta^2} - \frac{2\rho}{H}\ddot{w}_n(t) = 0, \tag{7}$$

for $n=2$ to $n=N-1$, where $p_n(t)$ and $\ddot{w}_n(t)$ are the wave forms of the pressure difference and acceleration of the cochlear partition of the nth element. The basal boundary condition, Eq. (5), can also be approximated by

$$\frac{p_2(t) - p_1(t)}{\Delta} - 2\rho\ddot{w}_{SR}(t) = 2\rho\ddot{w}_{SO}(t). \tag{8}$$

The apical boundary condition, Eq. (6), can be written as

$$p_N(t) = 0. \tag{9}$$

This discrete model is illustrated in Fig. 4. Equations (7)–(9) can be written in matrix form (Neely, 1981), as

$$\frac{H}{2\rho\Delta^2}\begin{bmatrix} -\frac{\Delta}{H} & \frac{\Delta}{H} & & & & & & 0 \\ 1 & -2 & 1 & & & & & \\ 0 & 1 & -2 & 1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 1 & -2 & 1 & & 0 \\ & & & & 1 & -2 & 1 & \\ 0 & & & & & 0 & -\frac{2\rho\Delta^2}{H} & \end{bmatrix} \times \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-1}(t) \\ p_N(t) \end{bmatrix} - \begin{bmatrix} \ddot{w}_{SR}(t) \\ \ddot{w}_2(t) \\ \vdots \\ \ddot{w}_{N-1}(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \ddot{w}_{SO}(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{10}$$

which may also be written as

$$\mathbf{F}\mathbf{p}(t) - \ddot{\mathbf{w}}(t) = \mathbf{q}, \tag{11}$$

where $\mathbf{p}(t)$ and $\ddot{\mathbf{w}}(t)$ are the vectors of pressure differences and cochlear partition accelerations; $\mathbf{F}$ is the finite-difference matrix; and $\mathbf{q}$ is the vector of source terms.

We now cast the dynamics of each micromechanical element on the cochlear partition, for $n=2$ to $N-1$, in the state space form

$$\dot{\mathbf{x}}_n(t) = \mathbf{A}_n\mathbf{x}_n(t) + \mathbf{B}_n p_n(t) \tag{12}$$

$$\dot{w}_n(t) = \mathbf{C}_n\mathbf{x}_n(t) \tag{13}$$

where $\mathbf{x}_n(t)$ is the vector of state variables associated with the internal behavior of the nth micromechanical model.

Similarly, the state space equation describing the loading properties of the middle ear (located in our state space system at $n=1$) can be written as

$$\dot{\mathbf{x}}_1(t) = \mathbf{A}_1\mathbf{x}_1(t) + \mathbf{B}_1 p_1(t) \tag{14}$$

$$\dot{w}_{\mathrm{SR}}(t) = \mathbf{C}_1\mathbf{x}_1(t). \tag{15}$$

Although specific examples are given below for the forms of these micromechanical state space equations, they are kept in their most general form for now to illustrate the structure of the formulation. All of these elemental, uncoupled, state space models can be gathered together in the combined matrix equations

$$\dot{\mathbf{x}}(t) = \mathbf{A}_E\mathbf{x}(t) + \mathbf{B}_E\mathbf{p}(t) \tag{16}$$

$$\dot{\mathbf{w}}(t) = \mathbf{C}_E\mathbf{x}(t), \tag{17}$$

where the vectors are defined as

$$\mathbf{x}^T(t) = [\mathbf{x}_1^T(t)\ \mathbf{x}_2^T(t)\ \cdots\ \mathbf{x}_{N-1}^T(t)\ \mathbf{x}_N^T(t)], \tag{18}$$

$$\dot{\mathbf{w}}^T(t) = [\dot{w}_{\mathrm{SR}}(t)\ \dot{w}_2(t)\ \cdots\ \dot{w}_{N-1}(t)\ 0], \tag{19}$$

$$\mathbf{p}^T(t) = [p_1(t)\ p_2(t)\ \cdots\ p_{N-1}(t)\ p_N(t)], \tag{20}$$

and the matrices, which are block diagonal, are defined as

$$\mathbf{A}_E = \begin{bmatrix} \mathbf{A}_1 & 0 & \cdots & & \\ 0 & \mathbf{A}_2 & & & \\ \vdots & & \ddots & & \vdots \\ & & & \mathbf{A}_{N-1} & 0 \\ & & \cdots & 0 & \mathbf{A}_N \end{bmatrix}, \tag{21}$$

$$\mathbf{B}_E = \begin{bmatrix} \mathbf{B}_1 & 0 & \cdots & & \\ 0 & \mathbf{B}_2 & & & \\ \vdots & & \ddots & & \vdots \\ & & & \mathbf{B}_{N-1} & 0 \\ & & \cdots & 0 & \mathbf{B}_N \end{bmatrix}, \tag{22}$$

$$\mathbf{C}_E = \begin{bmatrix} \mathbf{C}_1 & 0 & \cdots & & \\ 0 & \mathbf{C}_2 & & & \\ \vdots & & \ddots & & \vdots \\ & & & \mathbf{C}_{N-1} & 0 \\ & & \cdots & 0 & \mathbf{C}_N \end{bmatrix}. \tag{23}$$

The overall vector of state variables $\mathbf{x}(t)$ should not be confused with the longitudinal spatial variable x used in Eq. (3).

We now assume that the matrix $\mathbf{F}$ in Eq. (11) is invertible, and express this equation for the fluid coupling within the cochlea in the form

$$\mathbf{p}(t) = \mathbf{F}^{-1}\ddot{\mathbf{w}}(t) + \mathbf{F}^{-1}\mathbf{q}(t). \tag{24}$$

Note from Eq. (17) that $\ddot{\mathbf{w}}(t)$ is equal to $\mathbf{C}_E\dot{\mathbf{x}}(t)$, so that

$$\mathbf{p}(t) = \mathbf{F}^{-1}\mathbf{C}_E\dot{\mathbf{x}}(t) + \mathbf{F}^{-1}\mathbf{q}(t). \tag{25}$$

Substituting this into Eq. (16) allows an overall state space equation for the coupled cochlear with distributed micromechanics and dynamic boundary conditions to be written in the general state space form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{26}$$

where

$$\mathbf{A} = [\mathbf{I} - \mathbf{B}_E\mathbf{F}^{-1}\mathbf{C}_E]^{-1}\mathbf{A}_E, \tag{27}$$

$$\mathbf{B} = [\mathbf{I} - \mathbf{B}_E\mathbf{F}^{-1}\mathbf{C}_E]^{-1}\mathbf{B}_E, \tag{28}$$

$$\mathbf{u} = \mathbf{F}^{-1}\mathbf{q}. \tag{29}$$

The matrix $\mathbf{A}$ is called the system matrix, and it determines the transient response of the system. In particular the eigenvalues of the matrix $\mathbf{A}$ are the poles of the system's transfer function and the real parts of all these eigenvalues must be negative if the transient response is to decay away, so that the system is stable.

III. A SPECIFIC EXAMPLE

In this section we demonstrate how the micromechanical model and boundary conditions used by Neely and Kim (1986) can be set in state space form, and reproduce some illustrative frequency responses from their paper to demon-

strate the equivalence of the state space formulation to the conventional frequency domain formulation for a stable system.

The Neely and Kim (1986) micromechanical model is an active second-order system, which must have two state variables associated with each degree of freedom. The four state variables are chosen to be the displacements of the two masses in Fig. 1, $x_1(t)$ and $x_2(t)$, and their velocities, $\dot{x}_1(t)$ and $\dot{x}_2(t)$. The coupled differential equations describing the dynamics of this micromechanical model, including the active feedback loop, may be written in terms of $x_1(t)$ and $x_2(t)$, as

$$\ddot{x}_1 = \frac{1}{M_1}\{p(t) + g\gamma[C_4(\dot{x}_2(t) - \dot{x}_1(t)) + K_4(x_2(t) - x_1(t))] - \dot{x}_1(t)(C_1 + C_3) - x_1(t)(K_1 + K_3) + \dot{x}_2(t)C_3 + x_2(t)K_3\}, \tag{30}$$

$$\ddot{x}_2 = \frac{1}{M_2}\{-\dot{x}_2(t)(C_2 + C_3) - x_2(t)(K_2 + K_3) + \dot{x}_1(t)C_3 + x_1(t)K_3\}. \tag{31}$$

These equations may be written in the form

$$\dot{\mathbf{x}}_n(t) = \mathbf{A}_n\mathbf{x}_n(t) + \mathbf{B}_n p_n(t), \tag{32}$$

by defining the state variable vector for this nth micromechanical element to be $\mathbf{x}_n$, which is equal to $[\dot{x}_1\ x_1\ \dot{x}_2\ x_2]_n^T$. The pressure difference across the cochlear partition, $p(t)$ in Eq. (29), is $p_n(t)$ in this case and $\mathbf{A}_n$ and $\mathbf{B}_n$ are defined as

$$\mathbf{A}_n = \begin{bmatrix} -\frac{(C_1 + C_3 - g\gamma C_4)}{M_1} & -\frac{(K_1 + K_3 - g\gamma K_4)}{M_1} & \frac{C_3 - \gamma C_4}{M_1} & \frac{K_3 - \gamma K_4}{M_1} \\ 1 & 0 & 0 & 0 \\ \frac{C_3}{M_2} & \frac{K_3}{M_2} & -\frac{(C_2 + C_3)}{M_2} & -\frac{(K_2 + K_3)}{M_2} \\ 0 & 0 & 1 & 0 \end{bmatrix}_n, \tag{33}$$

$$\mathbf{B}_n = \left[\frac{1}{M_1}\ 0\ 0\ 0\right]_n^T. \tag{34}$$

The differential equation used by Neely and Kim (1986) to describe the middle ear dynamics can be written

$$p_1(t) = M_S\ddot{w}_{SR}(t) + C_S\dot{w}_{SR}(t) + K_S w_{SR}(t), \tag{35}$$

where M_s, C_s, and K_s are the lumped mass, damper, and stiffness. Defining $\dot{w}_{SR}(t)$ and $w_{SR}(t)$ to be the two state variables for this model, it can be written in the form

$$\dot{\mathbf{x}}_1(t) = \mathbf{A}_1\mathbf{x}_1(t) + \mathbf{B}_1 p_1(t), \tag{36}$$

where

$$\mathbf{x}_1(t) = [\dot{w}_S(t)\ w_S(t)]^T, \tag{37}$$

$$\mathbf{A}_1 = \begin{bmatrix} -\frac{C_s}{M_s} & -\frac{K_s}{M_s} \\ 1 & 0 \end{bmatrix}, \tag{38}$$

$$\mathbf{B}_1 = \left[\frac{1}{M_S}\ 0\right]^T. \tag{39}$$

The pressure release boundary conditions at the helicotrema can also be represented in the state space form of Eq. (32), where

$$x_N = \dot{w}_H, \tag{40}$$

$$A_N = 0, \tag{41}$$

$$B_N = 0. \tag{42}$$

Provided the system is stable, the frequency response of a system described by the state space equations

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{43}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \tag{44}$$

can be derived by assuming each element of $\mathbf{x}(t)$ is proportional to $e^{j\omega t}$, so that $\mathbf{x}(t)$ can be written as $\mathbf{x}(j\omega)$ and $\dot{\mathbf{x}}(t)$ as $j\omega\mathbf{x}(j\omega)$. The overall matrix of frequency responses can thus be expressed as

$$\mathbf{y}(j\omega) = [\mathbf{D} + \mathbf{C}(j\omega\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}]\mathbf{u}(j\omega). \tag{45}$$

Equation (44) is the general equation for any output quantity, $\mathbf{y}(t)$, as a function of the states, $\mathbf{x}(t)$ and input $\mathbf{u}(t)$. For the case of interest here, the outputs are the BM displacements along the length of the cochlea, which are already contained, as x_{1n}, in the state vector of the micromechanical model, Eq. (32). Thus in this case, the $\mathbf{C}$ matrix just selects the appro-

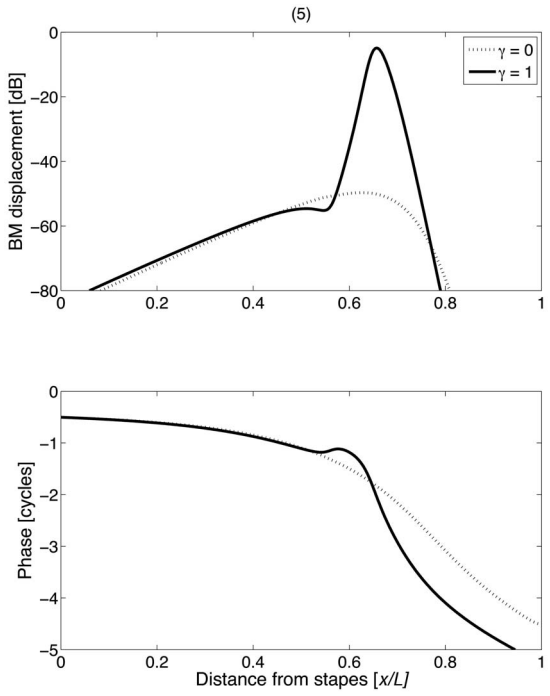


FIG. 5. Distribution of BM displacement along the cochlea calculated from the state space model for the parameters of Neely and Kim (1986) at 1.6 kHz with feedback gain $\gamma=0$ (dashed) and $\gamma=1$ (solid).

priate state variables and the **D** matrix is zero.

Equation (45) has been used with the formulation above and the parameters given in Table I to calculate the BM displacement at various frequencies as a function of position along the modeled cochlea. An example response is shown in Fig. 5, which is equivalent to Fig. 4 in Neely and Kim (1986).

The enhancement of the modeled active response can be varied in the Neely and Kim model by changing the value of the feedback gain, γ. The maximum overall enhancement in the frequency response at each position along the cochlea is plotted in Fig. 6 for the value $\gamma=1$, originally used by Neely and Kim. It can be seen that up to 90 dB enhancement is produced with this level of feedback gain. It has been argued, by Patazzi *et al.* (1989), for example, that this degree of enhancement overestimates that measured in the cochlea. Such a high degree of enhancement also indicates that the

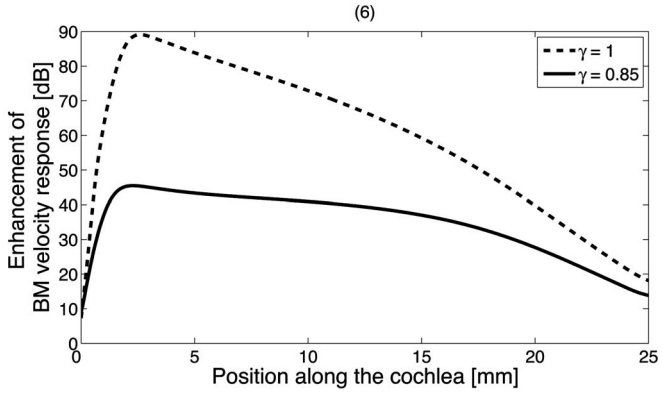


FIG. 6. Enhancement of the peak response in the BM velocity due to the cochlear amplifier for feedback gains of $\gamma=1$ and $\gamma=0.85$.

model is very close to instability (Franklin *et al.*, 2005). The stability of such a model might thus be considered unrealistically sensitive to perturbations in the model's parameters to be representative of the cochlea.

For these reasons, a value of feedback gain of $\gamma=0.85$ has been chosen to be the benchmark for the perturbation studies in this paper. The enhancement produced with this value of feedback gain is also shown in Fig. 6, which indicates that the maximum enhancement due to the cochlear amplifier is about 45 dB near the base and 20 dB near the apex, which is similar to the measured difference in response between live and dead cochlea (as reviewed, for example, by Robles and Ruggero, 2001).

Analyzing the stability of the coupled state space formulation of the cochlea is now readily accomplished by using the state space system to calculate the eigenvalues of the system matrix, **A**, as formulated in Eq. (43). Figure 7 shows these eigenvalues for the coupled system with feedback gains of $\gamma=0$, 0.85, 1.00, and 1.03. In this case, the number of elements in the discretized cochlear model was set to 500, as in the original Neely and Kim (1986) model, which includes the elements that account for the boundary conditions at each end. There are four poles per cochlear element, and three additional poles for the model of the boundary conditions, which makes a total of 1995 poles. Although individual poles can be distinguished on the bottom right-hand side of Fig. 7(b), the majority of the pole positions are so close together that they appear as a continuous line. Note that the distribution of poles in the lower part of Fig. 7(a) extends linearly below the range plotted, which was chosen to retain the detail in the other plots.

The coupled system is stable (since there are no poles with positive real parts) for both the passive model, with $\gamma=0$, and the active model with $\gamma=0.85$ or $\gamma=1$. As the model becomes active, the distribution of the poles nearest the frequency axis, which are the most lightly damped poles, moves further towards the frequency axis, and are so even more lightly damped. They are also stretched out in frequency, reflecting the increase in peak frequency for the active cochlea. When the gain in all the micromechanical elements is increased to 1.03, however, a number of poles appear with positive real parts, indicating that the system is unstable. As well as providing a convenient tool for establishing stability with uniform parameter distributions, the state space model also allows the stability of the cochlear model with nonuniform parameter distributions to be readily calculated, as considered in the following section.

IV. EFFECT OF SPATIAL INHOMOGENEITIES

It has been previously reported that variations in the physical parameters describing the basilar membrane along its length can result in cochlear instability. Although such spatial inhomogeneities could be present in any of the physical parameters in the micromechanical model, we use variations in feedback gain to illustrate some of the effects such inhomogeneities may have on the stability of the coupled system.

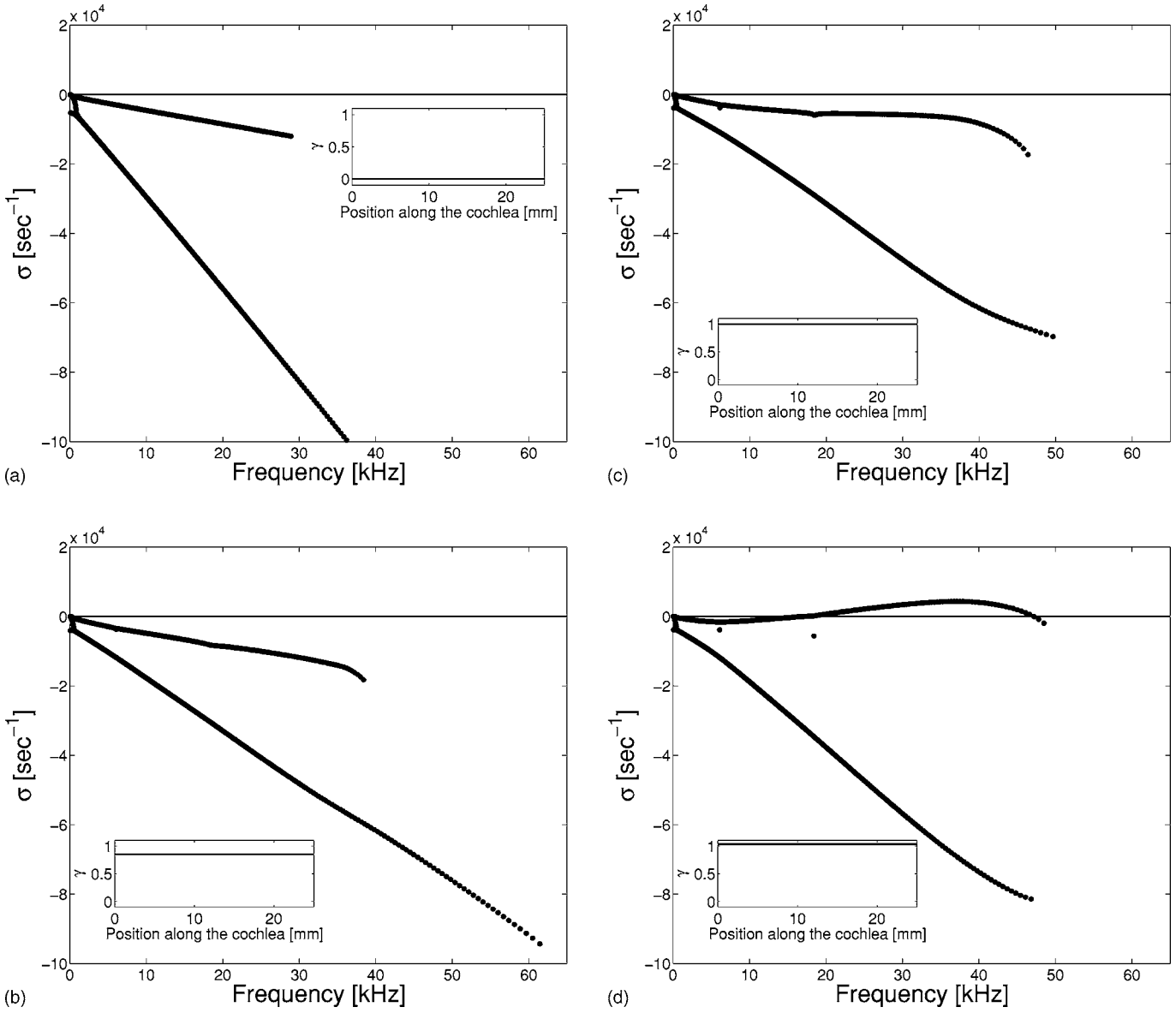


FIG. 7. Distribution of poles in the coupled cochlear model for various values of feedback gain in the micromechanical model, γ. Any pole with a positive real part denotes an unstable system. The cochlear model is clearly stable when it is passive, $\gamma=0$ (a), and also for feedback gains of 0.85 and 1 (b) and (c). The cochlear model is unstable if the feedback gain is increased to 1.03 (d).

A. Step changes in gain

Sharp discontinuities in the spatial distribution of gain can result in instability, even if the magnitudes of the changes are small. This is most evident when a step change in gain is introduced. Figure 8(a) shows the pole positions of a cochlea model with a 0.1% step reduction in gain at $x=5$ mm. The variation of feedback gain (γ) with position along the cochlear is also shown for reference.

In this case, a single pair of poles has a positive real part, indicating that the system is unstable. The frequency of the poles with the positive real part is close to the characteristic frequency of the position on the BM where the discontinuity is located. A 0.1% step increase of gain at 5 mm yields a similar result. The magnitude of the step change required to destabilize the cochlear model varies when discontinuities are introduced at different positions along the cochlea. The system's stability is most sensitive to discontinuities at the basal end, as illustrated above, and discontinuities of up to about 10% can be tolerated near the apical end, before the system becomes unstable.

It is generally accepted that cochlear instability, stabilized by nonlinearities into a limit cycle, is the source of spontaneous otoacoustic emissions (Bialek and Wit, 1984; Talmadge and Tubis, 1993; Nutall *et al.*, 2004). One theory regarding SOAEs suggests that they are not produced locally, but rather are the result of global wave scattering and active reinforcement of reflections by the active process, likening the cochlea to a laser oscillator (Talmadge and Tubis, 1993; Shera, 2003). A laser oscillator includes a resonant cavity which encloses a gain medium. The cochlear equivalent of a laser's resonant cavity exists between the region of a spatial discontinuity and the middle ear boundary; within that area, a minimum amount of energy must be reflected at each boundary in order to sustain a limit cycle. Here the outer hair

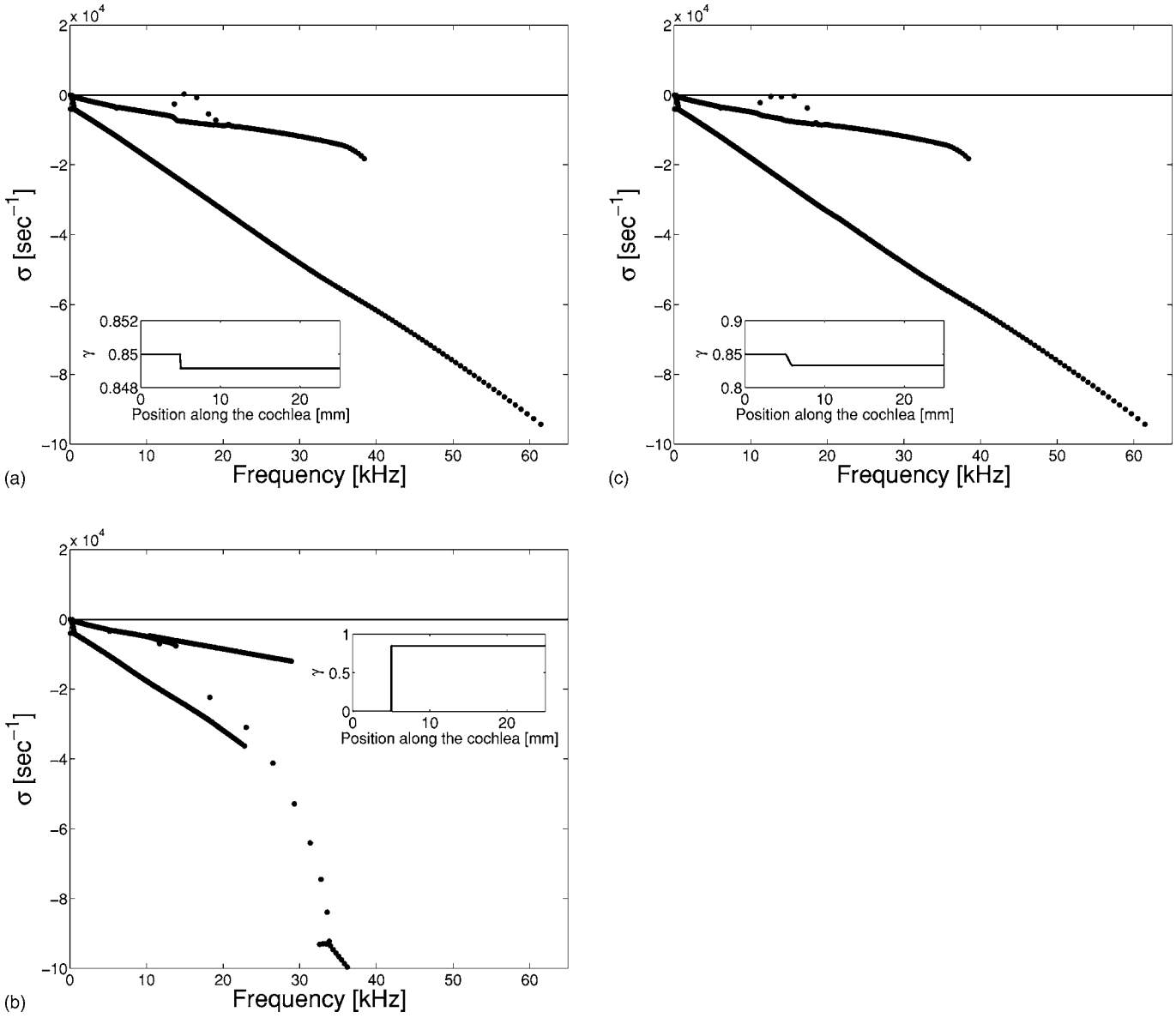


FIG. 8. Poles of the coupled cochlea for a gain distribution of feedback gains with: (a) a 0.1% step decrease at $x=5$ mm; system is unstable; (b) a step gain distribution of magnitude 0.85 at $x=5$ mm; system is stable; and (c) a smoothly varying gain distribution (quarter sin-squared wave dip) of magnitude 2% change at $x=5$ mm; system is stable. Note that the sketches of feedback gain with position along the cochlea have differing scales.

cells (OHCs) of the cochlea amplify waves traveling in each direction. While SOAEs are normally thought to be generated by random spatial perturbations in the dynamics of the cochlea (Zweig and Shera, 1995; Shera, 2003), pathologically abrupt discontinuities are also recognized as generating "atypical SOAEs" (Clark *et al.*, 1984; Lonsbury-Martin and Martin, 2001). Strictly speaking, it is oscillations of the BM motion that are being predicted here rather than the SOAEs, and modeling the detailed relationship between these oscillations and pressure in the external ear is the subject of current research.

Though small discontinuities in gain can drive the coupled cochlea unstable, one extreme spatial discontinuity in gain that does not generate instability is shown in Fig. 8(b), in which the cochlear amplifier gain is set to zero for all positions basal of $x=5$ mm. Note that the poles in the lower part of the graph continue below the plotted range, as in Fig. 7(a), and are very similar to those of the passive system in this frequency range, as expected. In light of the aforementioned theory of SOAE production, the system shown in Fig. 8(b) would have a zero roundtrip gain for any reflected waves between the gain discontinuity and the middle ear of the cochlea. As a result, this would be predicted not to produce enough energy to overcome the losses in that region, and consequently, no instability or SOAE would be produced; this is what is observed in the model here. Contrary to this observation, Fukazawa and Tanaka (1996) found that a similar gain distribution did result in system instability in their time domain model. It may be the feed-forward implementation of their cochlear model that produces this result.

B. Smooth variation in gain

Following the line of reasoning outlined above, it is possible that smoothly varying distributions of gain, associated with the compression of the cochlear response by a low pass

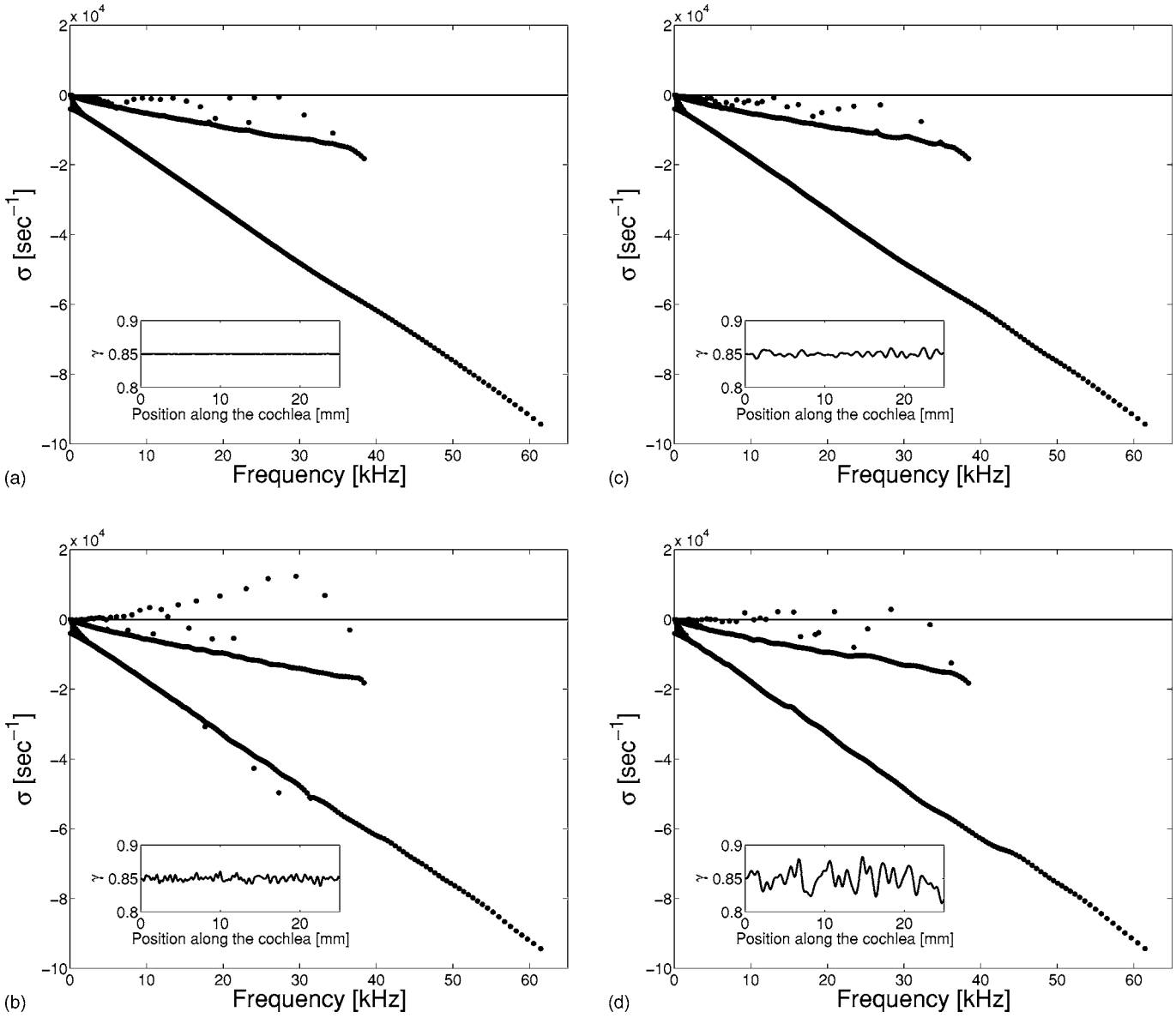


FIG. 9. Poles of the coupled cochlea for a random gain distribution of feedback gains: (a) with small magnitude (0.1%) and rough spatial variations; system is stable; (b) intermediate magnitude (~2%) and rough spatial variations; system is unstable; (c) intermediate magnitude (~2%) and smooth spatial variations; system is stable; and (d) large magnitude variations (~8%) and smooth spatial variations; system is unstable.

signal for example, might not reflect enough energy to sustain limit cycle oscillations. Figure 8(c) shows the pole positions for a 2% dip in gain over a 1 mm region of the cochlea centered on $x=5$ mm. The spatial distribution in the gain used in Fig. 8(c) is proportional to a quarter-$\sin^2(x)$ wave, and proves smooth enough to maintain stability. In contrast, a quarter-$\sin(x)$ wave form dip of the same magnitude in the same position, whose edges are more abrupt than the quarter-$\sin^2(x)$ dip, does not maintain stability. With the smooth quarter-$\sin^2(x)$ variation, even step changes of 100% in gain at the apical end of the cochlea do not destabilize the system.

V. SPATIALLY RANDOM VARIATION IN GAIN

Multiple spontaneous oscillations were observed in a nonlinear hardware model of a section of the cochlea by Zwicker (1986). Previous time domain simulations of active cochlear models have also produced internal oscillations that are believed to be representative of SOAEs (Fukazawa and Tanaka, 1996; van Hengel *et al.*, 1996; Talmadge and Tubis 1998). The active process has either been modeled using a feed-forward active force that is due to the incoming velocity from the base (Fukazawa and Tanaka, 1996) or a negative linear damper (van Hengel *et al.*, 1996; Talmadge and Tubis 1998).

In this section, linear systems with the Neely and Kim micromechanical model having randomly varying gain values are tested for stability. An entirely random wave form with a Gaussian distribution was band-pass filtered at different spatial frequencies in order to give smoothly or roughly spatially varying gain distribution. The filter applied is a fifth-order Butterworth filter, chosen for its characteristically flat passband (Lineton, 2001). The lower boundary of the filter is set at a spatial frequency given by the reciprocal of

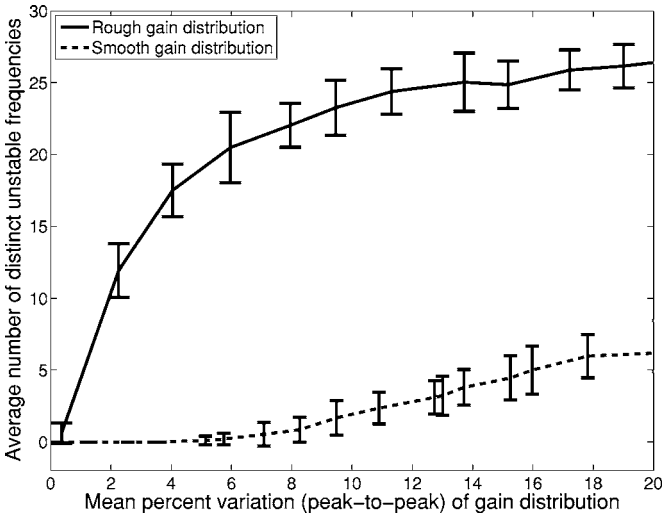


FIG. 10. The number of unstable poles in coupled cochlea models having different amplitudes of cochlear amplifier gain variation with either smooth or rough distributions along its length. The bars denote the range of 40 simulations with different random gain distributions and the circles denote the average value.

the length of the cochlea. The upper spatial frequency boundary of the filter is set within a range of 5–10% of the spatial sample rate, which corresponds to wavelengths on the order of several millimeters down to a fraction of 1 mm. Two upper spatial frequencies were used here: equivalent to wavelengths of 0.85 mm, which corresponds to a "rough" gain function and 1.7 mm, which corresponds to a "smooth" spatially varying gain function.

Figure 9 shows examples of randomly varying gain functions and their corresponding pole distributions. It can be seen that an increased magnitude of spatially varying gain fluctuation will cause instability, although lower percentage variations are required to cause the system to become unstable for rough spatially varying gains.

Different random gain variations will produce different distributions of poles, but the average number of unstable poles is dependent on both the smoothness of the gain distribution and its magnitude. This behavior was investigated in a series of simulations whose results are summarized in Fig. 10. The range of the number of unstable poles observed in 40 simulations is shown as a bar for different peak-to-peak percentage variations in gain. The average number of unstable poles is also shown, and these points are connected to illustrate the way in which the average number of unstable poles increases with percent variation in gain for both the rough and the smooth gain distributions. The system with the rough gain distribution becomes unstable even for 0.5% variations, whereas the system with the smooth gain distribution is stable for variations of up to about 5%. The average spacings between the poles when they are close to instability also cluster around a fixed value, as predicted by the theory of Zweig and Shera (1995), although the details of this behavior are still under investigation and will be the subject of a future publication.

VI. NONLINEAR TIME DOMAIN SIMULATIONS

Although an eigenvalue analysis of the state matrix can reveal whether a linear model is unstable, the exponentially

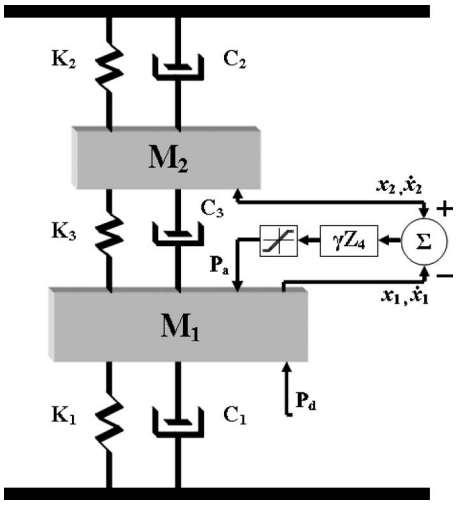


FIG. 11. Block diagram of the micromechanical model of Neely and Kim (1986) with a saturating nonlinearity in the active force.

diverging response of such a linear unstable system in a real cochlea would be limited by nonlinearities. In order to explore the complex behavior due to such nonlinearities, it is necessary to analyze the system in the time domain. Figure 11 shows the Neely and Kim micromechanical model with a saturation function operating on the active force, which might represent the nonlinearity of the outer hair cell response. This saturation is modeled here as a tanh function, which is linear for low-level excitation. It was noted in Sec. I that if the feedback gain, γ, was set to 1.1 in such an isolated model, the linear system would be unstable. The nonlinearity introduced in Fig. 11 will limit the amplitude of such an instability, as illustrated in Fig. 12, which shows the results of a time domain simulation of such an isolated nonlinear micromechanical element that is linearly unstable. The nonlinearity limits the amplitude of the oscillation, as predicted, so that in this case it settles down into a periodic limit cycle.

The state space formulation provides a convenient tool for such time domain analysis in the coupled cochlea model, using, for example, ordinary differential equation (ODE) solvers such as ode45 in MATLAB, which uses a Runge–Kutta

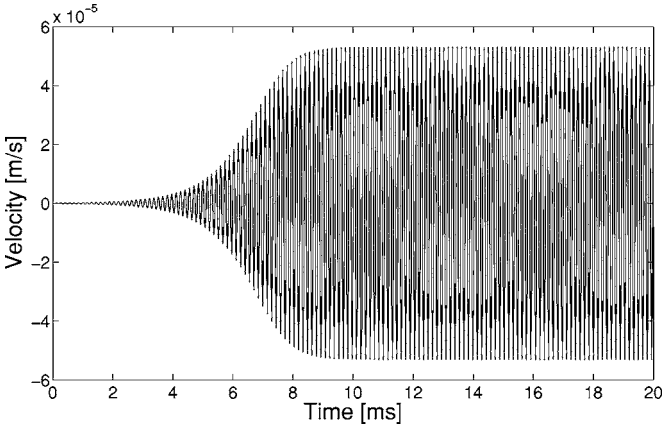


FIG. 12. The results of a time domain simulation of a single isolated nonlinear micromechanical element with $\gamma=1.1$, which is linearly unstable. The amplitude of the oscillation initially diverges exponentially, but is then limited by the saturation function so that it settles into a periodic limit cycle.

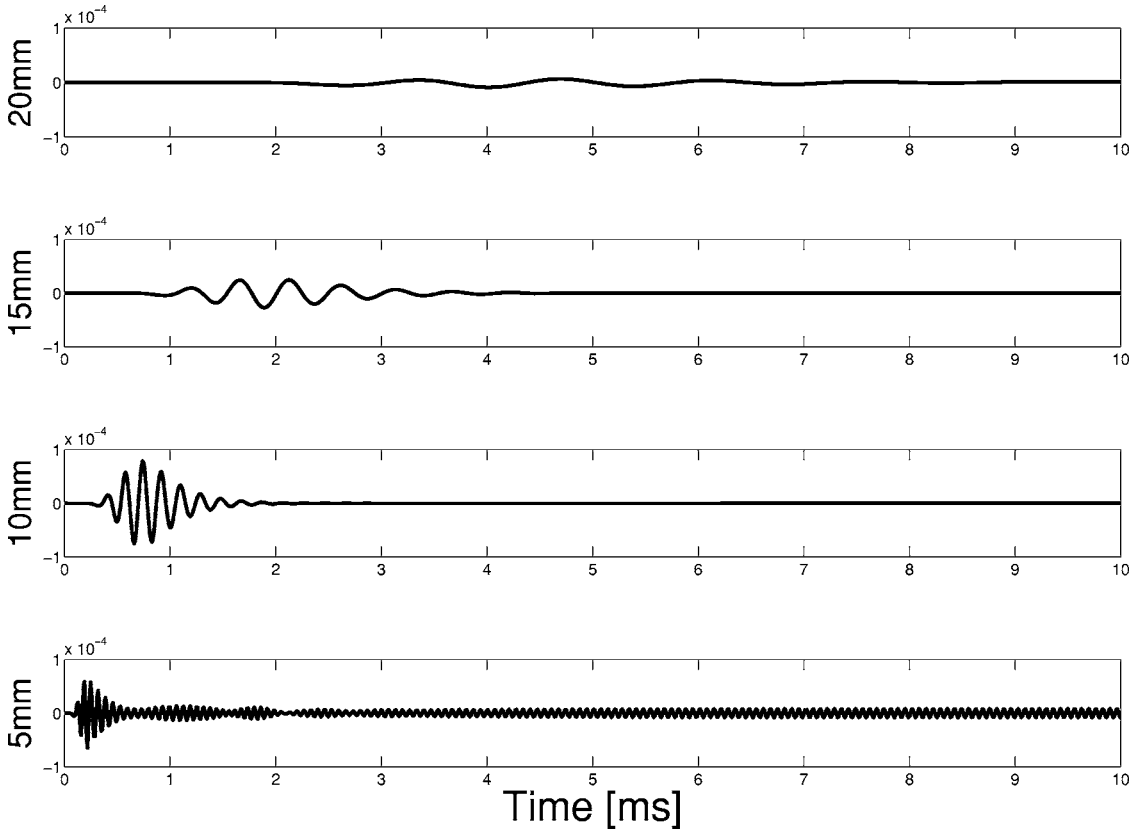


FIG. 13. BM velocity responses at different positions along the coupled cochlea as a result of a time domain simulation with saturating cochlear amplifiers and a distribution of feedback gains corresponding to Fig. 8(a).

algorithm with variable internal step size. This program has been used to simulate Eq. (26) in the time domain with an output sample rate of 100 kHz, which is necessary to avoid aliasing at the higher characteristic frequencies. The error tolerance in the ODE solver was set to 1e-10. Extremely similar results were obtained when the error tolerance was set to 1e-11, although the computation time was considerably greater in this case. The time domain simulations were initially validated against the frequency domain results by calculating the "steady-state" response at all positions along the cochlea to a 6 kHz tone after 100 ms of simulation. The results were almost identical to the frequency domain results, except where the response had fallen by about 200 dB at the apical end, in which case the finite window of the time domain simulations caused some spurious results.

Time domain simulations have also been performed with a nonlinear active model having a 0.1% step in the spatial gain distribution at $x=5$ mm, as used in Fig. 8(a), which is just sufficient to cause a single pair of poles to have positive real parts and is thus linearly unstable. The nonlinearity is incorporated into the simulation by replacing γ in the $\mathbf{A}_n$ matrix for each active micromechanical element, Eq. (33), with the instantaneous ratio of the input to output of the tanh function for that element. Figure 13 shows the evolution of the BM response at a number of positions along the cochlea for this gain distribution along the cochlea. In order to reduce the simulation time before the nonlinear model reaches a steady state, the cochlea is initially excited by a short pulse at the basal end. This causes a transient that dies away within 2 ms for positions less than about 10 mm along the cochlea, but takes longer to die away at more apical positions because of dispersion. This is more clearly seen in the grey scale plot shown in Fig. 14, which indicates the amplitude of the cochlear pressure as a function of time at all modeled positions along the cochlea. The original transient due to the impulsive excitation corresponds to an arc starting at the origin and

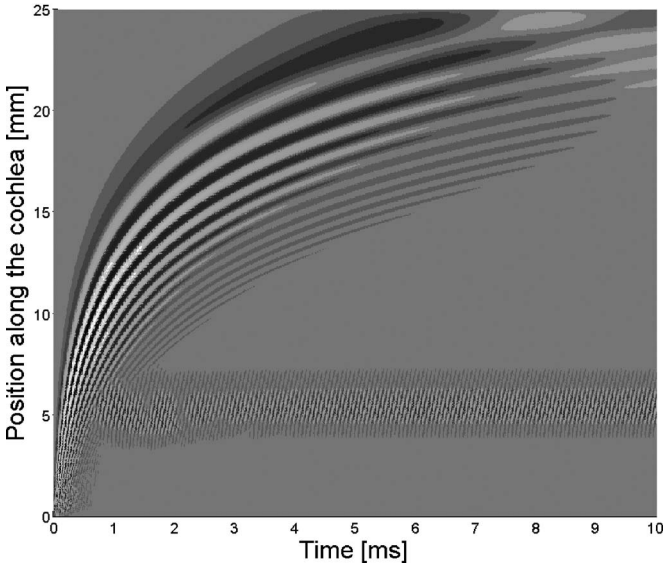


FIG. 14. Grey scale representation of the results of the time domain simulation described in Fig. 13 at all positions along the cochlea. The grey scale corresponds to the amplitude of the pressure (positive white, negative black) at each position at each time. An animation of this response is available online (Ku, 2007).

spreading out to 15 mm after about 3 ms. The strongest response after this transient has died away, however, is a single oscillation centered at $x=5$ mm along the cochlea. The frequency of this limit cycle is very similar to that of the imaginary part of the unstable pole, which is itself very similar to the characteristic frequency at the position of the step in gain.

VII. CONCLUSIONS

The main contribution of this paper is the formulation of a general state space model of the coupled cochlea. The prevalence of analytical and numerical tools for dynamic systems cast in this time domain form then allows several aspects of the behavior of cochlear models to be explored in more detail than is possible with an entirely frequency domain model.

Particular emphasis is placed on the assessment of stability for a linear model. The results of frequency domain models cannot be trusted unless the system can be demonstrated to be stable, and the pole positions derived from the state space model provide a convenient way of establishing whether such a coupled linear model is stable.

The stability of the cochlear model is illustrated using the model of Neely and Kim (1986). In particular, the effect of different distributions of micromechanical feedback gains along the cochlea is considered. It is shown that abrupt changes in feedback gain can cause instability even when the magnitude of the change is less than 0.1%. Smoother changes in gain can be much larger in magnitude before the stability of the model is compromised. The effect of various spatially random variations of feedback gain on stability is also investigated. It is found that there is a significant difference in the magnitude of such inhomogeneities that cause instability, depending on the smallest spatial wavelength of the feedback gain distribution.

Finally, we illustrate the use of the state space formulation to simulate the coupled system. Nonlinearities in the micromechanical feedback loops are included in order to begin to account for the finite active contribution of the OHCs. The limit cycle oscillations that are generated when the linearly unstable response is limited by the saturating nonlinearity are thought to be the source of SOAEs. The state space formulation may also be a convenient tool to study other forms of otoacoustic emission.

ACKNOWLEDGMENTS

We thank Professor Stephen Neely for his help in defining the parameters in his 1986 model. The work of Emery M. Ku was supported by a Fulbright Postgraduate Award.

Allen, J. B. (**1977**). "Cochlear micromechanics—a mechanism for transforming mechanical to neural tuning within the cochlea," J. Acoust. Soc. Am. **62**(4), 930–939.

Bialek, W., and Wit, H. (**1984**). "Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear," Phys. Lett. **104A**(3), 173–178.

Clark, W. W., Kim, D. O., Zurek, P. M., and Bohne, B. A. (**1984**). "Spontaneous otoacoustic emissions in chinchilla ear canals: Correlation with histopathology and suppression by external tones," Hear. Res. **16**, 299–314.

de Boer, E. (**1996**). "Mechanics of the cochlea: modelling effects," *The Cochlea*, edited by P. Dallas, A. N. Popper, and R. R. Ray, Chap. 5 (Springer-Verlag, New York, NY).

Franklin, G. F., Powell, J. D., and Emani-Naeini, A. (**2005**). *Feedback Control of Dynamics Systems*, 5th ed. (Addison Wesley, Reading, MA).

Fukazawa, T., and Tanaka, Y. (**1996**). "Spontaneous otoacoustic emissions in an active feed-forward model of the cochlea," Hear. Res. **95**, 135–143.

Furuta, K., Sano, A., and Atherton, D. (**1988**). *State Variable Methods in Automatic Control* (Wiley, New York).

Kanis, L. J., and de Boer, E. (**1993**). "Self-suppression in a locally active non-linear model of the cochlea: A quasilinear approach," J. Acoust. Soc. Am. **94**, 3199–3206.

Koshigoe, S., and Tubis, A. (**1983**). "Frequency-domain investigations of cochlear stability in the presence of active animals," J. Acoust. Soc. Am. **73**(4), 1244–1248.

Ku, E. M., ed. (**2007**). *Animations of time domain simulations*, Institute of Sound and Vibration Research, University of Southampton. Last viewed 8/14/2007. ⟨http://www.isvr.soton.ac.uk/SPCG/Cochlear_Mechanics/JASA2007/index.html⟩

Lindgren, A. G., and Li, W. (**2003**). "Analysis and simulation of a classic model of cochlea mechanics via a state-space realization." Last viewed 8/14/2007. ⟨http://www.ele.uri.edu/~liw/bm.pdf⟩

Lineton, B. (**2001**). "Testing a model of the stimulus frequency otoacoustic emissions in humans," PhD thesis, University of Southampton, Southampton, UK.

Lonsbury-Martin, B. L., and Martin, G. K. (**2001**). "Otoacoustic emissions," *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Singular, San Diego), 443–480.

Monderer, B., and Lazar, A. A. (**1988**). "Speech signal detection at the output of a cochlear model," Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP-88, 11–14 April, 1988, New York, NY.

Neely, S. T. (**1981**). "Finite difference solution of a two-dimensional mathematical model of the cochlea," J. Acoust. Soc. Am. **69**, 1386–1393.

Neely, S. T., and Kim, D. O. (**1986**). "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**(5), 1472–1480.

Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). "Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig," J. Assoc. Res. Otolaryngol. **5**, 337–348.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). "Outer hair cell receptor current and sensorineural hearing loss," Hear. Res. **42**(1), 47–72.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**(3), 1305–1352.

Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**(1), 244–262.

Shera, C. A., and Zweig, G. (**1993**). "Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emissions," *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore) 54–63.

Skogestad, S., and Postlethwaite, I. (**1996**). *Multivariable Feedback Control* (Wiley, New York).

Stibler, B. Z., Lewis, E. R., and Henry, K. R. (**1998**). "A state space model of gerbil cochlea," Proceedings of the 6th Annual Conference on Computational Neuroscience: Trends in research, 26–30 July, 1998, Big Sky, MT, 107–112.

Talmadge, C. L., and Tubis, A. (**1993**). "On modelling the connection between spontaneous and evoked otoacoustic emissions," *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 25–32.

Talmadge, C. L., and Tubis, A. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**(3), 1517–1543.

van Hengel, P. W. J., Dulfuis, H., and van den Raadt, M. P. M. G. (**1996**). "Spatial periodicity in the cochlea: The result of interaction of spontaneous emissions," J. Acoust. Soc. Am. **99**(6), 3566–3571.

Zweig, G. (**1991**). "Finding the impedance of the organ of corti," J. Acoust. Soc. Am. **89**, 1229–1254.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**(4), 2018–2047.

Zwicker, E. (**1986**). "A hardware cochlear model with active feedback," J. Acoust. Soc. Am. **80**(1), 146–153.

Zwislocki, J. (**1950**). "Theory of the acoustical action of the cochlea," J. Acoust. Soc. Am. **22**, 778–784.

Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay[a)]

Claus Elberling[b)]
Oticon A/S, "Eriksholm," DK-3070 Snekkersten, Denmark

Manuel Don
Electrophysiology Laboratory, House Ear Institute, Los Angeles, California 90057

Mario Cebulla
ENT Clinic, Julius Maximilians-University, D-97080 Würzburg, Germany

Ekkehard Stürzebecher
WDH Denmark, c/o D-15370 Petershagen, Germany

(Received 22 January 2007; revised 9 August 2007; accepted 22 August 2007)

This study investigates the use of chirp stimuli to compensate for the cochlear traveling wave delay. The temporal dispersion in the cochlea is given by the traveling time, which in this study is estimated from latency-frequency functions obtained from (1) a cochlear model, (2) tone-burst auditory brain stem response (ABR) latencies, (3) and narrow-band ABR latencies. These latency-frequency functions are assumed to reflect the group delay of a linear system that modifies the phase spectrum of the applied stimulus. On the basis of this assumption, three chirps are constructed and evaluated in 49 normal-hearing subjects. The auditory steady-state responses to these chirps and to a click stimulus are compared at two levels of stimulation (30 and 50 dB nHL) and a rate of 90/s. The chirps give shorter detection time and higher signal-to-noise ratio than the click. The shorter detection time obtained by the chirps is equivalent to an increase in stimulus level of 20 dB or more. The results indicate that a chirp is a more efficient stimulus than a click for the recording of early auditory evoked responses in normal-hearing adults using transient sounds at a high rate of stimulation. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783985]

PACS number(s): 43.64.Qh, 43.64.Bt, 43.64.Ri [WPS] Pages: 2772–2785

I. INTRODUCTION

A. Background

Recently we have focused our attention on the area of hearing screening and carried out a series of studies with the aim of improving response detection and stimulus condition for the recording of the auditory steady-state response (ASSR) in newborns (Stürzebecher *et al.*, 2005; Cebulla *et al.*, 2006; Stürzebecher *et al.*, 2006). In an attempt to improve the efficiency of frequency-specific ASSR stimulation a specific approach has been introduced that combines a new design method in the frequency domain with a compensation for the cochlear traveling wave delay (Stürzebecher *et al.*, 2006).

B. The chirp stimulus

Previous publications describe how the auditory compound action potential (ACAP) and the auditory brain stem response (ABR) can be recorded using chirp-stimuli that are designed to compensate for the cochlear traveling wave delay, e.g., Shore and Nuttall (1985), Lütkenhöner *et al.* (1990), and Dau *et al.* (2000). The traveling wave in the cochlea in response to a brief stimulus takes a considerable amount of time to reach from the base of the cochlea to the apex, i.e., from the highest to the lowest frequency responding area. The individual areas along the cochlear partition and the corresponding hair cells and nerve fibers of the auditory nerve will not be stimulated at the same time. Thus, the compound neural response will be temporally smeared. This temporal dispersion can be counteracted by delaying the higher frequencies relative to the lower frequencies of the stimulus. Such a scheme has to be based on an appropriate model of the traveling wave delay. A chirp stimulus attempts to compensate for the dispersion by aligning the arrival time of each frequency component in the stimulus to its place of maximum excitation along the basilar membrane. Such compensation will make the stimulus more efficient by achieving higher temporal synchronization between the evoked activities from the different neural elements that contribute to the formation of not only the ACAP but also the ABR and ASSR.

[a)]Portions of this work were presented in "Chirp stimuli–underlying assumptions," at the NHS-Conference, Como, Italy, May 2006; "ASSR stimuli based on design in the frequency domain and cochlear modelling," at the BAA-conference, Telford, United Kingdom, November, 2006; "ASSR stimuli for hearing testing of newborns and children," at the AAA-conference, Denver, CO, April, 2007, and "Chirp stimuli based on cochlear traveling wave delay," at the IERASG-conference, Bled, Slovenia, June 2007.

[b)]Electronic mail: ce@oticon.dk

C. Literature review

In the following overview several references are made to the high-pass masking and the notched-noise masking techniques. High-pass masking is used to obtain derived or narrow-band auditory evoked responses, e.g., Teas *et al.* (1962), Elberling (1974), Eggermont (1976), Parker and Thornton (1978), and Don and Eggermont (1978). Notched-noise masking is used to improve the place or frequency specificity of auditory evoked responses to tone burst stimuli, e.g., Picton *et al.* (1979) and Stapells *et al.* (1995).

One of the first attempts to compensate for the cochlear traveling delay with brief stimuli was made by Shore and Nuttall (1985), who used tone bursts of exponentially rising frequency to hypothetically activate synchronous discharges of the auditory nerve fibers along the cochlear partition by recording the ACAP in guinea pigs. The characteristics of the recorded ACAPs in response to a tone burst of rising and of falling frequency were compared. The rising frequency tone burst produced a narrower N_1-peak and a higher N_1-P_1-amplitude than the corresponding falling frequency tone burst. These results were further substantiated by a high-pass masking experiment and the authors concluded that the appropriate temporal organization of the spectral energy in accordance with the cochlea traveling wave properties increases the ACAP synchrony.

Lütkenhöner *et al.* (1990) constructed a chirp stimulus using the parameters for the cochlea traveling delay based on frequency specific ABR-data obtained by Pantev *et al.* (1985). Standard click and chirp ABRs were recorded at four different levels (range 5–35 dB nHL) in four normal-hearing human subjects. A comparative analysis revealed that the chirp resulted in higher neural synchronization, larger ABR amplitude, and better response detection than the corresponding click.

Dau *et al.* (2000) developed a chirp using the inverse delay-line characteristics of the cochlear partition based on the linear cochlear model by de Boer (1980). From the model the propagation time needed to arrive at the place of resonance was formulated and combined with the frequency-place mapping proposed by Greenwood (1990) in order to derive numerically the instantaneous phase of the resulting "exact" chirp. By using an asymptotic expression for the propagation constants an "approximated chirp" could be derived analytically. From the exact chirp a "flat-spectrum chirp" was constructed. ABRs were recorded in ten normal-hearing subjects and it was found that for most of the applied stimulus levels the flat-spectrum chirp generated ABRs with higher wave-V amplitude than the corresponding click stimulus. It was further concluded that in contrast to the click, the rising frequency chirp enables the inclusion of activity from lower cochlear frequency regions.

Bell *et al.* (2002) produced octave-band chirp stimuli based on the approximate chirp by Dau *et al.* (2000). The stimuli were used to estimate frequency-specific ABR thresholds in ten normal-hearing young adults. The obtained ABR thresholds were evaluated and found to be closer to behavioral thresholds and to yield smaller variance than reported for tone burst stimuli without masking, but not as close to the behavioral thresholds as reported for tone bursts in notched noise. The authors commented that band-limited chirps may have the disadvantage of a wider spectral spread than tone bursts and therefore elicit responses from unwanted frequency regions of the basilar membrane.

Wegner and Dau (2002) examined the usefulness of the chirp for retrieving frequency specific information. The version of the chirp having a flat magnitude spectrum (the flat-spectrum chirp) and a click were used in ABR experiments involving high-pass and notched-noise masking paradigms in nine normal-hearing young adults. The chirp generated higher wave-V amplitudes than the click in all high-pass masking conditions. However, neither the amplitudes of the narrow-band responses nor of the notched-noise responses differed significantly between the chirp and the click. The authors concluded that the applied one-octave bandwidths were not wide enough to demonstrate the higher neural synchronization with the chirp stimulus. Finally a low-frequency chirp was compared to a 250-Hz tone pulse with comparable duration and amplitude spectrum and it was found that at low and medium levels of stimulation the low-frequency chirp produced larger wave-V amplitudes than the tone pulse.

In a subsequent comparative paper, Fobel and Dau (2004) examined ABRs recorded to three different chirps (O, A, and M chirp) and a standard click in nine normal-hearing young adults. The O chirp used an estimate of the basilar membrane group delay derived from stimulus-frequency otoacoustic emissions (Shera and Guinan, 2000); the A chirp used an estimate of cochlear traveling time based on frequency-specific tone-burst ABRs (Neely *et al.*, 1988; Gorga *et al.*, 1988); and finally, the M chirp used the original flat-spectrum chirp (Dau *et al.*, 2000). All chirps elicited larger wave-V amplitudes than the click; only small differences were found between the O chirp and the M chirp ABRs at low and medium levels of stimulation; the A chirp generated the largest ABRs—especially at low levels.

Agung *et al.* (2005) used the chirp by Dau *et al.* (2000) in an experiment to investigate the post-auricular-muscle response (PAMR) in 12 normal-hearing young adults. The chirp was compared to a standard click and a /t/ stimulus using both a standard clinical earphone (EAR-3A) and one with extended high-frequency response (ER-2). The chirp was found to generate larger PAMRs than the click and the /t/ stimulus, and the PAMR was further augmented when the extended high-frequency response earphone was used.

Stürzebecher *et al.* (2005) developed frequency-specific stimuli for the recording of the ASSR. The stimulus design included, among other characteristics, compensation for the cochlear traveling wave delay based on the model of de Boer (1980). Four stimuli (centered at 500, 1000, 2000, and 4000 Hz) were constructed with a bandwidth of about 500 Hz. Each stimulus was compared to an equivalent counterpart designed without the compensation for the temporal dispersion in the cochlea. ASSRs were recorded in 70 normal-hearing young adults using a stimulus rate of 90/s and for each response the signal-to-noise ratio (SNR) was estimated. For all stimuli the compensated versions produced higher SNRs than the uncompensated ones. However, the compensation assumed that the propagation time from de

Boer's model, i.e., the latency-frequency function, to represent a phase delay instead of a group delay and therefore the compensation may have been less than optimal (see Sec. II B). Here it should be mentioned that Junius and Dau (2005) also recorded the ABR (ASSR) to higher rates of stimulation using chirps that were included in a specific stimulus paradigm in order to investigate neural adaptation of the ABR.

D. The Stacked ABR technique

Another technique to compensate for the traveling wave delay has been developed by Don *et al.* (1994). This technique can be explained as follows: First, the click-evoked ABR is decomposed into narrow-band contributions (the derived or narrow-band ABRs), which are obtained by successive high-pass masking. Next, the narrow-band ABRs are temporally aligned by time shifting in accordance with the observed traveling wave delay (i.e., peak wave V latencies of the derived bands), and third, the time-shifted derived-band ABRs are added together to produce the "stacked" ABR. In general, the Stacked ABR is significantly larger than the corresponding unmasked click ABR, which demonstrates that compensation for cochlear traveling time greatly affects the amplitude of the ABR. In subsequent clinical studies [Don *et al.* (1997, 2005b)], it was shown that this technique is particularly useful for the detection of small, intracanalicular acoustic tumors.

E. Input and output compensation

The Stacked ABR technique is based on the same fundamental concept that underlies the chirp stimulus, i.e., compensation of the cochlear response time in order to increase the temporal synchronization of the activity from the contributing neural elements. However, the Stacked ABR uses a traveling wave delay that is obtained from the narrow-band recordings in the *individual* subject and applies the compensation to the recorded *output* wave forms. In contrast to this, a chirp-evoked ABR uses an *average* traveling wave delay (via a cochlear or latency model) and applies the compensation to the *input* stimulus wave form.

In conclusion, both input compensation of the traveling wave delay (chirp stimulus) and output compensation (Stacked ABR) have demonstrated that appropriate compensation schemes increase the amplitude of early electrophysiological responses (ACAP, ABR and ASSR) by increasing the temporal synchronization between the contributing neural elements.

F. Models of the cochlear delay

Several models describe the cochlear traveling wave delay in humans. The effect on neural responses has been studied directly in a significant number of subjects with normal hearing by the recording of the ACAP (Eggermont, 1979) and of the ABR (Neely *et al.*, 1988; Donaldson and Ruth, 1993; Don *et al.*, 2005b). Characteristics of the traveling wave delay have been obtained by a cochlear model by, e.g., de Boer (1980) using the specific mapping parameters for the human cochlea described by Greenwood (1990).

G. The present study

In light of the findings by Fobel and Dau (2004) that chirps based on different estimates of the cochlear delay give different ABR amplitudes, we wanted to include still another estimate of the delay function. We expected that the improvement in response amplitude observed for the ACAP and the ABR when using a chirp instead of a click also applies to the ASSR using high rates of stimulation as indicated by the findings by Junius and Dau (2005). Therefore, we wanted to test the chirp stimuli constructed from different delay estimates with an ASSR paradigm on a significant number of normal-hearing young adults. These are the reasons for undertaking the studies presented herein.

The traveling wave delay constitutes one important aspect of the cochlear input/output characteristics. The input/output characteristics or transfer function of a linear system is fully described either in the frequency domain by the system's amplitude and phase response or in the temporal domain by the system's impulse response. However, the cochlea and the peripheral part of the nerve fibers constitute a nonlinear system and may not be approximated successfully by a linear system—even at a fixed stimulus level.

Therefore, in the first part of the present paper a linear system is used to evaluate the modification that is introduced by the traveling wave delay on the phase spectrum of the click stimulus. Based on this evaluation, it will be described how different chirps may be designed using a new design method in the frequency domain. Subsequently, the clinical efficiency of the different chirps will be examined by comparative electrophysiological ASSR testing in a group of 49 young adults with normal hearing.

II. CHIRP DESIGN

A. Cochlear traveling wave delay

In the literature on human cochlea physiology several studies present data that describe the cochlea traveling wave delay, i.e., the time it takes for the input sound to generate the neural output from different frequency locations along the cochlear partition. Since most of the data are based on electrophysiological experiments, the obtained values for the traveling wave delay also include mechanical, sensory cell, and neural activity. The active hair cells and their bandpass filter function are, therefore, included in these descriptions.

In the present study, the data will include reported delay values—also called latency-frequency functions—based on electrophysiological data described by Eggermont (1979), Neely *et al.* (1988), and Don *et al.* (2005a).[1] However, the observation of the mechanical response in the human cochlea as presented in a model by de Boer (1980) is also included because this model previously has been used for designing chirp stimuli (e.g., Dau *et al.*, 2000). The following data are thus included:

(1) The data from Eggermont (1979), which are based on *narrow-band* ACAPs recorded by trans-tympanic electrocochleography from 15 normal-hearing adults in response to click stimuli (90 dB p.e.SPL).

(2) The data from Neely *et al.* (1988) and Gorga *et al.* (1988), which are based on auditory brain stem responses to *frequency-specific stimuli* recorded by surface electrodes from 20 normal-hearing adults in response to tone-pulses (range: 20–100 dB SPL).
(3) The data from Don *et al.* (2005a), which are based on narrow-band ABRs, recorded by surface electrodes from 39 normal-hearing adults in response to click stimuli (60 dB nHL).
(4) The data from de Boer's model are based on the mechanical characteristics of the cochlea, de Boer (1980), using the frequency-to-location mapping by Greenwood (1990). The model is based on the mechanical properties, e.g., stiffness and wave patterns in human cochlea preparations originally observed by von Békésy (1960) and supplemented by later observations from the living cochlea in experimental animals by, e.g., Johnstone and Boyle (1967) and Rhode (1971).

The observed latency-frequency functions of the electrophysiological data can all be described by the same power function as suggested by Anderson *et al.* (1971) and subsequently used by Eggermont (1979) and Neely *et al.* (1988):

$$\tau = kf^{-d}, \tag{1}$$

where τ is the latency in seconds, f is the frequency in hertz, and k and d are constants (*in the Neely et al. data k varies with stimulus level*).

(1) The Eggermont data (Eggermont, 1979) were submitted to a curve-fitting procedure that gave the following constants: $k=0.6962$ and $d=0.77$. (Note that Eggermont subtracted a value of 0.8 ms for the synaptic gap delay prior to the curve fitting.)
(2) The Neely *et al.* data (Neely *et al.*, 1988) obtained at the stimulation level of 90 dB SPL were submitted to a curve-fitting procedure that gave the following constants: $k=0.0525$ and $d=0.413$. (Note that Neely *et al.* subtracted a value of 5 ms prior to defining their final power function [Eq. (1)]; 4 ms for the wave I-V latency difference and 1 ms for the synaptic delay.)
(3) The Don-data consist of observed, mean, narrow-band wave V-latency values that were obtained using an ER-2 earphone with a nominal tube length of 292 mm. In order to relate the values to the level of the cochlea both the acoustic delay of the earphone tube (0.86 ms) and the wave I-V latency difference (average 4.10 ms–Elberling and Parbo, 1987), i.e., a total of 4.96 ms, is subtracted from the observed latency values. (Note that no attempts have been made to correct for the synaptic gap delay.) The latency values are shown in Table I. The mean value corresponding to the highest band center frequency, i.e., 11.3 kHz is excluded, because this data point probably is invalid because the level of activation is much lower in this frequency region given the ear's sensitivity curve and the falling spectrum of the click above 8 to 9 kHz (Don *et al.*, 1979). The remaining four data points are fitted to the same power function as noted earlier [Eq. (1)], and the curve fitting procedure gives the following constants: $k=0.0920$ and $d=0.4356$ ($r=0.99$). If the fitted power function is shifted downwards by 0.8 ms (the synaptic gap delay used by Eggermont, 1979) the Don-data come close to the Eggermont data within in the frequency range 500–8000 Hz.
(4) For de Boer's cochlea model data (de Boer, 1980; Greenwood, 1990), the approximate formula, $\tau=k(f+a)^{-d}$, developed by Fobel and Dau (2004), uses the following constants: $k=4.78$, $a=165.4$, and $d=1.1$.

TABLE I. Derived-band ABR center frequencies and the corresponding observed wave V latencies for the Don-data ($N=39$). The latencies corrected by 4.96 ms are shown in the last column, which also indicates that only the corrected latencies for the four lower frequency bands are used for the construction of the final latency-frequency function in Fig. 1.

Band center frequency (Hz)	Observed mean values (ms)	Observed mean values −4.96 ms (ms)
Unmasked	7.10	2.15
11 300	7.06	2.10
5 700	7.15	**2.19**
2 800	7.78	**2.82**
1 400	8.75	**3.79**
700	10.44	**5.48**

The resulting latency-frequency functions are plotted in Fig. 1 and shown together with the obtained fitting formula and constants. For the Don latency-frequency function the four, adjusted, observed latency values are also plotted.

B. Modification of the phase response of the stimulus

A linear system is fully described by its *amplitude response*, $A(\omega)$, and *phase response*, $\theta(\omega)$ or by the system's

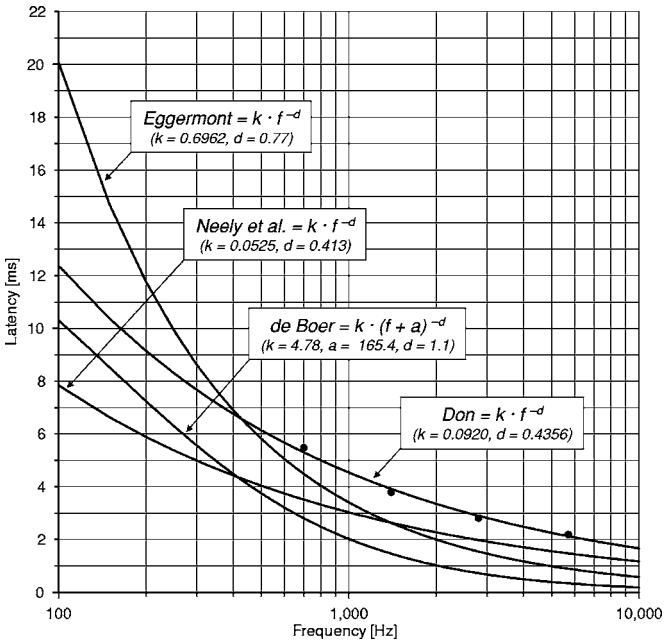


FIG. 1. Latency-frequency functions extracted from: (1) narrow-band ACAP (Eggermont, 1979), (2) tone-burst ABR (Neely *et al.*, 1988), (3) narrow-band ABR (Don-data, Don *et al.*, 2005a), and a cochlear model (de Boer, 1980). On each latency-frequency function the fitting formula and the retrieved parameter values are indicated. Also included are the four data points (Table I) that are used for the curve fitting of the Don-data.

impulse response, $h(t)$. (Here ω is the radian frequency $=2\pi f$, where f indicates frequency, and t indicates time.)

The *phase delay*, t_p, is expressed by the phase response by

$$t_p = -\frac{\theta}{\omega} \tag{2}$$

and describes how much a cosine of a specific frequency is delayed from the input to the output of the system.

The *group delay*, t_g, is defined as the negative slope of the phase response function:

$$t_g = -\frac{d\theta}{d\omega} \tag{3}$$

and describes how much the envelope of a signal formed by a narrow group of frequencies—as, for instance, an amplitude-modulated carrier frequency—is delayed through the system.

The electrophysiological latency-frequency functions included in the present description are all observed experimentally and are used as indirect observations of the traveling time. These data therefore constitute an outside model of the cochlea. The observed electrophysiological responses—either narrow-band ACAP or ABR or tone-burst ABR—can be regarded as being obtained by stimulation with a (narrow) group of frequencies. Pragmatically, we will assume that these latency-frequency functions represent the group delay of a linear system that modifies the phase response of the applied stimulus. The delay function retrieved from the traveling wave velocity in de Boer's model is based on direct measures and observations of wave patterns in the cochleae of humans and experimental animals and the constraints given by the chosen model approach. de Boer's delay function is therefore a result of an inside model. From de Boer's model, Dau *et al.* (2000) extracted the propagation time needed to arrive at the place of resonance and used this as a basis for defining the travel time. In the papers by Dau and colleagues (Dau *et al.*, 2000; Wegner and Dau, 2002) the retrieved delay function is also regarded as being the group delay.

Thus the corresponding phase delay can then be calculated by integration using Eqs. (2) and (3). The latency-frequency functions for the data by Eggermont, Neely *et al.*, and Don are all described by the same expression: $\tau = kf^{-d}$. This gives the following equations for the group delay, t_g, the phase response, θ, and the phase delay, t_p:

$$t_g = kf^{-d} = c\omega^{-d}, \tag{4}$$

$$\theta = -\int t_g d\omega + g = -c\frac{\omega^{1-d}}{(1-d)} + g$$

$$= \frac{-t_g\omega}{(1-d)} + g = -2\pi\frac{k}{(1-d)}f^{1-d}. \tag{5}$$

[If the phase response has to be realized in a real system with a real impulse response, the phase function, $\theta(\omega)$, has to be an *odd* function i.e., $\theta(-\omega) = -\theta(\omega)$. Therefore $\theta = 0$ for $\omega = 0$ and thus the integration constant, $g = 0$],

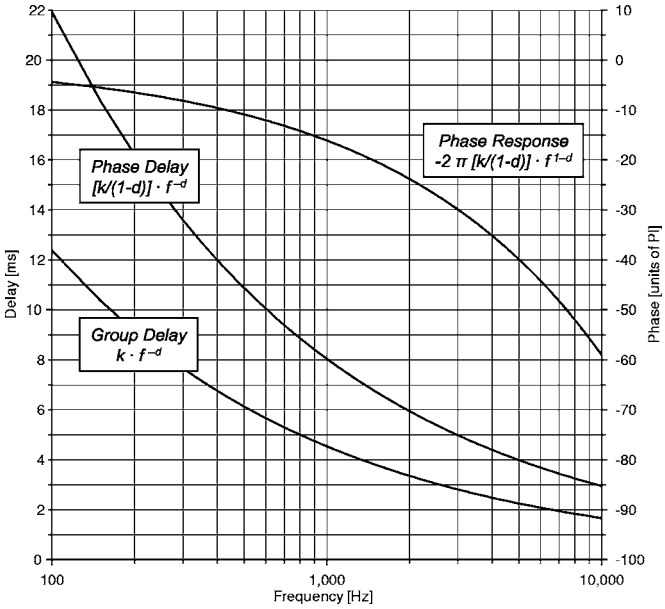


FIG. 2. The Don latency-frequency function (Fig. 1) applied to a linear system. The calculated group delay, phase delay, and phase response are shown together with the corresponding equations.

$$t_p = -\frac{\theta}{\omega} = \frac{k}{(1-d)}f^{-d}. \tag{6}$$

The latency-frequency function for de Boer's model data is described by $\tau = k(f+a)^{-d}$. The same steps as noted earlier give the following equations for the group delay, the phase response, and the phase delay:

$$t_g = k(f+a)^{-d}, \tag{7}$$

$$\theta = -2\pi\frac{k}{(1-d)}[(f+a)^{1-d} - a^{1-d}], \tag{8}$$

$$t_p = \frac{k[(f+a)^{1-d} - a^{1-d}]}{f(1-d)}. \tag{9}$$

For the Don-data, the group delay, phase response, and phase delay are plotted in Fig. 2, using the parameter values from Sec. II A: $k = 0.0920$ and $d = 0.4356$.

C. Formation of a chirp

From Fourier analysis theory we know that any (brief) stimulus which is repeated with a specified repetition rate (i.e., number of stimuli per second) can be described by a number of pure tones (cosines) having fixed frequencies corresponding to whole number multiples of the stimulus repetition rate (Stürzebecher *et al.*, 2006). For a broad-band stimulus the cosine frequencies will cover a large bandwidth (e.g., from 50 to 10 000 Hz). Correspondingly, as an example, a click stimulus with a repetition rate of 20/s (interstimulus-interval =50 ms) and covering the approximate frequency range 100–8000 Hz, can be constructed by adding together a harmonic series of cosines with the frequencies 100,120,140,…7980,8000 Hz all having equal amplitude and a zero phase delay at time=0.

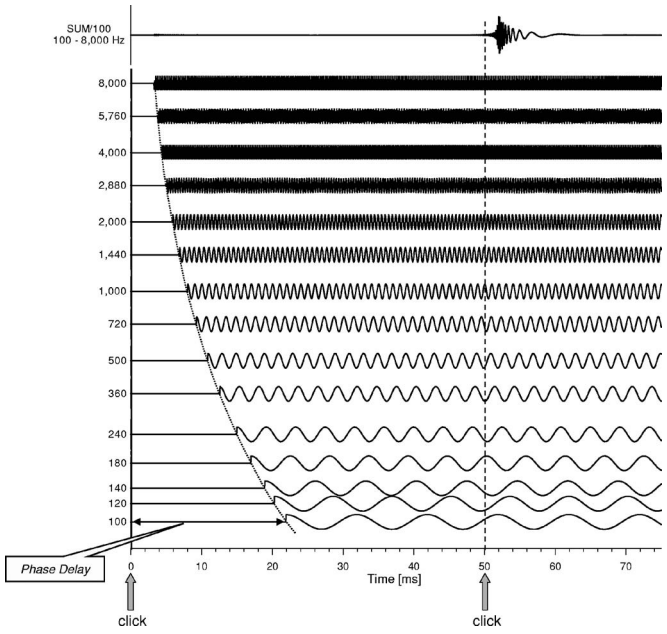


FIG. 3. Demonstration of how a broadband stimulus (100 Hz→8000 Hz) with a repetition rate of 20/s is obtained by taking the sum of a harmonic series of cosines with the frequencies 100,120,140,…,7980,8000 Hz. (From the harmonic series are shown the first three cosines followed by the cosines that are spaced by approximately half an octave). All cosines have equal amplitude and are given a phase delay in accordance with the values in Fig. 2 (Don-data).

Figure 3 demonstrates how the broadband stimulus is constructed if all the cosines are given a phase delay corresponding to the values in Fig. 2 (Don-data). The resulting broadband stimulus clearly presents the effect of temporal dispersion of the different spectral parts of the stimulus. The broadband stimulus is shown in more detail as the modified click in Fig. 4, and can be regarded as an approximation to the impulse response in the frequency range 100–8000 Hz of a linear system characterized by the parameters in Fig. 2. The stimulus clearly starts with the high frequencies (8000 Hz) and ends with the low frequencies (100 Hz). If the linear system did not introduce a phase delay (phase delay=0) then the broadband stimulus would resemble the original click, as shown as the click in Fig. 4. However, in order to compensate for the temporal dispersion, the negative values of the actual phase delay should be applied to each cosine of the click stimulus. This would result in a stimulus wave form shown as the chirp in Fig. 4.

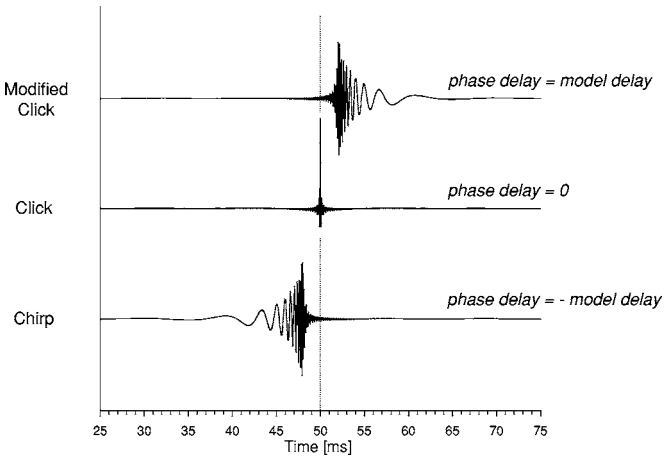


FIG. 4. Broad band stimulus (100 Hz→8000 Hz) with a repetition rate of 20/s obtained by adding together a harmonic series of cosines with the frequencies 100,120,140,…,7980,8000 Hz all with equal amplitude. Three different sets of phase delay values are used: (top, "modified click") Phase delay=model delay (Fig. 2); (middle, "click") phase delay =0; and (bottom, "chirp") phase delay=negative value of the model delay (Fig. 2).

D. Designing chirp stimuli in the frequency domain

Design in the time domain is normally preferred for the construction of traditional auditory stimuli for electrophysiological testing, when the *temporal* characteristics are most important—for instance, for generating stimuli for transient evoked responses.

Similarly, design in the frequency domain may be preferred when the *spectral* characteristics of the stimulus are most important—as, for instance, the generation of frequency-specific stimuli for steady-state evoked responses. By adjusting the amplitude and/or the phase of each cosine, the frequency spectrum of the final acoustic stimulus can be varied to produce, for instance, a flat, a rising, or a falling amplitude spectrum (e.g., a stimulus with equal energy per 1/3 octave), to compensate for the frequency characteristics of the earphone or other parts the sound transmission system or to design frequency-specific stimuli. By adjusting the phase of each cosine by the negative phase delay of the cochlea traveling delay, a chirp can be constructed that will attempt to compensate for the cochlea delay as shown by the example in the lower curve of Fig. 4.

For the study of the acoustic reflex threshold in both normal-hearing subjects and subjects with sensori-neural hearing loss, Müller-Wehlau *et al.* (2005) also applied a similar frequency-domain design of tone complexes with different phase characteristics.

E. Three different chirps

For the clinical experiments reported in Sec. III three different chirps are chosen. One is based on latencies obtained from narrow-band ABRs described by the Don-data, one is based on latencies from tone burst ABRs by Neely *et al.* (1988), and one is based on the cochlear model by de Boer (1980). The three chirps are referred to as the "*Don chirp*," the "*Neely et al. chirp*," and the "*de Boer chirp*."

The three chirps are designed using the traveling wave delays described earlier, which represent different group delay estimates. The Don-data were obtained with a click level of 60 dB nHL (approximately 101 dB p-p.e. SPL),[2] whereas the Neely *et al.* data provide different latency values depending on stimulus level; however, in the present experiments only the latency-frequency function corresponding to a tone burst level of 90 dB SPL (approximately 65 dB nHL) has been used.[3] de Boer's latency data are not level specific, but most likely represent the cochlear parameters at relatively high levels of stimulation. The underlying estimates of the traveling wave delay correspond to those shown in Fig. 1 (the Eggermont latency-frequency function is not used in the next part of the present study).

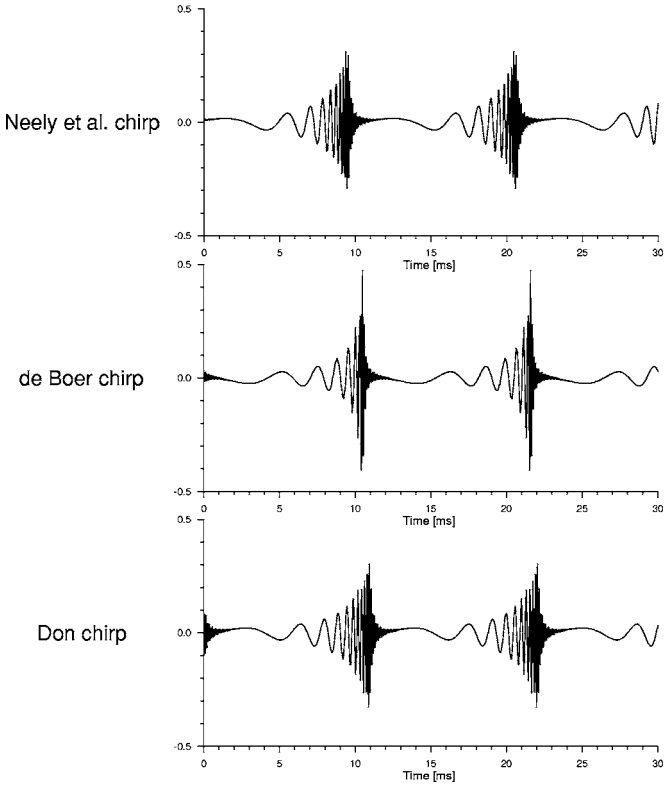


FIG. 5. Three different chirps calculated similarly to the broadband chirp in the bottom curve of Fig. 4, but based on a repetition rate of 90/s. The broadband covers the frequency range 180 Hz→7920 Hz. From the original latency-frequency functions by Neely *et al.* (1988), de Boer (1980), and the Don-data shown in Fig. 1, the negative values of the corresponding phase delays are used for the calculations. The chirps are called: Neely *et al.* chirp, de Boer chirp, and Don chirp.

Since the three chirps are to be tested in a clinical ASSR experiment using a stimulus rate of 90/s (see Sec. III) and an approximate frequency range 200–8000 Hz, they are all constructed by adding together a harmonic series of cosines with the frequencies 180,270,360,…,7830,7920 Hz, and using the phase delays corresponding to the three original group delay functions. Thus, the resulting chirp stimuli are broadband stimuli with flat amplitude spectra spanning the frequency range from about 200 to 8000 Hz and having an interstimulus-interval of 11.1 ms (the inverse of the stimulus rate =90/s). The final chirp wave forms are presented in Fig. 5.

F. Discussion

As stated previously, Neely *et al.* (1988) provide latency values as a function of tone burst level given in dB SPL. Gorga *et al.* (1988) suggested using a conversion factor of 25 dB in order to compare their results to frequency-specific ABR-latency values obtained by others using stimulus levels in dB HL or dB SL. Therefore, on average across frequencies, the stimulus levels in dB SPL in the Neely *et al.* data when compared to corresponding levels in dB nHL are about 25 dB higher. Thus 90 dB SPL would be approximately equivalent to 65 dB nHL. This means that the latency data used herein by Neely *et al.* (1988) and the Don-data were obtained at comparable levels above the psychoacoustic threshold in subjects with normal hearing (60–65 dB nHL). However, there is one difference that may play an important role in this comparison: The Neely *et al.* (1988) data are obtained with frequency-specific tone bursts, which only excite a limited part of the cochlea, whereas the Don-data are obtained with broadband clicks that excite most of the cochlear partition. Therefore, the Neely *et al.* (1988) data most likely represent a higher effective stimulus level in each frequency band than the corresponding Don-data. The equivalent rectangular bandwidth of the tone bursts used in the Neely *et al.* study can be calculated from the specifications given by Gorga *et al.* (1988, Table I) and varies as a function of tone burst frequency. For the frequencies 250, 1000, and 4000 Hz the bandwidth corresponds to approximately 160, 320, and 640 Hz. The click distributes its energy across a broad frequency band and by means of a filtering program we have calculated the (electrical) peak-to-peak amplitude and the rms amplitude of the click within each of the three above-mentioned frequency bands. The peak-to-peak amplitude varies from −18 to −30 dB (mean=−24 dB) and the rms amplitude varies from −11 to −17 dB (mean=−14 dB) relative to the corresponding broadband values. This leads to the following estimation: For each band the tone burst level is 90 dB SPL (65 dB nHL), whereas the click level is about 77 dB p-p.e. SPL (46 dB nHL) (101−24=77;60−14=46). Therefore the tone burst data by Neely *et al.* (1988) probably reflect the cochlear behavior at a level which is about 10–20 dB higher than the Don-data. Another difference is related to the upward spread of excitation provided by the tone bursts at higher levels of stimulation. Thus the neural activity is less restricted than suggested by the tone burst spectrum and the amount is level dependent. This would add to the shorter latencies that are observed in the Neely *et al.* (1988) data at higher levels of stimulation.

Based on the tone burst ABRs, Neely *et al.* (1988) and Gorga *et al.* (1988) formulated a set of latency-frequency functions which changed significantly with stimulus level in the range from 20 to 100 dB SPL. The latency-frequency functions changed in absolute and relative latency corresponding to changes both in latency offset and in form, i.e., the higher the stimulus level the shorter the latency and the less curvature of the latency-frequency function. When constructing a chirp, a difference in latency offset will shift the location of the chirp on the time axis, whereas differences in curvature of the latency-frequency function will change the wave form of the chirps as seen in Fig. 5. Therefore chirps constructed from the different latency-frequency functions by Neely *et al.* (1988) will be quite different in their form. It is therefore interesting to observe that with the click-evoked, narrow-band ABRs, Don *et al.* (1998) did not find any significant level dependence of the curvature of the latency-frequency function (i.e., difference in latency between 5.7 and 1.4 kHz) across the range of click levels from 93 to 53 dB p.e. SPL (60 to 20 dB nHL with the TDH-50P-earphone) in normal-hearing subjects. Contrary to what was found for the Neely *et al.* (1988) data, this has the consequence that when the narrow-band latency data are used for constructing a chirp, the temporal wave form of the chirp will be constant across levels.

Conceptually there is no difference between the de Boer chirp reported here and the flat-spectrum chirp by Dau *et al.* (2000). However, the two different ways used to design the stimuli give rise to at least subtle differences. The de Boer chirp, as well as the other two chirps, is designed in the frequency domain. Therefore, they are specified accurately in the spectral domain but not in the temporal domain. This can, for instance, be observed by the extended trailing edge of our chirps (Fig. 5). Contrary to this, the Dau *et al.* chirp is designed in the temporal domain by calculating the phase of the instantaneous frequency by integration from time=0 to t. This chirp is therefore specified accurately in the temporal domain but not in the spectral domain. The abrupt ending of the temporal wave form of the chirp will, for instance, affect its frequency spectrum (frequency splatter). Dau *et al.* (2000) do not provide details of how the flat-spectrum-chirp is generated from the exact chirp in order to change its amplitude spectrum from a falling to a flat spectrum. However, in the paper by Fobel and Dau (2004) it appears that changes of the amplitude spectrum are introduced by multiplying the instantaneous phase function by a time-variant amplitude function; this will affect the resulting phase spectrum of the chirp. It is not clear to us whether these differences between our de Boer chirp and the flat-spectrum-chirp by Dau *et al.* (2000) will have any practical consequence.

III. ASSR TESTING

A. The ASSR

The ASSR is an electrophysiological response that is evoked by a periodically repeated auditory stimulus. The response is stable over time for as long as the stimulation is turned on. If the recording time window can be regarded as being infinitely long, then the response will consist of a series of discrete frequency components that are constant in both amplitude and phase over the time window, e.g., Picton *et al.* (2003).

Auditory steady-state responses are to be distinguished from auditory transient responses. According to the above, ASSRs are evoked by a series of sound stimuli that are presented at a high repetition rate (greater than 60–70/s) or by amplitude and/or frequency modulated carriers, whereas auditory transient responses are evoked by an individual, brief sound stimulus or a series of brief sound stimuli that are presented at a low repetition rate (lower than 20–30/s). At medium repetition rates there is a grey zone where the differentiation between the two response types is difficult to make. However, if the stimulus repetition rate is so high that the electrophysiological response to one stimulus overlaps with the response to the next stimulus, then the recorded activity can be meaningfully classified as a steady-state response. At repetition rates close to 40 stimuli/s, components of the middle latency response overlap and the recorded ASSR is named the 40-Hz response (Galambos *et al.*, 1981). Although the ABR often can be identified in this response, it is dominated by evoked activity from the higher auditory pathways in the thalamus (middle brain) and the cortex and is therefore influenced by test subject conditions such as attention, arousal, and sleep (anesthesia and sedation). At repetition rates higher than 60–70 stimuli/s, components of the ABR begin to overlap. This ASSR is dominated by early evoked activity from the brain stem and is therefore not significantly influenced by the above-mentioned test subject conditions, e.g., Picton *et al.* (2003).

Previously we have applied different ASSR test paradigms using stimulus rates around 90/s in order to evaluate the efficiency of a series of objective response detection algorithms, statistical test strategies and novel stimuli (Cebulla *et al.*, 2006; Stürzebecher *et al.*, 2005, 2006). The methods evaluated in these publications were used in the experiment described in the following.

A total of nine different stimuli were presented at a high stimulus rate in an experiment where the ASSR was recorded at two presentation levels in a group of normal-hearing subjects. The presence of the ASSR was detected objectively by a detection algorithm (see Sec. III D) and the response characteristics were extracted. The efficiency of each stimulus condition was evaluated by the observed detection time, i.e., how fast the corresponding ASSR could be detected by a specific detection algorithm, and the signal-to-noise ratio (SNR) of the ASSR. However, the present report deals only with the comparison of the three chirps described in Sec. II and a click stimulus.

B. Normal-hearing test subjects

The test group consisted of $N=49$ young adult subjects (age: 17–34 years) with normal hearing. The pure tone thresholds were 10 dB HL or better at the frequencies of 500, 1000, 2000, and 4000 Hz. Each subject was tested only in one ear that was chosen randomly. Due to the size of the total experiment, as explained earlier, it was not possible to test all test subjects with all stimuli; therefore, for each stimulus condition the effective number of test subjects was about 15% lower than 49. The experiment was carried out with a general approval by the Institutional Review Board of the Julius Maximilians-University in Würzburg. Prior to the testing, the purpose and procedure of the study were orally presented to the test subjects and any questions the subject had regarding his/her participation were explained.

C. Stimuli and calibration

Calibration of the three chirps and a click (with the same amplitude spectrum as the chirps) was obtained from a subgroup of $N=10$ subjects with normal hearing. The stimuli were delivered by a set of EAR-3A earphones and presented in a random order across the group of test subjects at a rate of 90/s. For each stimulus, the individual, subjective thresholds were established using a bracketing method (ISO 8253-1, 1989) with a step size of 1 dB, and the mean threshold levels across the group were then computed. The levels of the four stimuli were finally expressed in dB nHL.

As stated in Sec. I, our general goals are to find ways to improve ASSR detection at low SNRs for newborn hearing screening and therefore the four stimuli were evaluated at two relatively low levels viz. 30 and 50 dB nHL. In the ASSR experiment the stimulus conditions were presented in a random order across the group of subjects.

D. Recording

Details of the recording setup and procedures can be found in Stürzebecher *et al.* (2006) and Cebulla *et al.* (2006). However, a brief summary is given in the following. The test subjects were placed on a comfortable couch in a sound treated room and were instructed to relax and, if possible, to sleep during the ASSR testing. The two active electrodes were placed—one at the vertex (C_z) and the other at the ipsi-lateral mastoid—whereas the ground electrode was placed at the forehead. All recordings were made with the MB11-2 instrument (Maico Diagnostics GmbH), equipped with separate EAR-3A earphones and a preamplifier.

The EEG was bandpass filtered from 25 Hz (6 dB/octave) to 1.5 kHz (24 dB/octave) and during the recording the EEG was continuously stored on hard disc for later retrieval and analysis. An artifact rejection level of ±20 μV was applied as a standard; however, in a few cases it was necessary to increase the level to ±40 μV. During the recording sessions the EEG was transformed to the frequency domain by means of a Fast fourier transformation (FFT) and to detect the ASSR a modified Rayleigh test (Mardia, 1972) was used. Originally this test uses only the spectral phases but the modification described by Cebulla *et al.* (2006) includes both spectral amplitudes and phases. In the present study, this later modification was applied to the first harmonic of the response using a sequential test strategy and an error probability of α=0.1% controlled for the effect of repeated testing (Stürzebecher *et al.*, 2005). The individual recording was continued for at least 100 s but never longer than 300 s. If no ASSR was detected by 100 s, the recording was continued until detection was made and then terminated approximately 10 s later.

Normally, ASSR is detected in the frequency domain and therefore the averaged temporal wave form of the response is not readily available. However, from the resulting complex frequency spectrum the harmonic response frequencies can be isolated (here multiples of the stimulus rate, i.e., 90,180,270… Hz) and by an inverse FFT an estimate of the temporal wave form can be obtained.

E. Description of the results

The number of test subjects that actually were tested in each stimulus condition is shown in Table II.

The *detection rate* (%) for each stimulus condition is computed as the ratio between the number of test subjects in which the ASSR was detected within 300 s divided by the total number of test subjects that were tested in that condition. In Table II the detection rates are given both as percentages and as fractions.

The *detection time* (s) for each stimulus condition and for each test subject is calculated. Due to the detection algorithm design the lowest possible detection time is 10 s, and due to the test protocol the largest detection time is 300 s. For each stimulus condition the *median* detection time and the corresponding 5 and 95 *percentiles* are given in Table II.

The SNR for each stimulus condition is estimated for each test subject. Regardless of whether detection was obtained or not, the resulting frequency spectra after 100 s were used to calculate the SNR of the first response harmonics. The SNR was defined and estimated as follows:

$$\mathrm{SNR} = \sqrt{\frac{S^2}{\bar{N}^2} - 1},$$

where S is the rms amplitude of the first response harmonics in the frequency spectrum, and $\bar{N}$ is the rms amplitude of the corresponding averaged background noise at the first response harmonics, estimated from 60 frequency bins, i.e., 30 bins taken from each side of the bin that corresponds to the first response harmonics. For each stimulus condition the *median* SNR and the corresponding 5 and 95 *percentiles* are given in Table II.

By inverse FFT an estimate of the wave form is obtained for each ASSR. However, since each recording was terminated either after 100 s or shortly after a positive detection was made ($p<0.001$) no clear definition of the wave forms is necessarily available. Therefore the estimated wave forms are averaged across subjects and these *grand average wave forms* are presented in Figs. 6(a) and 6(b). The peak-to-peak amplitude of each grand average is calculated and displayed in the two figures, and all eight amplitude values are also plotted as input-output functions in Fig. 7.

TABLE II. For each stimulus condition (stimulus type and level), the number of subjects tested, the detection rate (in % and fractions), the detection time within 300 s (median and 5%–95% percentile), and the SNR at a recording time of 100 s (median and 5%–95% percentile) are shown.

	Detection rate (%)	Detection time (s)	SNR
	30 dB nHL		
Click	**83.3**	**72**	**3.1**
N=42	35/42	23–230	0.2–8.4
Neely *et al.*	**95.3**	**41**	**5.0**
N=43	41/43	17–107	1.2–14.6
de Boer	**97.6**	**32**	**5.2**
N=41	40/41	14–187	2.2–12.5
Don	**97.7**	**30**	**6.0**
N=43	42/43	14–127	1.5–17.8
	50 dB nHL		
Click	**85.7**	**40**	**5.2**
N=42	36/42	16–207	0.8–10.7
Neely *et al.*	**97.7**	**27**	**6.6**
N=43	42/43	13–107	1.9–16.7
de Boer	**100**	**27**	**6.6**
N=43	43/43	13–69	3.0–15.4
Don	**100**	**23**	**7.4**
N=43	43/43	13–55	3.2–15.3

F. Analysis of the results

Statistical analysis is performed on the obtained results (detection time and SNR) using a nonparametric paired comparison test, i.e., *Wilcoxon Matched-Pairs Signed-Ranks test*, Siegel (1956). The number of test subjects that contribute to each paired comparison testing may be different from the number used for the descriptive statistics in Table II. For the statistical testing, the unknown values of the detection time (for those recordings where no response was detected within 300 s) are (conservatively) assigned a value of 300 s.

1. Comparison between chirp and click

At both stimulus levels all three chirps produce detection times that are significantly shorter and SNRs that are significantly larger than those produced by the click ($p < 0.001$).

Between the detection times and the SNRs obtained by the click at 50 dB nHL and the detection times and the SNR obtained by each chirp at 30 dB nHL, significant differences are found ($p < 0.05$), i.e., each chirp results in lower detection times and larger SNRs at 30 dB nHL than the click at 50 dB nHL.

2. Comparison between the chirps

At both stimulus levels no significant differences are found between the detection times or the SNRs obtained by the different chirps.

A new variable is calculated by combining the detection time and SNR.[4] At 50 dB nHL this variable (=detection time/SNR) is found to be significantly smaller ($p < 0.05$) for the Don chirp than for the de Boer and Neely *et al.* chirps, whereas no significant difference between de Boer chirp and Neely *et al.* chirp is found.

G. Discussion

As expected, the chirps are more efficient than the click at both levels of stimulation. This indicates that a more synchronous excitation of the cochlea and the peripheral part of auditory nerve fibers has the same positive effect on the ASSR amplitude as observed previously by other investigators for the ACAP and the ABR.

The extracted parameters for each chirp—the detection rate, the median detection time, the median SNR, and the peak-to-peak amplitude of the grand average waveforms—provide a systematic rank-order of the three chirps (1) Don, (2) de Boer, and (3) Neely *et al.*. However, this order of performance can only be confirmed statistically at 50 dB nHL where the Don chirp is found to be significantly more efficient than the other two chirps.

In general, our findings are in accordance with the findings by Fobel and Dau (2004) although our experiment is limited to a stimulus range of 20 dB, whereas Fobel and Dau (2004) used a range of 50 dB. Another difference between the two studies is that Fobel and Dau (2004) used the level-dependent latency information to construct their A chirps (Neely *et al.* chirps) for the comparison with the M chirp (the flat-spectrum de Boer chirp) and a 100-μs click. However, when converting the tone burst levels from dB SPL to dB SL, Fobel and Dau (2004) subtracted 40 dB from the dB SPLs given by Neely *et al.* (1988); this conversion factor is different than the one suggested by Gorga *et al.* (1988) and the one we used in our discussion in Sec. II F, but was based by Fobel and Dau (2004) on a comparison with the M chirp. This means that the latency information that underlies the A chirp at 50 dB SL in the study by Fobel and Dau (2004) is the same as we have used to construct our Neely *et al.* chirp, i.e., the latency-frequency function at 90 dB SPL.

At lower levels of stimulation, i.e., 10–20 dB SL, Fobel and Dau (2004) found that the A chirp (level-adjusted Neely *et al.* chirp) produces significantly higher ABR amplitudes than the M chirp (de Boer chirp), whereas at higher levels of stimulation no amplitude differences were found. As argued by Fobel and Dau (2004) the difference they observed in the ABR amplitude at lower levels was probably due to the marked wave form difference between the A chirp and M chirp—a difference which becomes more and more prominent the lower the level used for constructing the A chirp. Since we have not tested at levels lower than 30 dB nHL(SL), we are not able to confirm this finding. In the present study no level adjustments of the Neely *et al.* chirp were included because the temporal wave form of the A chirp used by Fobel and Dau at 30 dB nHL deviates only slightly from the wave form of the Don chirp. Thus at 50 dB nHL the Neely *et al.* chirp used herein is comparable to the A chirp used by Fobel and Dau, whereas at 30 dB nHL the Don chirp is comparable to the A chirp. The result of the present study where no statistical difference is found at 30 dB nHL between the ASSR in response to the de Boer and the Don chirp is therefore in line with the findings by Fobel and Dau (2004).

As found for the ABR by Fobel and Dau (2004), we find that the chirps increase the ASSR amplitudes by a factor of about 2–2.5, compared to the amplitudes produced by the click at 30 and 50 dB nHL. As expected, all our ASSR amplitudes are somewhat smaller (about 1.5 times) than the corresponding ABR amplitudes obtained by Fobel and Dau (2004); this is partly due to (1) the use of the amplitudes of the grand average ASSR wave form instead of the average amplitude of the individual recordings and (2) an effect of neural adaptation due to the significantly higher stimulus rate used for the recording of the ASSR than for the ABR (90/s vs 20/s).

Previous publications have shown that a chirp evokes a stronger PAMR than a corresponding click (e.g., Agung *et al.*, 2005; Dau *et al.*, 2000). In ASSR recordings, which are analyzed in the frequency domain, the response wave forms are normally not inspected or available and a response is defined only by the spectral energy at the harmonic frequencies of the stimulus rate. Especially at higher stimulus levels, the chirp ASSR could therefore easily contain strong myogenic components without being apparent to the tester. However, in the present experiment the stimulus level is restricted to 30 and 50 dB nHL, and we do not see any trace of the PAMR in the grand average response wave forms. These wave forms look similar to classical ABR waveforms—where one ABR partly overlaps the next one that follows in rapid succession.

ASSRs to stimulus repetition rates greater than 60–70/s are conveniently analyzed in the frequency domain, where the response contains significant energy at the first six to eight harmonics of the stimulus rate (e.g., Cebulla *et al.*, 2006). For hearing screening or hearing diagnostics this analysis technique makes it possible to record the ASSR from the right and left ear simultaneously. By using slightly different repetition rates for the two ears, the left and right ASSRs will be located at two spectrally different harmonic patterns and can therefore be recovered and analyzed independently. Based on the same principle, simultaneous stimu-

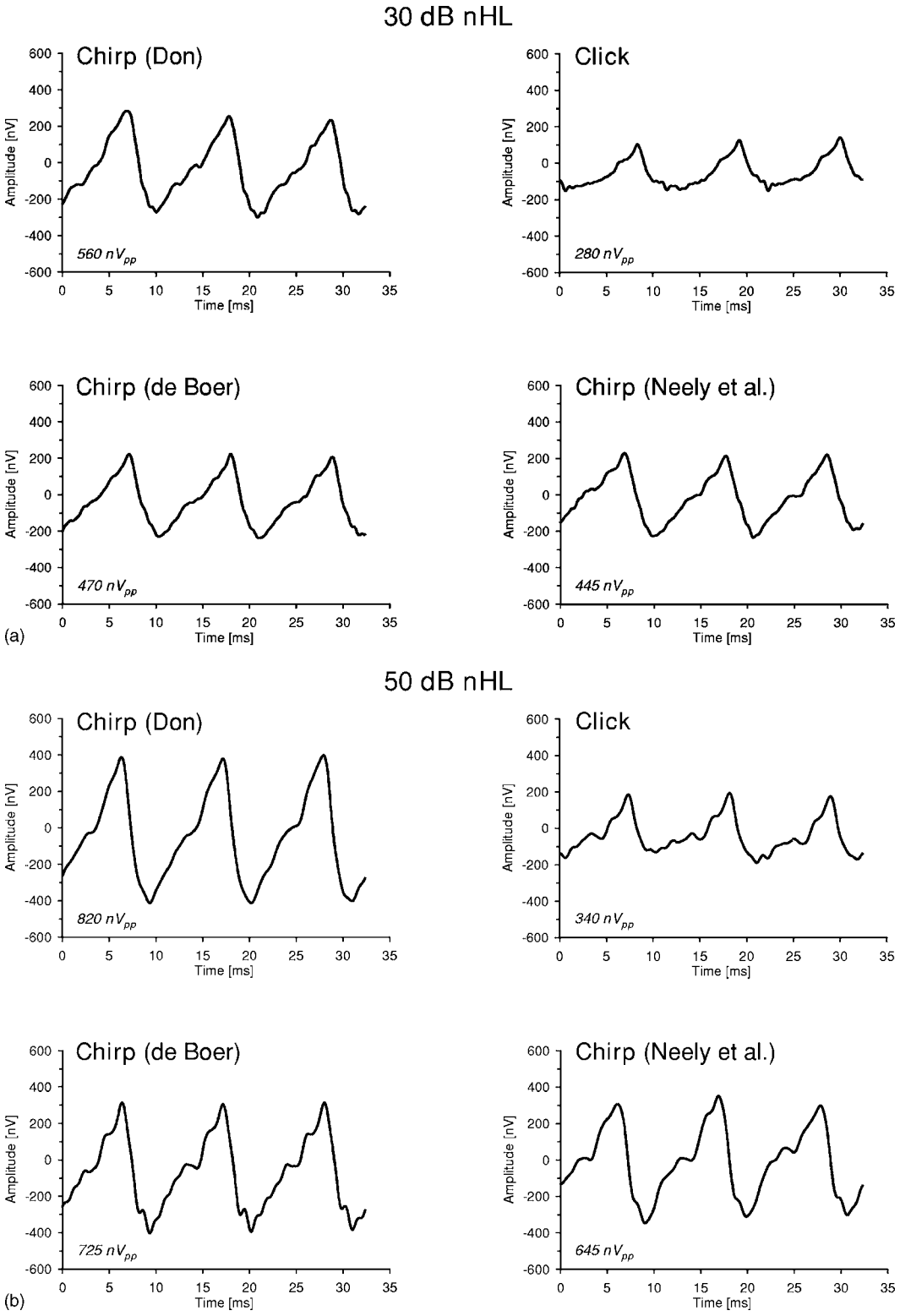


FIG. 6. (a) Grand average ASSR wave forms at 30 dB nHL for the three chirps and the reference click stimulus after a recording time of 100 s. The calculated peak-to-peak amplitudes are shown for each wave form. (b) Grand average ASSR wave forms at 50 dB nHL for the three chirps and the reference click stimulus after a recording time of 100 s. The calculated peak-to-peak amplitudes are shown for each wave form.

lation on the same ear is also possible using frequency-specific stimuli presented at slightly different repetitions rates. The application of multiple simultaneous stimuli may significantly reduce test time when the ASSR is used for hearing evaluation. However, when simultaneous stimulation is applied to the same ear, masking and interaction between the stimuli may take place and under some conditions affect the efficacy and accuracy of the threshold evaluation (John *et al.*, 2002).

From the experimental results obtained here, the efficiency of the present ASSR recording technique can be described as follows: At 30 dB nHL, Table II gives the median

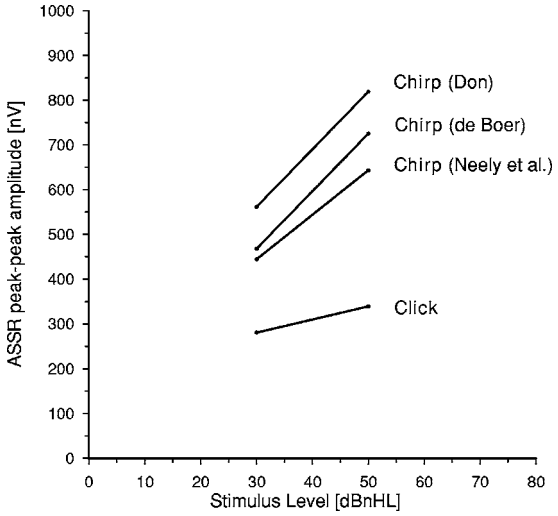


FIG. 7. Peak-to-peak amplitudes of the grand average ASSR wave forms [Figs. 6(a) and 6(b)] plotted vs stimulus level.

detection times for the click: 72 s and for the Don chirp: 30 s. Compared to a standard ABR recording setup using a stimulation rate of 20/s these numbers translate to 1440 sweeps and 600 sweeps, respectively. This would mean that in 50% of normal-hearing test subjects a click response would be detected within 1440 sweeps and a chirp response within only 600 sweeps with a probability of only 0.1% that a response was falsely detected.

The ASSR detection algorithm used in the present study is a modified Rayleigh test (Cebulla *et al.*, 2006). It is applied as a one-sample test, which uses only the first harmonic of the stimulus rate for the detection (i.e., 90 Hz) and does not include a higher number of harmonic frequencies (up to six) as done in some of the *q*-sample tests reported previously by Cebulla *et al.* (2006). Therefore only the energy of the lowest frequency part of the ASSR is utilized for the response detection as well as to calculate the SNR. The lowest detection time is limited to 10 s and is most likely the reason why we observe that the inverse of the response detection time is approximately proportional to $SNR^{1.5}$, rather than to SNR^2, which would be expected if the detection algorithm used an estimate of the SNR.[5]

The three chirps and the click used in the present experiment have the same bandwidths and amplitude spectra and are presented at equal hearing levels (dB nHL). If, however, the goal is to evaluate the electrophysiological effect of compensating for the traveling wave delay, the stimuli should be presented with the same spectrum levels (power per Hz-bandwidth). This would ensure that the differences between the stimuli would be limited to differences in the temporal distribution of power in accordance with the latency model applied. In a preliminary experiment, the Don chirp and the click (corresponding to the lower two wave forms in Fig. 4) were presented with equal spectrum levels from an EAR-3A earphone at a rate of 90/s to 10 normal-hearing test subjects. The psychoacoustic thresholds were obtained by a bracketing method (ISO 8253-1, 1989) using a step size of 1 dB and were found to differ by less than 0.5 dB ($p<0.05$). Thus, it seems reasonable to conclude that when the two stimuli are presented at equal spectrum levels they are also presented at equal hearing levels. Therefore, our main results directly reflect the effect of the applied compensation strategy.

IV. SUMMARY AND CONCLUSION

In the literature it has been shown that both input compensation of the traveling wave delay using chirp stimuli and output compensation by the method of the Stacked ABR increase the amplitude of early electrophysiological responses (ACAP, ABR, and ASSR) by increasing the temporal synchronization between the contributing neural elements in the auditory periphery.

Chirp stimuli that are constructed to compensate for the traveling wave delay are based on a model delay function or a latency-frequency function that characterizes the cochlear transfer. Several latency-frequency functions in humans are described in the literature: Some of these functions are retrieved from electrophysiological experiments in normal hearing individuals and others are provided by model experiments based on observed response characteristics of the cochlea in human preparations and from experimental animals.

In the present study the latency-frequency functions were assumed to reflect the group delay of a linear system that modifies the phase spectrum of the applied stimulus. Three different estimates of the group delay, which differed in both latency offset and latency curvature, were used to construct three chirps. These differences affected both the exact temporal location as well as the wave form of the constructed chirps.

The three chirps were compared to a click stimulus in a clinical evaluation where the ASSR was recorded from 49 young normal-hearing individuals and where the stimuli were presented at two relatively low levels at a rate of 90/s. The ASSR was detected significantly faster to the three chirps than to the click. At 30 dB nHL the necessary recording time was approximately halved. The results also indicated significantly higher response SNRs and amplitudes evoked by the chirps than by the click. It was shown that for response detection all three chirps were more efficient than the click even when the chirps were presented at a level that was 20 dB below that of the click in normal-hearing adults.

All parameter values of the ASSRs to the different chirps appeared with the same rank order. Despite this apparent ranking, only at 50 dB nHL was it possible to demonstrate statistically the efficiency of the Don chirp over the two other chirps. Large within-subject and between-subject variance could be part of the reason for this lack of statistical differentiation.

The clinical results indicated that it would be advantageous to substitute the standard click, which normally is used in the recording of peripheral auditory evoked responses, i.e., ACAP, ABR, and ASSR, with an appropriately designed chirp.

ACKNOWLEDGMENTS

The authors want to thank Viola Hirschvogel and Martina Busch for their significant contribution to the management of the test subjects and to the collection of the electro-

physiological data. The authors also want to acknowledge Søren Laugesen, Ph.D., for his constructive criticism and support to the work presented herein.

[1]In this reference the high-pass masked ABRs were recorded, but the corresponding normative, narrow-band ABR latencies were not presented. In the present work these unpublished latency data are used and referred to as the Don-data.

[2]The Don-data (Don *et al.*, 2005a) are obtained using a 100 μs rarefaction click presented at a rate of 45/s from an ER-2 earphone. Richter and Fedtke (2005) give a reference zero calibration value of 43.2 dB p-p.e. SPL (their table 4) for a 100 μs *alternating* click delivered at a rate of 20/s from an ER-2 earphone and measured in an occluded ear simulator (Brüel & Kjær 4157). Correction for polarity equals +0.4 dB (rarefaction versus alternating, their Table 2) and correction for rate equals −2.2 dB (45/s vs 20/s, their Fig. 4), resulting in a reference zero value of 41.4 dB p-p.e. SPL for the stimulus condition for the Don-data. Thus the 60 dB nHL corresponds to 101.4 dB p-p.e. SPL in the occluded ear simulator (711-coupler).

[3]This latency-frequency function was used by Fobel and Dau (2004) to generate a chirp that was presented at 50 dB SL in normal-hearing subjects. In the present study two presentation levels are applied: 50 dB and 30 nHL. Therefore at 50 dB nHL the present results can be directly compared to the results from Fobel and Dau (2004). At 30 dB nHL Fobel and Dau (2004) used a Neely *et al.* chirp based on the latency-frequency function of 70 dB SPL. This latency-frequency function is almost the same as the Don latency-frequency function (difference less than 0.5 ms at 500 Hz) and the corresponding Neely *et al.* chirp can hardly be distinguished visually from the Don chirp in the frequency range 500–8000 Hz. (See discussion in Sec. III G.)

[4]The detection time and the SNR are both dependent on the ASSR and the background noise and in each recording condition they are both extracted from the recorded data. Although the response detection algorithm includes spectral amplitude it is predominantly dependent on the *phase* of the first response harmonics (Rayleigh test). In contrast to this, the SNR is dependent on the spectral *amplitude* of the first response harmonics and the mean spectral *amplitude* of the neighboring frequency bins after a fixed recording time (100 s). The detection time and the SNR across all subjects, stimuli, and levels are only weakly (negatively) correlated (r= −0.45), which indicates that they provide (at least partly) independent descriptions of the response quality. It is therefore hypothesized that a new metric, which combines detection time and SNR, may provide higher statistical leverage than either of them alone. Because higher stimulus efficiency results in *lower* detection time and *higher* SNR, the combination: detection time/SNR has been chosen. This metric characterizes the response quality in each recording condition.

[5]With stationary physiological input noise, the output noise after averaging is equal to the input noise divided by the square root of N (i.e., the number of averaged sweeps, which is proportional to the recording time—see, e.g., Don and Elberling, 1994). Thus in the individual recording the output SNR^2 is proportional to the input SNR^2 multiplied by the recording time. If the detection algorithm detects a response when a given SNR value has been reached then the detection times across individuals would be inversely proportional to the input SNR^2 or to the SNR^2 obtained after a specific recording time (e.g., 100 s).

Agung, K., Purdy, S. C., Patuzzi, R. B., O'Beirne, G. A., and Newall, P. (**2005**). "Rising frequency chirps and headphones with extended high-frequency response enhance the post-auricular muscle response," Int. J. Audiol. **44**, 631–636.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects," J. Acoust. Soc. Am. **49**, 1131–1139.

Bell, S. L., Allen, R., and Lutman, M. E. (**2002**). "An investigation of the use of band-limited chirp stimuli to obtain the auditory brainstem response," Int. J. Audiol. **41**, 271–278.

Cebulla, M., Stürzebecher, E., and Elberling, C. (**2006**). "Objective detection of auditory steady-state responses: Comparison of one-sample and q-sample tests," J. Am. Acad. Audiol **17**, 93–103.

Dau, T., Wagner, O., Mellert, V., and Kollmeier, B. (**2000**). "Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion," J. Acoust. Soc. Am. **107**, 1530–1540.

de Boer, E. (**1980**). "Auditory physics. Physical principles in hearing theory. I," Phys. Rep. **62**, 87–174.

Don, M., and Eggermont, J. J. (**1978**). "Analysis of click-evoked brainstem potentials in man using high-pass masking," J. Acoust. Soc. Am. **63**, 1084–1092.

Don, M., Eggermont, J. J., and Brackmann, D. E. (**1979**). "Reconstruction of the audiogram using brainstem responses and high-pass noise masking," Ann. Otol. Rhinol. Laryngol. **88**, 1–20.

Don, M., and Elberling, C. (**1994**). "Evaluating residual background noise in human auditory brain-stem responses," J. Acoust. Soc. Am. **96**, 2746–2757.

Don, M., Kwong, B., and Tanaka, C. (**2005a**). "A diagnostic test for Meniere's disease and cochlear hydrops: Impaired high-pass noise masking of auditory brainstem response," Otol. Neurotol. **26**, 711–722.

Don, M., Kwong, B., Tanaka, C., and Brackmann, D. (**2005b**). "The stacked ABR: A sensitive and specific screening tool for detecting small acoustic tumors," Audiol. Neuro-Otol. **10**, 274–290.

Don, M., Masuda, A., Nelson, R., and Brackmann, D. (**1997**). "Successful detection of small acoustic tumors using the stacked derived-band auditory brain stem response amplitude," Am. J. Otol. **18**, 608–621.

Don, M., Ponton, C. W., Eggermont, J. J., and Kwong, B. (**1998**). "The effects of sensory hearing loss on cochlear filter times estimated from auditory brainstem response latencies," J. Acoust. Soc. Am. **104**, 2280–2289.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1994**). "Auditory brainstem response (ABR) peak amplitude variability reflects individual differences in cochlear response times," J. Acoust. Soc. Am. **96**, 3476–3491.

Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I. Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.

Eggermont, J. J. (**1976**). "Analysis of compound action potential responses to tonebursts in the human and guinea pig cochlea," J. Acoust. Soc. Am. **60**, 1132–1139.

Eggermont, J. J. (**1979**). "Compound action potentials: Tuning curves and delay times," in *Models of the Auditory System and Related Signal Processing Techniques*, edited by M. Hoke and E. de Boer [Scand. Audiol. Suppl. **9**, 129–139].

Elberling, C. (**1974**). "Action potentials along the cochlear partition recorded from the ear canal in man," Scand. Audiol. **3**, 13–19.

Elberling, C., and Parbo, J. (**1987**). "Reference data for ABR's in retrocochlear diagnosis," Scand. Audiol. **16**, 49–55.

Fobel, O., and Dau, T. (**2004**). "Searching for the optimal stimulus eliciting auditory brainstem responses in humans," J. Acoust. Soc. Am. **116**, 2213–2222.

Galambos, R., Makeig, S., and Talmachoff, P. J. (**1981**). "A 40-Hz auditory potential recorded from the human scalp," Proc. Natl. Acad. Sci. U.S.A. **78**, 2643–2647.

Gorga, P. M., Kaminski, J. R., Beauchaine, K. A., and Jesteadt, W. (**1988**). "Auditory brainstem responses to tone bursts in normally hearing subjects," J. Speech Hear. Res. **31**, 87–97.

Greenwood, D. D. (**1990**). "A cochlea frequency position function for several species – 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

ISO 8253-1. (**1989**). "Acoustics - Audiometric test methods. 1. Basic pure tone air and bone conduction threshold audiometry," International Organization for Standardization, ISO, Switzerland.

John, M. S., Purcell, D. W., Dimitrijevic, A., and Picton, T. W. (**2002**). "Advantages and caveats when recording steady-state responses to multiple simultaneous stimuli," J. Am. Acad. Audiol **13**, 246–259.

Johnstone, B. M., and Boyle, A. J. F. (**1967**). "Basilar membrane vibration examined with the Mössbauer technique," Science **158**, 389–390.

Junius, D., and Dau, T. (**2005**). "Influence of cochlear traveling wave and neural adaptation on auditory brainstem responses," Hear. Res. **205**, 53–67.

Lütkenhöner, B., Kauffmann, G., Pantev, C., and Ross, B. (**1990**). "Verbesserung der synchronisation auditorisch evozierter hirnstammpotentiale durch vervendung eines die kochleären laufzeitunterschiede kompensierenden stimulus," Arch. Otolaryngol. Suppl. II, 157–159.

Mardia, K. V. (**1972**). *Statistics of Directional Data* (Academic, London).

Müller-Wehlau, M., Mauermann, M., Dau, D., and Kollmeier, B. (**2005**). "The effects of neural synchronization and peripheral compression on the acoustic-reflex threshold," J. Acoust. Soc. Am. **117**, 3016–3027.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-

burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.
Pantev, C., Lagidze, S., Pantev, M., and Kevanishvili, Z. (**1985**). "Frequency-specific contributions to the auditory brain stem response derived by means of pure tone masking," Audiology **24**, 275–287.
Parker, D. J., and Thornton, A. R. D. (**1978**). "Frequency specific components of the cochlear nerve and brainstem evoked responses of the human auditory system," Scand. Audiol. **7**, 53–60.
Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (**2003**). "Human auditory steady-state responses," Int. J. Audiol. **42**, 177–219.
Picton, T. W., Ouellete, J., Hamel, G., and Smith, A. D. (**1979**). "Brainstem evoked potentials to tonepips in notched noise," J. Otolaryngol. **8**, 289–314.
Richter, U., and Fedtke, T. (**2005**). "Reference zero for the calibration of audiometric equipment using 'clicks' as test signals," Int. J. Audiol. **44**, 478–487.
Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**(4), 1218–1231.
Shera, C. A., and Guinan, J. J. (**2000**). "Frequency dependence of stimulus-frequency-emission phase: Implications for cochlea mechanics," in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Tasake, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 381–387.
Shore, S. E., and Nuttall, A. L. (**1985**). "High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts," J. Acoust. Soc. Am. **78**, 1286–1295.
Siegel, S. (**1956**). *Nonparametric Statistics for the Behavioural Sciences* (McGraw-Hill, London).
Stapells, D. R., Gravel, J. S., and Martin, B. A. (**1995**). "Thresholds for auditory brainstem responses to tones in notched noise from infantand young children with normal hearing and sensorineural hearing loss," Ear Hear. **16**, 361–371.
Stürzebecher, E., Cebulla, M., and Elberling, C. (**2005**). "Automated auditory response detection: Statistical problems with repeated testing," Int. J. Audiol. **44**, 110–117.
Stürzebecher, E., Cebulla, M., Elberling, C., and Berger, T. (**2006**). "New efficient stimuli for evoking frequency-specific auditory steady-state responses," J. Am. Acad. Audiol **17**, 448–461.
Teas, D. C., Eldredge, D. H., and Davis, H. (**1962**). "Cochlear response to acoustic transients: An interpretation of whole-nerve action potentials," J. Acoust. Soc. Am. **34**, 1438–1459.
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Wegner, O., and Dau, T. (**2002**). "Frequency specificity of chirp-evoked auditory brain stem responses," J. Acoust. Soc. Am. **111**, 1318–1329.

Validation of the auditory hazard assessment algorithm for the human with impulse noise data

G. Richard Price[a)]
Auditory Hazard Analysis, P.O.B. 368, Charlestown, Maryland 21914

(Received 26 January 2007; revised 25 August 2007; accepted 27 August 2007)

Predicting auditory hazard from intense acoustic impulses, such as weapons fire or airbags, has been an intractable problem. The U.S. Army developed a theoretically based mathematical model of the ear designed to predict such hazards [the Auditory Hazard Assessment Algorithm for the Human (AHAAH)]. To validate it as a predictor of hazard, data from the literature (wave forms and changes in hearing sensitivity) were processed with the model in order to predict the onset of unacceptable threshold shift (25 dB or more) in the 95th percentile human ear. For comparison, alternate standards MIL-STD-1747D and A-weighted energy were also used to compute hazards for the same data. The primary dataset was that of the US Army's "Albuquerque studies" (53 different cases) and other impulses from the literature (19 additional predictions). The AHAAH model predicted correctly in over 95% of the cases, the MIL-STD-1474D was correct in 42% of the cases, and A-weighted energy was correct in 25% of the cases. Errors for all methods tended to be in the direction of overprediction of hazard. In addition to greatly increased accuracy, the AHAAH model also has the advantage of being theoretically based and including novel diagnostic features.
 [DOI: 10.1121/1.2785810]

PACS number(s): 43.64.Wn, 43.50.Yw, 43.64.Bt, 43.50.Qp [BLM] Pages: 2786–2802

I. INTRODUCTION

A. Background

Rating the hazard of intense impulse noises has long been a perplexing technical problem. The recent consensus of the scientific community has been that none of the existing standards is accurate (Chan *et al.*, 2001; Kardous *et al.* 2005; NATO, 1987, 2000, 2003). A long-term effort at the US Army Research Laboratory has produced the auditory hazard assessment algorithm for the human ear (AHAAH) (Kalb and Price, 1987; Price, 1988b, 2007; Price and Kalb, 1991, 1998). AHAAH is essentially a theoretically based electro-acoustic analog of the ear designed to predict hazard from any very intense sound, such as those produced by firearms or airbags, for example, where the loss mechanisms within the inner ear are thought to be essentially instantaneous mechanical stress. At lower levels characteristic of most industrial sources, in contrast, the loss mechanisms are more likely to be those associated with metabolic stresses. The final output of AHAAH is in auditory risk units (ARUs) that yield a prediction of immediate threshold shift, which in turn also provide a prediction of permanent threshold shift and hair cell loss (Price, 2005). A copy of this software, supporting documentation, and instructions about its operation are now downloadable at the Army Research Lab's website (2007).

As of this writing the model has been made available to the scientific community for use and comment for almost 10 years and it is gaining acceptance internationally as the basis for a noise standard. It has been reviewed and accepted as a standard procedure for evaluating airbag noise by the Society of Automotive Engineers (2003), it is being used to evaluate unprotected exposures by the US Army and is being proposed for use in the evaluation of protected hearing and is the basis for a proposed revision of the current military standard MIL-STD-1474D (2007) (interface standard for noise limits). Given that the AHAAH model is being used as a standard in some venues and is being considered in others, it is appropriate that the validation of the model with human data should be published in the scientific literature. The purpose of this article is to present the relevant background of the model and make the validation procedures explicit.

B. Need for accuracy

Ideally, a noise standard should be theoretically based, easy to use, unambiguous in its application, and most importantly, it should predict hazard accurately. The need for accuracy is critical from two perspectives. If the true hazard is underestimated, hearing loss will be produced in the exposed population. In addition to the great personal loss associated with deafness are the operational costs associated with hearing-impaired ears (Price, 2005). The other perspective is just as compelling, however. If the true hazard is overestimated, then there will be costs associated with overprotection of hearing and unnecessary design limitations. Neither inaccuracy is acceptable. These problems are perhaps most obvious for the military applications; but a current example from the civilian world would be the automobile industry and its design of airbags. Underestimation of hazard will result in loss of hearing in those exposed (Fleischer *et al.*, 2002; Price, 1998a, 2006; Saunders *et al.*, 1998; Yaremchuck, 1998; Yaremchuk and Dobie, 1999) and overestimation of hazard could result in ineffective airbag designs that would fail to protect from injury.

[a)]Electronic mail: ahanalysis@comcast.net

C. Design and development of the model

The AHAAH model is essentially an electro-acoustic analog of the ear structured to match the physiology of the ear element for element (reported more fully in Price, 2007; Price and Kalb, 1991). It was developed first for the cat ear where temporary threshold shift (TTS), permanent threshold shift (PTS), and cellular changes could be examined (Price, 2003a). Briefly, it reproduces the measured transfer functions from the free field to the stapes and translates stapes motion into basilar membrane displacements. It keeps track of the displacements at 23 locations (roughly 1/3 octave intervals) and derives a dose at each location by squaring the peak amplitude of each upward displacement of the basilar membrane (in microns) and summing them for the analysis interval. The result is in ARUs:

$$\mathrm{ARU} = \mathrm{sum}\ (D^2)$$

where D=upward basilar membrane displacement (in microns).

Further, it has been empirically established in the cat ear that ARUs can be converted into compound threshold shift (CTS) at 30 min (Price, 2006, Fig. 1):

$$\mathrm{CTS} = 26.6 \times \ln(\mathrm{ARU}) - 140.1$$

where CTS includes both temporary and possible permanent components of threshold shift.

This function is presumed to be the generic relationship between ARUs and threshold shift for at least the cat and human cochleas; thus this formula can be used to predict specific amounts of CTS for any number of ARUs.

With these intense exposures, where the loss mechanism is primarily mechanical stress, recovery begins only very slowly and is not likely to go to completion when a dose of 500 ARUs (25 dB CTS) is exceeded. When the 500 ARU limit is exceeded, the amount of PTS is usually about 0.6 of the loss at 30 min (Price 2006, Fig. 2). Further, it has also been shown that the predicted PTS corresponds to hair cell loss (Price, 2006, Fig. 3).

The foregoing relationships were established with a model developed for the cat ear and tested with exposure to explosive sources. The correlation between ARUs and mean threshold shift for 12 groups of ten animals exposed to different impulses was very high (r=0.94) which implied that little variance was left to be explained. Given the high accuracy of prediction with the cat ear, a parallel model was developed according to the same principles for the human ear. Only the coefficients in the equations were changed to reflect the differences in physical sizes of the two ears.

In this article the human model is being challenged with impulse noise data from the human database. Initially it was assumed that the model would have to be adjusted to improve the fit to the data, but so far no adjustment has been made.

The model includes the possibility of middle ear muscle contractions occurring either before the onset of the stimulus being analyzed (a warned exposure) or a muscle contraction that is elicited by the noise and includes a latency and a growth to full effect (unwarned exposure). In the case of free-field exposures, it also includes a transfer function from the free field to the eardrum appropriate to a wave approaching from the frontal quadrant (a worst-case angle). The model also provides an adjustment for susceptibility.[1] For use as a damage risk criterion (DRC), it has been adjusted to provide an analysis representative of the 95 percentile ear (most susceptible). The model also makes a movie of the action of the sound in the cochlea as a means of providing analytical insight into the damage processes in the time domain.

D. Data analysis

AHAAH operates by processing a digitized wave form; hence high-resolution digital recordings provide the best analysis. Failing that, pictures of wave forms can be digitized, but in the process there is a great loss in dynamic range. Therefore, formal analysis with AHAAH was normally restricted to those wave forms for which digitized recordings were available. Of course, in order to confirm the prediction of hazard, it was also necessary that human threshold shift data also had to be available for exposures to the wave forms.

Today there are about 70 exposures that meet these requirements. Most of these (53) come from the U.S. Army's experiments using human volunteers (Johnson, 1994, 1998). The remainder come from the impulse noise literature of the 1960s and more recent studies. An additional critical point is that all the newer data from experiments are from subjects with protected ears, whereas the earlier studies were typically done with subjects wearing no hearing protection. Because of modern human use considerations, such unprotected exposures of human ears are not likely to be repeated in experiments, although they may occur in practice. Therefore these data represent a particularly interesting test of the model.

Even when the digitized wave forms and human threshold shift data are available, comparisons are complicated by a lack of a common basis in reporting data. Studies often use unique measures of effect, e.g., number of impulses to reach some criterion [some amount of threshold shift in an ear(s) at some frequency(s), number of subjects reaching a criterion threshold shift, number of subjects not recovering in a given time period, etc]. In addition, the damage criteria themselves do not use a common data set to predict hazard from an impulse.

To cut through this Gordian knot of idiosyncratic information reporting and use, in this article we adopted a common method of comparison. DRCs are now commonly designed to protect the 95th percentile ear (most susceptible). So the problem for comparison of rating methods becomes that of determining whether any given exposure will produce an unacceptable threshold shift in the 95th percentile ear, which for gunfire type impulses is now commonly held to be 25 or more dB at any frequency (NATO, 1987, 2000). Experience in the Albuquerque studies supports the contention that 25 dB of threshold shift is tolerable and would be expected to produce no immediate permanent threshold shift from at least occasional exposure (Johnson, 1998).

TABLE I. Exposure matrix for the Albuquerque studies.

	NUMBER OF IMPULSES				
LEVEL	6	12	25	50	100
7		x	x	x	x
6					
5					
4					
3					
2					
1					

In this article, we consider the Army's recent experiments first, then the individual studies.

II. ANALYSIS OF THE EXPOSURE DATA

A. The Albuquerque studies

The U.S. Army has conducted the most extensive set of studies of impulse noise exposure of protected ears ever undertaken (Johnson, 1994, 1998; Patterson *et al.*, 1997). Given that they were conducted at a facility near Albuquerque, NM, they have become collectively referred to as "the Albuquerque studies." The studies represent the greatest source of data about protected human exposure to intense impulses of which we are aware. In them, the exposure pressures were also far higher than in most studies, ranging from 95 dB peak pressure level (PPL)—higher even than those around modern weapon systems. Given the difficulty and cost in conducting such research with human volunteers, it seems unlikely that this database will be superseded.

1. Design of the studies

The design of this series of studies was based on the concept that exposure to impulses produced by exploding charges of differing weights and physical locations would simulate exposures to generic large caliber weapons used by the Army. The exposure scheme was delineated as diagrammed in Table I. Each level of the exposure was intended to be a 3-dB increase in the peak pressure of the impulse and a doubling of its energy. A particular exposure was considered dangerous for a subject if it produced a 25 dB or greater threshold shift at any test frequency.

Volunteer subjects with normal hearing were exposed to only one type of impulse (11 exposures to complete the simplest matrix). Each group began the exposure with about 60 Ss. They entered the exposure matrix (Table I) at the lower left corner (Level 1, six impulses) wearing a circumaural hearing protective device and if they had no significant threshold shift they progressed to the next higher exposure, Level 2, six impulses, and so on up the left side of the matrix as indicated by the "up" arrow. When they reached Level 7, six impulses, and passed it, they dropped back to Level 6 and were exposed to increasing numbers of impulses, as indicated by the right-pointing arrow. If they received no significant threshold shift, they were exposed to the next higher number of impulses. If they reached Level 6, 100 impulses, they had successfully completed the course of exposures. For the most part, this was the pattern of exposure. If, however during the course of traveling this path, a subject experienced a significant threshold shift, he was dropped down in level and moved toward more impulses at the lower level and so on. The apparently reasonable assumption behind this approach—that lower peak pressures were associated with less risk—will create interpretive problems that will be covered in discussion of the data.

In examining the data, it will be apparent that the size of the group in the cells did not remain constant. This resulted because the Ss were volunteers who could opt out of the test, and because an individual might be dropped for some reason unrelated to the purpose of the test.

In the initial conception of the study, it was presumed that failures would occur as the levels increased and that the subjects would "migrate" on a "just safe" path through the matrix at lower levels and increasing numbers of impulses. It was also presumed, on the basis of existing DRCs, that double hearing protection would be needed in order to complete the matrix, but it was not necessary. In fact, the ear appears to be surprisingly robust and such failures occur less often than might be expected. As a result, only a few subjects were tested on the interior cells of the matrix (Levels 1–5 and 12, 25, 50, and 100 impulses) and the data there are only suggestive.

2. The exposures

The subjects were located 5, 3, or 1 m from the source and the size of the charge was varied to achieve desired impulse. In traditional acoustic terms, durations of the first positive peaks of the impulses were about 2.6, 1.5, and 1.0 ms, respectively. The pressure histories of the impulses and their energies, etc. have been reported in Patterson *et al.* (1997) and in Fig. 3 of Chan *et al.* (2001).

Reverberant impulses were produced in a steel-walled chamber (3 m × 3 m × 2.44 m) and had envelope durations of about 300 ms. They were intended to simulate weapons fired from bunkers, within rooms, or the pressure history within the crew compartments of fighting vehicles. An explosive charge was fired outside the chamber at the end of a barrel that extended into the chamber. The wave forms for these impulses have been published in Johnson (1998). For the reverberant impulses, exposure was limited to one round at all levels of Table I and two and three rounds only at level 6.

For the nonreverberant impulses, subjects were seated during all exposures with their heads in a chin rest and their right ear facing the impulse source. For the reverberant im-

pulses, the left and right ears were tested about equally. The nontest ear (typically the left ear) always wore double hearing protection. Impulses were delivered 1 min apart following a countdown (ten, nine, eight, seven —-) that was audible because of a peak-limiting talk-through circuit in the muff. Subjects were well aware of when the impulse was coming, and in fact depended on the countdown.

3. Audiometry

Thresholds were determined by automatic audiometry just before exposures, then immediately afterward, and changes were followed until they had resolved. Two types of "audiometric failure" were recognized in the study. Any threshold shift greater than 25 dB at any one frequency was defined as a "full audiometric failure." Threshold shifts between 15 and 25 dB, however, were taken as evidence that the ear was being stressed and the experimenters were justifiably reluctant to continue by exposing that ear to the next higher condition. These were considered "conditional failures" at the next higher level.

This level of shift was apparently well chosen. It avoided false-positive results as well as permanent threshold shift. All subjects recovered to their pre-exposure thresholds.

4. The hearing protector

The Racal muff was originally chosen because it represented a moderately effective hearing protector that had a talk-through circuit and could be worn under an infantry helmet (used in the tests). The unexpected result from the first 5-m study was that the muff alone provided adequate protection for all subjects at all exposures. The experimenters then made the critical decision to defeat the seal on the right muff (exposed ear toward the impulse source) by inserting eight plastic tubes (2.3 mm inside diameter, four open to the front and four open to the rear). This leak was intended to simulate a badly fitting ear seal and perhaps more closely replicate fit during field conditions. The 5-m study was rerun with these muffs followed by the 3-m and 1-m conditions as well as a reverberant room exposure to simulate firing from an enclosure.

5. Problems with the studies

a. The hearing protector For the four sets of tests in which the modified muff was used, the protector's performance became a major and variable element in the exposure, accounting for almost as great a range in stimulus strength at the ear canal entrance as the deliberate manipulation of peak sound pressure level in the exposures. The attenuation of the muff as a function of level and exposure condition can be seen in Fig. 1. The data in this figure have been taken primarily from the report by Patterson *et al.* (1997) in which they reported the sound pressure levels (SPLs) outside and inside the muffs during the exposures and also from a CD issued by USAARL with the recorded pressure histories. The data plotted on the ordinate in Fig. 1 are attenuation in A-weighted energy; but essentially the same picture would have emerged if sound exposure level were used. The abscissa is in dB with respect to the base level for the exposures. If the muff were linear in its operation all the lines would have been horizontal. The picture that emerges is

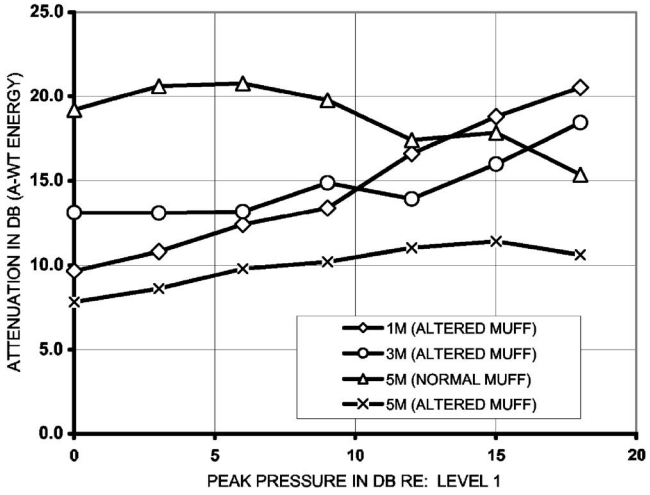


FIG. 1. Attenuation of the muffs used in the Albuquerque studies for three different exposure conditions displayed as a function of the test level (in dB with respect to the lowest exposure level in the tests). The base level(s) were 178 dB (1 and 3 m) and 173 dB (5 m). By way of comparison, there is also a line for a muff with an intact seal for the 5 m condition. The designation of conditions by "*m*" is a convention from the Albuquerque studies and refers to the distance from the impulse source.

quite different, however. It is apparent that the normal muff attenuated best at lower intensities (a little over 20 dB) but got worse by about 5 dB as the pressure rose above Level 3. In contrast, the altered muff for the 5-m condition attenuated much less initially (just under 8 dB at Level 1), but got *better* by a little more than 4 dB as the peak pressure rose. For both the 3- and 1 m impulses the same pattern of improved attenuation with higher peak pressures can easily be seen. Attenuation for the 1-m condition grew by 11 dB.

The reduction in attenuation at high levels with the intact muff might be expected on theoretical grounds (Buck, 2000). On the other hand it appears that the holes installed in the defeated muff progressively increased attenuation at higher SPLs, a trend that was accentuated as the spectral peak of the impulse rose. The holes as configured (normal to the wavefront and next to the reflecting surface of the head) were apparently producing an effect like that seen in hearing protective devices (HPDs) deliberately designed to be nonlinear (Hamery *et al.*, 1997).

The import of these nonlinearities on the interpretation of the data is critical. Depending on the specific conditions, the pressure under the muff to which the ear was actually exposed rose anywhere between 7 and 23 dB, for a nominal 18-dB change in the peak pressure of the incident wave form. Changes in level in the free field might have been 3 dB/step; but at the ear they were truly between 1 and 4 dB/step.

An analytical technique or impulse noise standard, such as MIL-STD-1474D, which uses the pressure measured in the free field (Chan *et al.*, 2001) and/or makes a single value correction for any hearing protector, obviously misses this detail in these or other similar data. On the other hand, AHAAH and the A-weighted energy measure use the wave form measured under the protector for their calculation and can in principle evaluate all the impulses in the Albuquerque tests accounting for the bias introduced by the muffs.

b. Limitations of the data set A second concern is that we have no full exposure measured under the muff for an individual subject. We are fortunate to have the sample re-

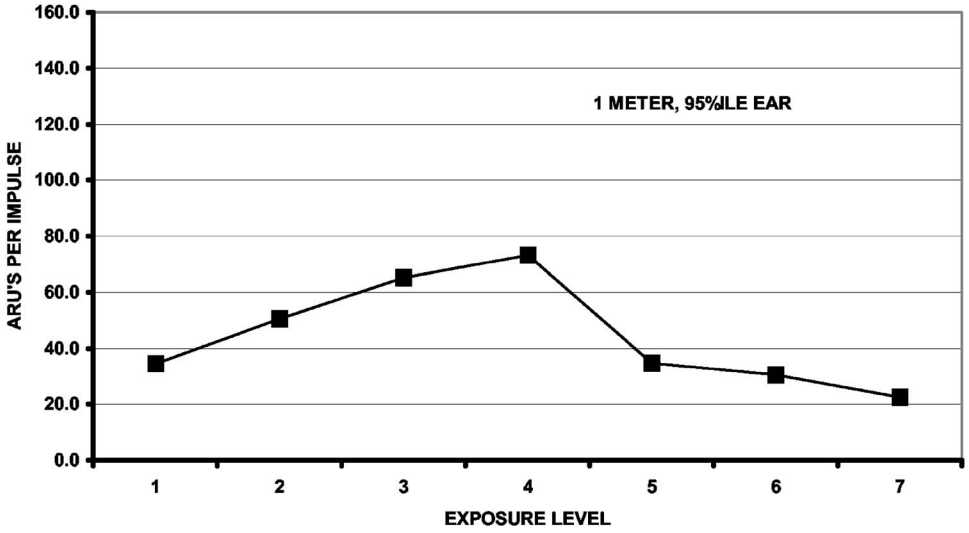


FIG. 2. Plot of ARUs per impulse for pressure histories under the muff for the 1 m exposure, 95 percentile ear.

cordings distributed with the Albuquerque reports; however, it is not possible to follow a full exposure for any individual subject. Levels above 180 dB are truly intense noise exposures and have the ability to produce unexpected effects, such as the nonlinearity in Fig. 1 or movement of the muffs during the exposure [actually observed in the tests (Johnson, 1994)]. Estimating the effect of 50 or 100 impulse exposures for more than 240 subjects, based on data from a small sample of impulses includes some uncertainty. Change in hearing at these levels sometimes occurs very suddenly, as though one impulse were particularly bad. It is also true that sound fields around explosive sources can have occasional "hot" spots. Given the possible importance of rare events, it would have been interesting to have full recordings of the exposures. In lieu of that, samples of the five or six impulses that were available for each condition were analyzed.

6. Interpretation of the data sets

Evaluating the Albuquerque data set with the AHAAH model was for the most part a straightforward exercise. The one exception is that we had no pressure histories under the muff for subjects wearing the defeated muff in the 5-m condition. Fortunately, data were available from preliminary work done with the experimenters themselves in which data had been recorded under the muff for this condition.

a. Interpretation of failures An intuitive approach to the data sets might assume that within a data set (5, 3, 1 m, etc.) there is a similarity between all the impulses and that a failure at some given peak level would naturally imply a failure at higher levels as well. However, a detailed analysis of the ear's response to a pressure history with the AHAAH model suggests that such an assumption is not warranted, e.g., wiggles in the wave form during very high pressures (when peak clipping in the middle ear was strong) would not be transmitted to the inner ear at full strength (see Price, 2005, for discussion of this type of effect) or in some cases the shape of the pressure history changed in significant ways as a function of level. The model's analysis was sensitive to these details and calculated that the hazard from the impulses did not grow monotonically with peak level. The result was that for the 1 and 3-m series of exposures, the greatest hazard was predicted to come from the middle levels of peak pressures and actually became *lower* as the peak pressure rose higher.

This state of affairs is presented in Fig. 2 for the 1-m exposure condition, 95th percentile ear. The prediction of inverted effect is unexpected, but given the model's analysis, it is not irrational. If true, it follows that failure at a lower level of this condition in the Albuquerque studies does *not* necessarily imply failure at higher levels. Hence, to be consistent with the model's implications in scoring failures, failure at higher levels was not presumed unless the higher exposure level actually contained a larger number of ARUs. Other analyses of the Albuquerque dataset, in keeping with their premises, have propagated failures at lower levels to all higher levels (Chan *et al.*, 2001) while others have simply ignored the issue even when using the model (Patterson and Ahroon, 2005).

A statistical problem, inherent in the design of the study, is that once a subject failed, he did not remain part of the exposure group; hence the group had lost its most susceptible subject (at least for that particular exposure). Had the impulses truly been an ascending series, assuming failure at the higher levels might have dealt with the question of the loss of the susceptible subject. In an ideal world there would have been a new group of subjects for each level; but such an approach was simply not practical.

In the end, the question of whether or not to consider failure at lower levels as indicating failure at higher levels turns out to be something of a tempest in a teapot. The data have been analyzed using the different approaches and in the end, there is only a small difference in assessment of the accuracy of the ear model's prediction (Price, 2003b).

b. Establishing the 95 percentile failure A statistical argument was developed (Johnson *et al.*, 1990; Patterson and Johnson, 1994) that in a sample size of 60, one can be roughly 95% confident that the true 95 percentile subject lies between the exposures in which one and six subjects show a significant threshold shift. Alternatively put, when there is no threshold shift exceeding 25 dB, we can be 95% confident that the exposure is safe for the 95 percentile subject. Likewise, when six ears show a threshold shift of 25 or more dB, we can be 95% confident that the 95 percentile threshold shift also exceeds 25 dB. When 1–5 subjects show threshold excess threshold shifts the exposure is indeterminate—it could be either safe or hazardous. In such cases, the prediction was given the benefit of the doubt and not counted as incorrect.

TABLE II. Summary of tests for the 1-, 3-, and 5-m exposure conditions (number of ears/number of failures).

	1 M defeated muff					3 M defeated muff					5 M defeated muff				
Number	6	12	25	50	100	6	12	25	50	100	6	12	25	50	100
L7	56/2					41/0					54/0				
L6	59/0	56/2	56/4	49/7	37/11	64/2	64/2	61/3	57/9	36/11	57/0	57/1	57/1	57/1	56/2
L5	63/1	6/1	9/3	10/6	19/10	67/1	10/2	8/2	9/3	22/4	57/0				
L4	63/0					68/0					58/0				
L3	64/0					68/0					59/0				
L2	65/0					68/0					59/0				
L1	65/0					68/0					59/0				

c. The AHAAH calculation From the standpoint of the model, unacceptable threshold shift was predicted to occur in the 95 percentile ear when an exposure totaled 500 ARUs. Therefore, using wave forms recorded under the hearing protector, five or more impulses from each exposure condition (as available) were run through the model with input in the model set at the ear canal entrance and their average was taken to be characteristic of that exposure level. The total exposure, of course, was the number of ARUs per impulse multiplied by the number of impulses. Because the Albuquerque studies included an audible countdown before each impulse, we presume that the subjects were "braced" for the event and that their middle ear muscles had contracted before the exposure. In support of this contention is the fact that the warning was very important to the subjects. When the experimenters tried one series of experiments without the clear warning, the subjects did not want to complete the exposure series.

d. Predictions by other methods For comparison purposes, the predictions of MIL-STD-1474D have been calculated by Patterson *et al.* (1997) and are presented. There has also been interest in using A-weighted energy, even for intense impulses (Dancer, 2000). If such a measure worked, it would have great advantages in measurement and in the ability to combine exposures from a variety of sources and levels, with and without hearing protection. It would also be compatible with ISO-1999. For a free field measure (subject absent) the application of an 85-dB L_{AEQ8} criterion is clear. However, where the measure is under a protector, some adjustment should be made for the fact that the measure includes higher pressures in some frequency regions due to the acoustic effects of the head and external ear. And the proposal to use such an energy measure usually contains a caveat against using it above 160 dB peak pressures because of a possible critical level (Dancer, 2003). This would eliminate virtually all the exposures in the Albuquerque studies even when measured under the protector. Nevertheless, it would be instructive to do an A-weighted energy calculation just to see over what range such a measure might apply. In terms of energy, an 85-dB L_{AEQ8} exposure contains 8.7 J/m^2. Exposures less than that were rated safe, more than that, hazardous.

7. Pass/fail analysis of the Albuquerque data

Unfortunately, the data on subjects' threshold shifts (TSs) were not immediately available; however the exposure charts documenting each Ss path through the exposure matrix were (Johnson, 1994, Figs. E1–E45). Therefore, each subject's path through the exposure was individually traced. As outlined earlier, failures at lower levels were not counted as failures at higher levels unless they contained more ARUs; but *within* a level, failures at lower *numbers* of impulses were counted as presumptive failures for higher *numbers* of impulses. And because Ss were dropped when they showed an unacceptable TS or had some other issue—often not related to the experiment—not all data groups contained the target number of 60 Ss.

The data summaries developed in this analysis appear in Table II. The reverberant exposures and the 5-m exposures with the good muff produced no losses; therefore their data are omitted from Table II. If for the 1-m exposures we focus on the condition producing more than five failures (clear indication that 95 percentile exposure is exceeded), the data in Table II clearly indicate for Level 6 a threshold between 25 and 50 impulses for the 1-m exposure. The smaller number of Ss associated with 100 impulses in Level 5 are not large enough to take the results too seriously; however they do represent a very high failure rate—even higher than 100 impulses at Level 6. It should be noted in this regard that when the number of failures at Level 6 was seen to be high (which implied that there was no point in further exposures there), the experimenters stopped the exposures at Level 6 and began exposing Ss to 100 rounds at Level 5, rather than at Level 6. So the Ss tested here were not primarily susceptible Ss dropped down from a failure at Level 6.

In Table II it is clear that for the 3-m exposure the 95 percentile exposure for Level 6 lies somewhere between 25 and something less than 50 impulses. The values in Level 5 are too small to be conclusive; but they do represent an extraordinarily high rate of failure and are included for interest.

For the modified muff and the 5-m condition, the pattern of failures at Level 6 is a bit problematic. A conditional failure at *L*6/6 is a presumptive failure at the 12 and 25 impulse conditions, with additional failures happening at the 100 impulse condition. It is surprising that no additional failures occurred between the six and 100 impulse exposures.

8. Evaluation of predictions by three methods

A convenient summary of the data can be seen in Table III. In each section, the predicted hazard of an exposure (either "safe" or "hazardous") is compared with the actual outcome of the exposure (again either "safe" or "hazardous"). Evaluations by MIL-STD-1474, A-weighted energy, and AHAAH are presented. Table III contains three matrices with the outcomes. Entries in the upper left and lower right quad-

TABLE III. Outcomes for analysis with the three methods.

MIL-STD-1474D

PRED \ OUTCOME	SAFE	HAZ
SAFE	11	0
HAZ	33	9

A-WEIGHTED ENERGY

PRED \ OUTCOME	SAFE	HAZ
SAFE	4	0
HAZ	40	9

AHAAH

PRED \ OUTCOME	SAFE	HAZ
SAFE	41	0
HAZ	3	9

rants are correct predictions (safe prediction/safe outcome or hazardous prediction/hazardous outcome). Entries in the upper right and lower left quadrants represent errors (safe prediction/hazardous outcome or hazardous prediction/safe outcome).

In Table III we see that MIL-STD-1474D was correct in its evaluation of 22 of the exposures and incorrect in 31, an accuracy of 42%. Its errors were all in the direction of *over*predicting hazard. Next are the results for A-weighted energy. It was successful in 13 cases out of 53, an accuracy of 25%. This method also erred in *over*predicting the true hazard. The amount of error in the overprediction was sizable, often 10–20 dB. Finally the results for the AHAAH model indicate that it was correct in all but three cases for an overall accuracy of 94%. Its three errors were also in the direction of *over*predicting hazard.

B. Additional studies

The following studies are important because they allow a test of the model's algorithms appropriate for unprotected exposure to free-field impulses, in some cases produced by real weapons, provisions bypassed by the Albuquerque studies, where pressures were measured under the muff. Also, in this set of impulses are spark-gap discharges with spectral peaks above the midrange that test the model's high-frequency response. Together they represent a valuable test of the model's capabilities.

The impulses sources include the 7.62-mm rifle, the Belgian FNC rifle, the German G3 rifle, and a 120-mm mortar. Studies using the M72 LAW and spark gap discharges, while lacking in some details, also provide general tests of the fit of the model to human data. And from the clinical literature we also have some indication of exposures that have produced permanent threshold shifts in human subjects. Given that each is unique, we will consider them individually.

In doing the hazard analyses of these impulses, the AHAAH algorithm and the A-weighted energy methods can be used. MIL-STD-1474D does not allow any exposure without hearing protection when the peak level is above 140 dB; hence its accuracy for most of these impulses cannot be tested.

1. Analysis of other impulses

a. Rifle impulses: 7.62 mm Hodge and associates at the U.S. Army's Human Engineering Laboratory (Hodge *et al.*, 1964, 1965; Hodge and McCommons, 1966) conducted an extended series of pioneering studies. In those studies the unprotected left ears of soldiers were exposed to 25 or 50 impulses produced by a weapon firing a 7.62-mm round at 155- or 158-dB peak pressures. There was no specific signal to the subject that an impulse was coming; but the rounds were fired at regular 5 s intervals, so it is reasonable to suppose that subjects knew when at least the second and successive rounds were coming. The reports of these studies included a careful rendering of only the initial peak of the pressure history and could not be used for analysis with AHAAH. Fortunately, we had more recently recorded several impulses from the M-14 rifle, which also fired the same 7.62-mm round (Fig. 3). The peak levels of the recorded rounds were adjusted to those levels used in the studies. It is probably not an exact reproduction of the wave form; but we believe that it is very close to that used in the studies. They used varying numbers of subjects (7, 12, 28) and reported the upper and lower limits, quartiles, means, medians, and standard deviations of the threshold shifts.

In essence, three exposure conditions, 50 rounds at 155 and 158 dB and 25 rounds at 158 dB, repeatedly produced maximum TTS_2 of 40, 50, or even 70 dB. The third quartile data in Hodge *et al.* (1964) were also well above 25 dB, and in other cases (Hodge and McCommons, 1966), the mean plus 1.64 SD (95% of standard normal distribution) was also above 25 dB at 4 and 6 kHz. All these exposures

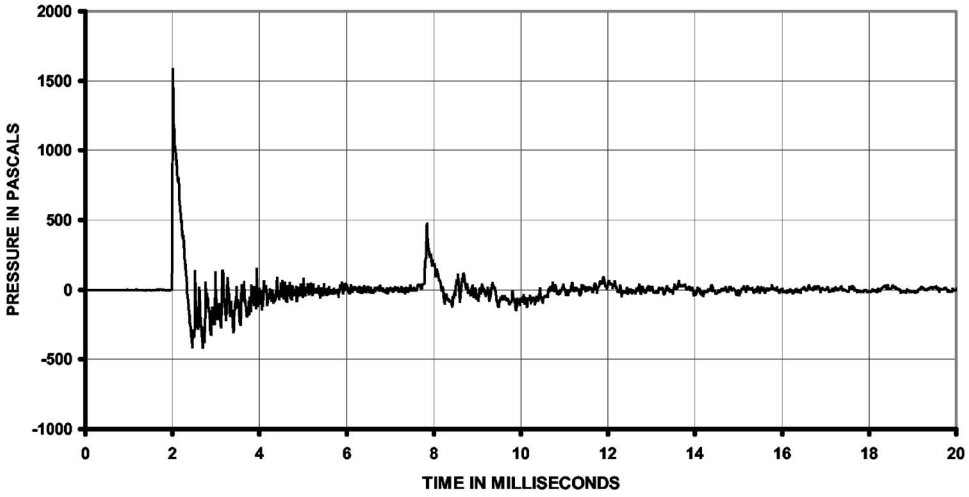


FIG. 3. Pressure history of 7.62 mm rifle impulse at the subjects' ear position used in the exposures used by Hodge *et al.* (1964, 1965, 1966).

would be rated as hazardous by the rules of the current analysis. They also noted that 25 rounds at 158 dB produced a little less threshold shift than 50 rounds at 155 dB (no statistical test, however).

Although there was no countdown, the timing of the impulses was regular and when the exposures were repeated with the same subjects, the result and its variability were the same. This would argue for the use of a "warned" calculation. The 155-dB impulses produced 16.6 ARUs/impulse (warned) and 118 ARUs unwarned. Therefore, 50 rounds at 155 dB would contain 830 ARUs, which would rate the exposure as hazardous. It would also predict a threshold shift of 39 dB at 1/2 h. This falls between the third Quartile (16 dB) and the upper limits of 68 and 75 dB at 4 and 6 kHz, respectively (measured at 2 min.) (Fig. 2, Hodge *et al.* 1965). In terms of A-weighted energy, each impulse contained 0.27 J/M^2 and a 50 round exposure, 13.5 J/M^2, which would also be rated as hazardous on the basis of its energy content. At 158 dB each impulse resulted in 22.3 ARUs (warned) and 153 ARUs (unwarned). This means that a 50-impulse exposure would produce 1115 ARUs and a 25-impulse exposure, 558 ARUs. Both are considered hazardous. AHAAH predicts 47 and 28 dB of CTS, respectively (at 1/2 h) and that is close to the result in the experiments. In keeping with the conclusion of Hodge and McCommons (1966) that the 50 round exposure at 155 dB was somewhat worse than the 50 round at 158 dB, AHAAH's analysis concurs: 830 ARUs for the 50 round exposure and 558 for the 25 round exposure.

In terms of A-weighted energy, the 25 round exposure contained 13.3 J/M^2 and the 50 round exposure, twice that. Both would correctly be rated as hazardous on this basis.

b. Rifle impulses: FNC rifle (5.56 mm) Brinkmann (2000) reported two exposures conducted in Germany in 1978 using the Belgian FNC rifle (Fig. 4). In one experiment, 51 troops fired six rounds from a standing position with no hearing protection and in the other, 53 soldiers fired five rounds similarly. In the first case, 11 subjects showed threshold shifts greater than 25 dB and in the second, nine subjects showed shifts greater than 25 dB. In both cases, recoveries of these shifts took from 30 min to 16 h. Given the size of the shifts and the number of subjects showing them, both exposures would be rated as hazardous by the standard being used in this analysis.

There were no wave forms available from the 1978 study; however, as part of the work of NATO RSG 29, Panel 8, Brinkmann recorded impulses from the FNC rifle for use with AHAAH. Because the soldiers fired their own weapons, a "warned" middle-ear muscle calculation was appropriate. The analysis showed an average of 156.8 ARUs per impulse (seven impulses analyzed). The two exposures would have

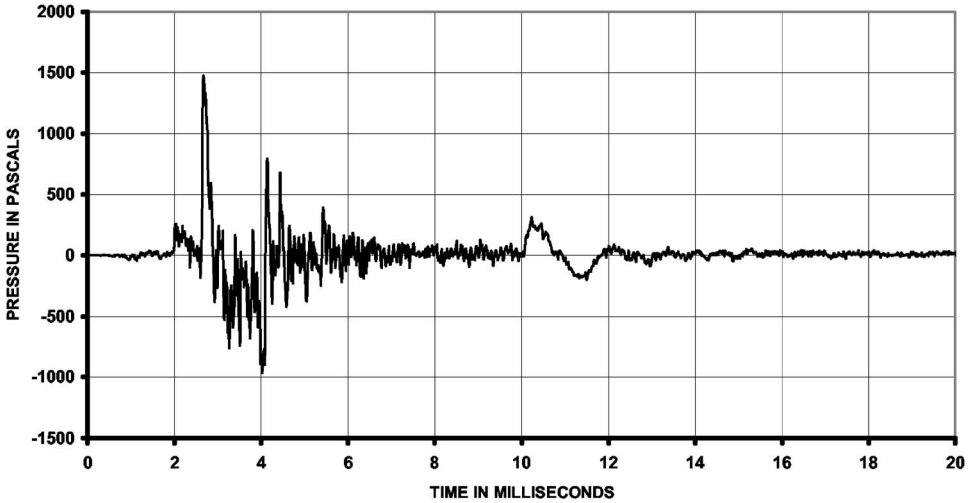


FIG. 4. Pressure history of FNC rifle impulse at firer's ear position.

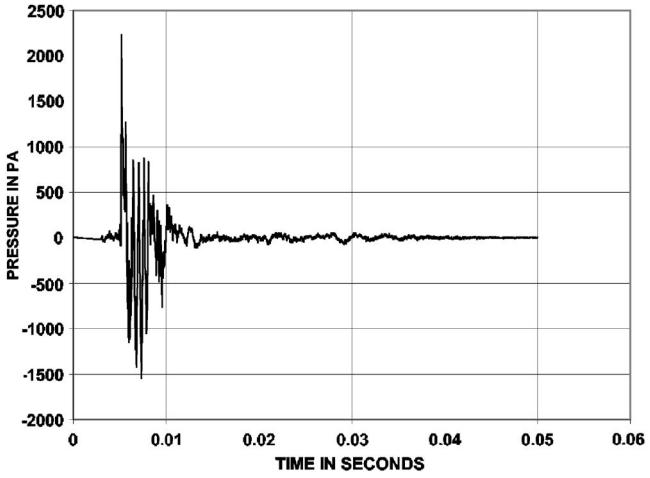


FIG. 5. Pressure history for G-3 rifle at firer's ear.

been to 940 and 784 ARUs, respectively, both in the hazardous range and consistent with the threshold shift data.

In terms of A-weighted energy each impulse contained an average of 1.4 J/M^2, which means that the exposures contained 8.4 and 7 J/M^2, respectively. Both of these would be rated as safe on the basis of A-weighted energy. This is the only instance in which any method has rated an exposure as safe that was in fact hazardous. On the other hand this error was very small (less than 1 dB).

c. Rifle impulses: The G3 rifle The standard weapon in the German army was the G3 rifle. Pfander (1975) reported that as part of their training, 78 soldiers had fired five rounds without hearing protection and were tested just before and just after the firing. Its pressure history appears in Fig. 5. Pfander's method of reporting data grouped soldiers based on how long it took them to recover; hence it is not possible to determine exactly how many soldiers had a particular threshold shift. However, from the data reported, we can say that some soldiers in four subgroups had threshold shifts of 30 dB, and others in two more subgroups had threshold shifts of 50 dB [the groups had long recovery times (3–6 days) and included 16 soldiers] (Pfander, 1975, Fig. 38). Clearly, this is a hazardous exposure. In fact, these data were instrumental in changing German training. This study was the last in which German soldiers in training were allowed to fire with no protection.

No pressure history from the original study was available; but as part of the NATO RSG's work, Brinkmann also supplied an impulse from the G3 rifle for analysis that is presented in Fig. 7. The soldiers had fired their own weapons; therefore a warned exposure is appropriate. The impulse contained 178.3 ARUs. In the tests, an exposure consisted of 892 ARUs, which is in the hazardous range and consistent with the threshold shift data.

In terms of A-weighted energy the G3 rifle impulse at 161 dB peak, contained 4.2 J/M^2; therefore the exposure was to 21 J/M^2, which would correctly be considered hazardous.

d. M-72 light antitank weapon (LAW) exposure The following study is not an ideal test of the AHAAH model because we lack the original pressure histories and because the experimenters excluded what may have been the most susceptible subjects. However, the outcomes are not "subtle" and the data do represent a check on the reasonableness of the model's predictions in extreme cases. Garinther and Hodge (1971) and Hodge and Garinther (1970) exposed subjects with no hearing protection to an impulse from the M-72 light anti-tank weapon (a shoulder-fired rocket). Exposures were at grazing incidence and at a peak pressure of 161 dB, 8 m to the left and right rear of the rocket. The experimenters fired the rocket remotely following a short countdown, presumably audible by the subjects. They also allowed subjects to fire one round *unprotected* at the firer's position (179-dB peak pressure). The subjects had volunteered, saying that they had done the same thing before in combat. The same study also had them fire with hearing protection (V-51R plugs, experimenter fit). The primary problem with the analysis is that we lack the pressure histories needed. We do have one digitized impulse from the M-72 LAW (presented in Fig. 6), but it differs in appearance from the pressure histories reported in the studies. A second concern is that the authors note that some subjects were dropped from the study "because of slow recovery" which would imply that the more susceptible ears were dropped. Therefore, only the most tentative comparisons are warranted from these data. The TSs from unprotected exposures at 161 dB were in the safe range. It was also clear that the 179-dB exposure with unprotected ears was *very* hazardous and exposures were stopped early. The protected exposures up to 184 dB peak were shown to be safe.

Given the one digitized impulse, we were able to change its apparent pressure by scaling within the computer. Real changes in pressure history would of course involve

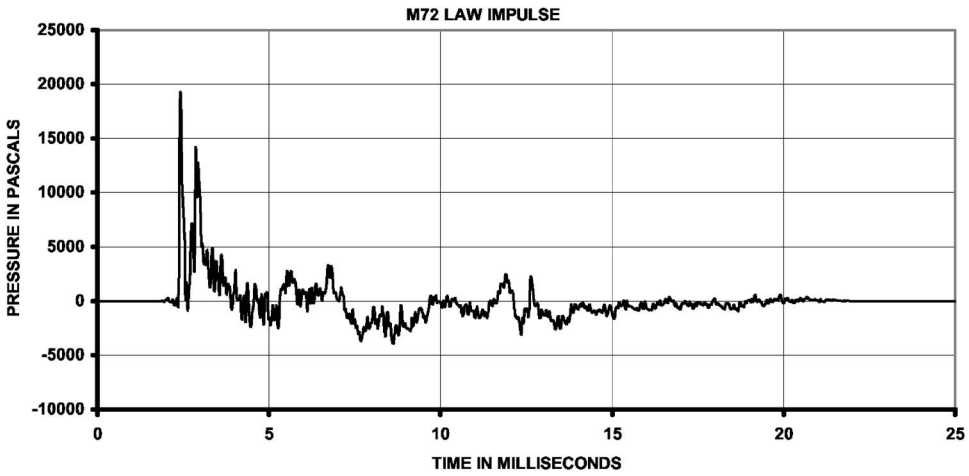


FIG. 6. Pressure history of LAW impulse at the firer's ear position.

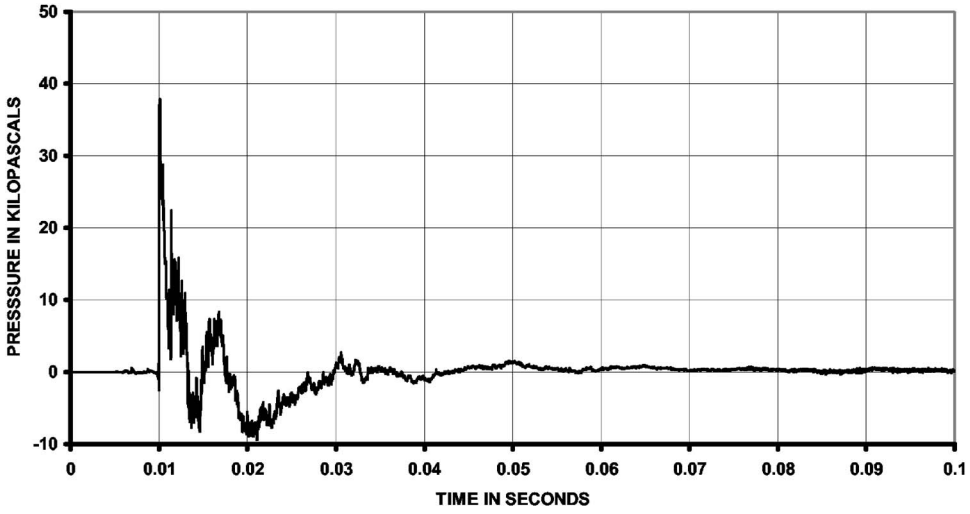


FIG. 7. Free field pressure history of 120 mm mortar impulse at the firer's head location. The peak pressure is 185.5 dB SPL.

different distances with different arrival times for the ground reflections, etc. Given all the caveats, it appears that AHAAH is at least in the right ballpark. The 161-dB impulse produced 105 ARUs warned and 516 ARUs unwarned. Such an impulse could be "just safe" as the data suggest. The 179-dB impulse produced 921 ARUs warned and 4217 ARUs unwarned, somewhere between dangerous and very dangerous (as the data suggest). And for the protected ears, AHAAH says that the firing should be safe (as it was).

In terms of A-weighted energy, the impulse analyzed contained 1.5 J/M^2 when the peak pressure was 161 dB and 114 J/M^2 when it was at 179 dB. These agree with the first being safe and the second, hazardous.

e. Mortar impulses This study is also not an ideal test of the AHAAH model because it was primarily a test of a new nonlinear earplug design as worn by mortar crews. Even though the number of ears tested was small, the data were carefully collected for an intense exposure under nearly true operational conditions. The French army tested a nonlinear earplug by having mortar crews (seven crews of four men) fire a 120-mm mortar while wearing the new earplug. The nonlinear plug was also inserted in the ear canals of an acoustic manikin specifically designed for such tests. The plugs were designed to attenuate minimally at low sound pressure levels (>110 dB), to promote communication and situation awareness, and to attenuate much more at high sound pressure levels in order to provide protection from intense impulses. The troops were professional mortar crewmen and in the test they fired as a crew normally does, i.e., as loader, ammo handler, section chief, and gunner. They did fire the mortar with a lanyard (rather than just dropping the round), which means that the instant of firing was known. They fired seven rounds (peak pressures of 185 dB at the firer's head) while wearing the nonlinear earplug. The free field pressure at the firer's location is portrayed in Fig. 7. The section chief and ammo handler were exposed to somewhat lower peak pressure levels than the other crewmen. The number of subjects actually tested was relatively small (seven soldiers at the area of maximum pressure). On the other hand, these were real troops doing their jobs, not precisely positioned, immobile subjects, which adds at least face validity to this test.

From the standpoint of earplug design, the test was a success because no subject experienced a significant threshold shift.

Digitized recordings of the pressures at the eardrum position of the acoustic manikin were made available (Dancer, 2003) for analysis with the ear model (Fig. 8). The hazard calculated for the round in the figure was 7.2 ARUs for a warned ear and 36.8 ARUs for an unwarned ear. For a seven round exposure the total dose would have been 50.4 and 257.6 ARUs, respectively; both rated as safe exposures.

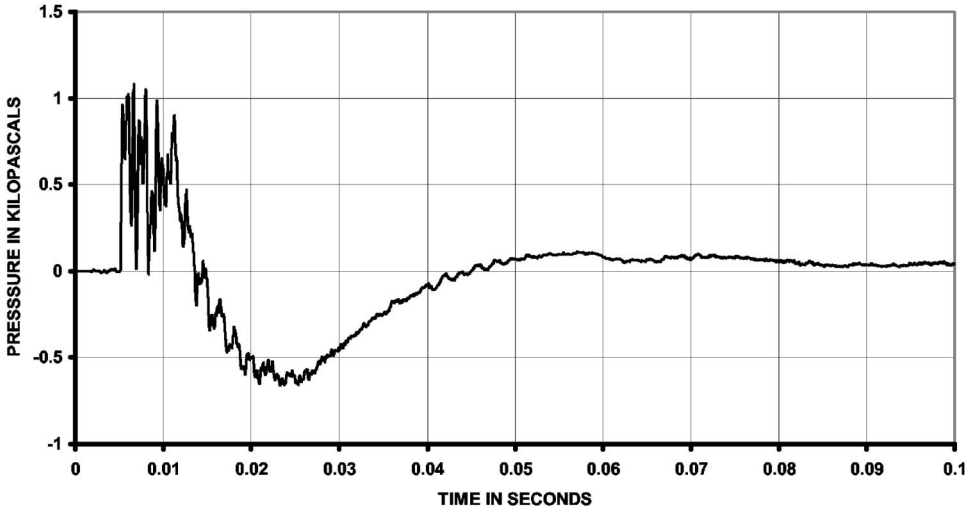


FIG. 8. Pressure history of 120 mm mortar impulse at the manikin's eardrum position under the nonlinear earplug. Manikin located at firer's head position.

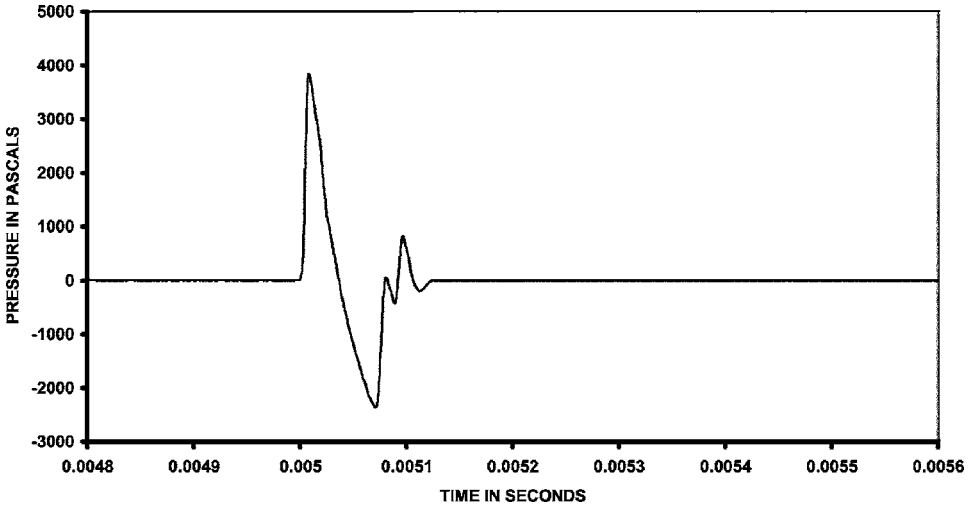


FIG. 9. Pressure history of three-gap impulse used by Loeb and Fletcher (1968).

Individual impulses under the hearing protector had A-weighted energies of about 1 J/m^2, the total for seven rounds being less than the 8.7 J/m^2 that would have been rated hazardous. A-weighted energy would have correctly rated this exposure as safe.

Given that hearing protection was worn, the MIL-STD-1474D allows an estimate of hazard. It would allow no rounds to be fired; but seven were shown to be safe.

f. Spark gap impulses Loeb and Fletcher (1968) developed a spark gap impulse generator and were able to produce impulses at 166 dB peak. By combining various numbers of gaps sequentially fired, they produced initial positive peaks of 36 μs (three gaps)–96 μs (six gaps). Subjects were exposed to 1/s in the free field in an anechoic chamber. The pressure histories were detailed enough to be digitized by scanning and the fact they were presented in an anechoic chamber meant that reflected waves were essentially not present (Figs. 9 and 10). In terms of spectrum, the six-gap impulse produced its peak energy at about 3 kHz, which is where the ear is tuned best and is maximally susceptible. The three-gap impulse had its spectral peak at about 8 kHz. From a theoretical standpoint, these particular tests of the model were interesting. All the other impulses we have examined had their spectral peaks below the midrange, which is common for most impulses in real life. So in effect we have tested the low-frequency to midfrequency portion of the model. These two sets of impulses, however, test the midfrequency to high-frequency portion of the model.

In these experiments, 72 subjects were tested. Exposure began with one impulse and continued with increasing numbers of impulses until a criterion shift (30 dB) had been reached at some frequency. Not surprisingly, shifts tended to be highest at 4.0 kHz and above. Because of the pattern of data taking and reporting, it is not possible to make a simple statement regarding the state of the 95 percentile ear for a given number of impulses. However, an interesting observation or two can be made. First, the six-gap impulse was *very* effective at producing threshold shift. One impulse was enough for some subjects to reach the 30-dB criterion and the median subject reached it in four impulses! Second, the three-gap impulse was much less hazardous—the median subject reached the criterion shift with 88 impulses.

The exposures came at 1 impulse/s. Presumably the subjects would have known just when the second and successive impulses were coming and the middle ear muscles were involved. The model showed that each six-gap impulse produced 176.1 ARUs (warned) and 995.4 ARUs (unwarned). AHAAH predicts that the six-gap impulse is indeed hazardous. Thus, one impulse, unwarned, is predicted to produce 44 dB CTS in the 95 percentile ear, and even warned, four impulses would be predicted to produce 34 dB of CTS.

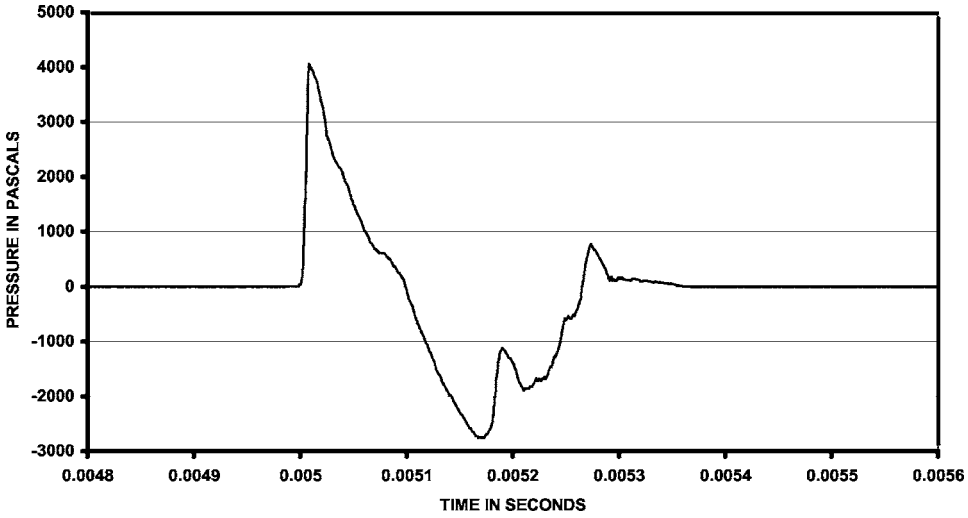


FIG. 10. Pressure history of six-gap impulse used by Loeb and Fletcher (1968).

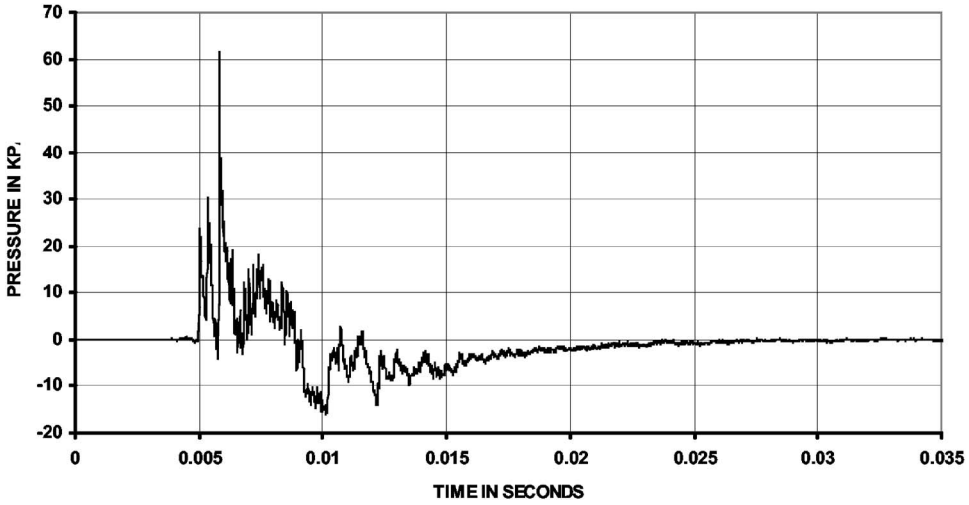


FIG. 11. Pressure history of AT4 rocket firing, gunner's ear position, kneeling. Peak pressure level is 189.8 dB.

This is in keeping with the TTS data from Loeb and Fletcher (1968) in which they report that 0.5 impulses caused a criterion shift in at least one ear. Certainly, AHAAH's predictions are close to what occurred for the six-gap impulse.

The three-gap impulse, for the 95 percentile ear, produced 31.6 ARUs (warned) and 186 ARUs (unwarned). Loeb and Fletcher (1968) report that a criterion shift was reached with as few as 11 impulses for this condition. AHAAH predicts a 15.6-dB of shift for a 95 percentile warned ear. If the first impulse were unwarned, the total becomes 502 ARUs, or 25 dB of CTS.

The close correspondence between the predicted shift and the actual shift is, to a first approximation, a confirmation that the model is doing what it should at its high-frequency end. On the other hand, the data of Fletcher and Loeb showed alot of loss at very high frequencies (8–16 kHz) (as might be expected), but the model predicted somewhat less loss than occurred. Fortunately there are few impulses with such high-frequency energy.

In terms of A-weighted energy, the six-gap impulse contained 2.3 J/M^2 and the three-gap impulse 0.38 J/M^2. Four six-gap impulses at 9.2 J/M^2 fall just over the line and are rated as hazardous. For the three-gap impulse, 11 impulses contain 4.2 J/M^2 and 88 impulses, 33.4 J/M^2. These values are also in the right general area.

g. AT4 exposure (antitank rocket) There is one case in the literature that provides a consistency check on the model (Vause and LaRue, 2001). A soldier in training inadvertently fired one round from an AT-4 rocket launcher from a kneeling position while wearing no hearing protection. He immediately experienced vertigo, tinnitus, and hearing loss so severe that he had to communicate by writing during the audiometric examination the following day. Fortunately for science, his baseline audiogram was available from his entrance physical, and of course full audiometric workups were available from the postexposure examinations. By 21 days postexposure his hearing levels remained at 50–60 dB in the midrange and sad to say, he was released from active duty with the Army because of his hearing disability. A single case study such as this is hardly definitive; but it can be argued that an accurate hazard rating system should predict that this exposure could be very hazardous for at least a susceptible ear. We are also fortunate to have a digitized wave form of an AT-4, recorded during weapons testing (Fig. 11).

1. Analysis with AHAAH. The analysis of the wave form in Fig. 11 calculated a dose of 5774 ARUs in the unwarned ear and 856 ARUs in the warned ear, a hazardous rating in either case. However, it is interesting to note that the higher number of ARUs would be expected to produce about the PTS that in fact resulted. AHAAH appears to be in the right ballpark.

2. Analysis with an A-weighted energy measure. The wave form in Fig. 11 contains 783 J/m^2, which is clearly in the hazardous region.

h. Exposure to a toy horn Finally, a case has been reported by Royster *et al.* (1999) of a 39 month old boy who exposed himself to several "toots" from a bicycle horn which was a small version of the classic British taxi horn (rubber bulb blowing a reed at the base of a coiled brass cone). The horn was capable of producing more than 150-dB peaks in its mouth for about 80 ms. The child immediately complained of pain and tinnitus and at 6 days showed a 50-dB hearing loss (HL) in the exposed ear (HL had been 10 dB 3 months earlier) and at 6 months postexposure he had a 20-dB PTS. The exact details of the exposure are not known (the child was playing with the horn, the parents heard the toots). Fortunately we did have a recording of the horn's sound. This case is not critical for the evaluation of standards, given that it involves an indefinite exposure in one young ear [which might well be more susceptible than an adult ear (Price, 1976)]. On the other hand, it would be unfortunate if a method did not indicate that such a PTS was possible from such an exposure. As reported in Royster *et al.* (1999), AHAAH calculated that several exposure circumstances could have produced the PTS experienced. A-weighted energy also would have considered this noise dangerous (easily having more than 10 J/m^2 in the exposure).

i. Airbag noise exposure Exposure to the noise of an airbag deploying, given accident statistics and the pervasive installation of airbags, seems likely to be an experience everyone will have at some time in their lives. With peak pressures in the 160–170 dB region airbag noise exposures are highly energetic (Price and Kalb, 1999). The AHAAH model has recently been used to predict the incidence of hearing loss that might result in the US population from airbag noise exposure (Price, 2006). This was truly a *pre*diction of incidence, so the data are still developing in the population and the present analysis cannot be definitive with respect to the model's accuracy. However, the existing data are indicative and warrant consideration. The predictions apply to US ve-

hicles, which presently have a more aggressive airbag design than European cars (mandated by the US Congress).

Over all conditions of exposure the AHAAH model predicts some permanent hearing loss in 17% of accidents in which airbags deploy (7% if side airbags are excluded). In contrast, with an A-weighted energy criterion, virtually all airbag deployments would be considered hazardous. Some hearing loss is in fact being reported in the clinical literature (Buckley *et al.*, 1999; Garman, 1998; Huelke *et al.*, 1999; Saunders *et al.*, 1998; Yaremchuk, 1998; Yaremchuk and Dobie, 1999). However there has not been the epidemic of hearing loss that the A-weighted energy measure would indicate, nor perhaps as much as the model indicates. Whether or not the percentages calculated with the AHAAH model are correct is of course an issue that remains to be established. A priori it might be expected that the true hearing loss would be under-reported. Measurement of hearing level is not typically part of emergency room procedure and even if it were, it is unlikely that a pre-exposure measure would be available to establish loss associated with the exposure. Losses of 20 dB or so might not be noticed and reported, especially if they are monaural. Nonetheless, we note that the predictions of the model are consistent with the data, perhaps more so than the prediction from A-weighted energy. It is ironic that the test of the model's predictions may be confounded by the fact that automotive engineers are now using the model to guide airbag design in an effort to make them less hazardous.

III. DISCUSSION AND CONCLUSIONS

A. Comparison of accuracy of the methods

Given all the human exposures reviewed, we conclude that AHAAH has been correct in its predictions in over 95% of the cases, which included a relatively wide range of exposures: protected exposures with up to 100 impulses at over 180 dB under the muff as well as rifle impulses and spark gap exposures with no hearing protection used. In contrast, the traditional methods of hazard analysis have been much less accurate. A-weighted energy was accurate in about 25% of the cases and MIL-STD-1474D was accurate in about 40% of the cases for which it made a prediction. In general, the errors tended to be in the direction of over-rating the hazard and in the case of impulses at very high levels having a lot of low-frequency energy, the errors with the A-weighed energy measure were very large (10–20 dB). The AHAAH model would contend that this type of error is inherent in the traditional methods of analysis because they make no allowance for the middle ear becoming nonlinear at high displacements, peak clipping the wave forms. Given a nonlinear middle ear, it is unlikely that *any* linear weighting scheme would manage to rate the hazard correctly at very high levels.

It is perhaps worth pointing out specifically that no simple adjustment of the A-weighted criterion from 85 dB A (8.7 J/m^2) to some other level solves the problem. The human data support the idea that an A-weighted limit of 8.7 J/m^2 is about right for a few rifle impulses in an unprotected ear (peak level less than 160 dB) when the exposure is controlled by the firer (individual knows when the impulse is coming). However, that level is grossly in error for the Albuquerque data, which includes high-level impulses with a lot of low-frequency content. A safe A-weighted level for those impulses is nearer to an A-weighted 1000 J/m^2 for the 1- and 3-m data; and for the intact muff and the 5-m impulse, the ear can accept A-weighted levels of over 2000 J/m^2 with no unacceptable threshold shift. A-weighting simply misses the mark in these data by more than 20 dB.

B. Prediction of the improbable

One of the tests of a useful theory is that it makes some predictions accurately that existing methods could not or do not handle. There are several cases in which the model has done this. The result for the different impulses just discussed is one such example, i.e., 2000 J/m^2, is safer than 8.7 J/m^2 or in terms of peak pressures, 100 impulses at 173 dBP are safe and five at 157 dBP are hazardous.

A second related example is the unusual outcome that higher peak pressures of an impulse at the same spectral location may not always mean higher risk. The traditional mind-set has understandably held that higher peak pressures must be associated with increased damage potential. In fact, the Albuquerque studies incorporated this assumption into their design and analysis (Chan *et al.*, 2001). Therefore the AHAAH model's specific prediction that the hazard in the 1- and 3-m exposures would *not* increase monotonically with peak pressure (because of a nonlinear middle ear) was controversial (plotted in Fig. 2 of this article). Because the inner cells of the exposure matrix were not fully run, the studies did not provide a definitive test. Not all those that were tested there were susceptible Ss that "dropped" in from higher levels (the experimenters saw that the higher exposure level was clearly hazardous, so they stopped testing there and ran the lower level). There are too few tests to provide certainty; but as we have seen, the data were at least in the direction of the model's prediction of increased hazard at the lower intensity. It is interesting to note that in this context, the data analysis used by Chan *et al.* (2001) (very different from the ear model) produced the same outcome. Their analysis incorporated various adjustments, but it was essentially a correlational comparison between peak pressure and incidence of threshold shift. When they plotted their results, the data points predicted by the AHAAH model to be most hazardous generally stood above the envelopes plotting their 95% confidence levels using different criteria; moreover, in the one case for which there were sufficient data, for the 1-m data at Level 5, losses clearly stood above losses for the Level 6 data (higher peak pressure). The reduced hazard from a higher intensity impulse was thus not a function of a quirk in a particular method of data analysis.

There is another human dataset that fits the improbable outcome of lower risk with higher peak pressures and greater energies. Sommer and Nixon (1973) in analyzing threshold shifts following simulated airbag noise exposures found less threshold shift when the pressure resulting from the buildup in a closed passenger compartment was present. Based on middle ear nonlinearity and analysis of airbag impulses (Price, 2006), the AHAAH model agrees with this outcome.

Finally, in developing the model, the cat ear was tested

with two rifle impulses with the same spectral shape but differing in PPL by about 9 dB and a nearly ten-fold increase in energy. The model's prediction of *equal* hazard (because of details in the pressure history) was in fact borne out (Price, 2003a).

C. Additional features of the AHAAH model

AHAAH also has the advantage of having a theoretical base rather than being empirically developed. This provides at least the hope that it will be more likely to correctly rate the hazard of new impulses with shapes that differ from the ones evaluated in the present report. Furthermore, at least for impulses in the pressure range in this validation, it is capable of rating hazard from essentially all impulses and all exposure conditions. For example, MIL-STD-1474D predicts only for protected ears, and the use of the A-weighted energy criterion is limited to less than 140 dB peak in the US. If the measurements are made with a manikin, hearing protectors of all designs—both linear and nonlinear—may be accommodated by the model as well as with A-weighted energy. Alternatively, a mathematical model of an HPD or a digital filter could be created that reproduces the performance of a protector and the free-field impulse could be processed and the wave form under the protector calculated. The US Army is presently pursuing such an approach.

In addition to providing a numerical rating of hazard, the AHAAH model also generates a "movie" of the action of the sound within the cochlea. This display makes it possible to correlate specific features in the wave form with their damaging action in the cochlea and in the process generate engineering insight into solutions to the problem. For instance, the hazard from one particular airbag wave form was largely a function of the fact that the bag bounced away from the driver's face during deployment (Price, 2006). The reversed motion of the bag produced a momentary reversal of pressure that allowed the middle ear to better conduct the damaging stimulus. Thus, damping the motion of the bag could reduce such a hazard even with the same basic inflation noise. Such an approach would have the benefit of maintaining the effectiveness of the airbag while preventing hearing loss.

D. Why the model's prediction rate is higher than that of A-weighted energy

By making use of the model's movie of the action to follow the action of the impulse in the time domain, it is possible to gain some insight into the features of the model that are key in its success in predicting hazard at high intensities. Certainly the fact that the model reproduces the transfer functions from the free field to the eardrum and the eardrum to stapes displacement is critical. Thus the low and high frequencies are attenuated properly and the midrange is emphasized. The physiological basis for the A-weighting function is in fact largely due to these same transfer functions, hence the AHAAH model and an A-weighted energy measure at lower SPLs agree, to a first approximation.

The second key to the model's success would be the frequency-dependent peak-limiting aspect of the middle ear transfer function at very high intensities. This has a major impact on energy transmission at high SPLs and explains why a simple A-weighted energy measure cannot work accurately over a range of high SPLs. Therefore, the AHAAH evaluation and the energy evaluation match reasonably well at levels typical of small arms and disagree so much when the pressure is higher as in the case of airbags and cannons. Given a peak-clipping middle ear, any linear method that works at one intensity must fail when the level changes significantly.

E. Limits of the AHAAH model

As is the case with any model or theory, one should always seek to know the limits of its application. The AHAAH model was developed to predict the onset of unacceptable risk for the 95 percentile ear from intense sounds at pressures like those in gunfire. But because it is theoretically based, one would suppose that its utility might extend beyond the initial aims.

Is the model too complex to apply in a field setting? The model does encompass a wide range of complexities to include multiple frequency and amplitude nonlinearities operating in the time domain. However, there is no inherent reason, given modern processing technology, that the AHAAH model could not be incorporated into a meter-like device. On the other hand, making good measurements over a wide range of intensities and frequencies remains something of a technical challenge for the transducer world as well as for the technician. There is no substitute for getting an accurate wave form. But once the wave form is acquired, the model provides a calculation relatively free of operator judgment. The current MIL-STD-1474D, for example, requires expert opinion to decide just how a wave form shall be considered or even if it is appropriate to use the standard on it. We conclude that if a good measurement can be obtained, the application of the model is simple, even if the processes within it are complex.

If it is simple to apply, might it serve as the basis for a rating hazard from sounds at all levels? Given the will and with sufficient wit, the model could conceivably be expanded to include whatever physiological processes operate in the ear and could serve as the basis for a method that would work for all levels and patterns of sound exposure. A-weighted energy has been proposed as one such measure (Dancer, 2003), but as we have seen, it fails (and from a theoretical standpoint *must* fail) at high intensities. The fact is that A-weighted energy is not an unqualified success at lower intensities either. Dixon Ward has commented in this regard, "the total energy principle—has survived not because of any compelling evidence, but because of its seductive simplicity. That is, it would be so wonderful if it were true that it is hard to keep from pretending that it is true." So, seduced by simplicity (and practicality), A-weighted energy is now commonly used around the world, at least for SPLs below 115 dB.

What then are the limits of the AHAAH model? As we have seen, it works well at high SPLs (at 150 dB and higher). But how much *lower* in pressure might it also be an

effective hazard measure? There is a shortage of research with the human ear in the 110–150 dB intensity region; therefore testing the validity of any predictive scheme will have to await more work. There is, however, one study done by Schori (1976) that set out to test safe limits for pink noise exposure at high levels. One of the test conditions was 10 s exposure at 130 dB (rms) under an earphone. With 40 subjects tested binaurally, one showed a 20+dB shift in both ears at 3.0 kHz that grew larger with time (TTS 2 min versus TTS 30 min). This delayed recovery is characteristic of losses at higher levels (Hamernik *et al.*, 1988) and taken by Pfander (1975) as a sign of unacceptable risk. There are problems in establishing the wave forms and calibration levels for use with the AHAAH model in evaluating Schori's exposure, but it would be fair to say that these findings are consistent with the model's prediction. It would also predict no loss at 30 min postexposure for Schori's 1 s exposure at 130 dB or to any of the lower level exposures at 125 and 120 dB PPL, which was also what Schori found. So the model is not grossly incompatible with data in the 130-dB region as measured in the external ear. It is interesting to note in this regard that rms pressures at 125–130 dB may have peaks that reach 140 dB, the dividing line commonly taken to set apart "really intense" exposures (MIL-STD-1474, 2007; OSHA, 1995). Furthermore, because of the transfer function of the head and ear canal, free field pressures of 115–120 dB may produce pressures of 130 DB within the ear. Thus there is reason to suspect that the AHAAH model, as it is currently configured, may apply to free field exposures above 115–120 dB.

At present the AHAAH model does not automatically account for changes in latency and efficiency of middle ear muscle contractions. This is not much of a problem at very high intensities where the shortest latency and greatest effect would apply. At lower intensities, however, predictive capacity might improve with the addition of appropriate algorithms for an adaptive and fatiguing middle ear muscle response.

As of this writing, there are no definitive data regarding whether the middle ear muscles in man are conditioned in people firing weapons, i.e., do the middle ear muscles contract in advance of an impending impulse, if the subject is aware that it is coming? In developing the model, the preponderance of the evidence was held to be that such a response exists for the human ear. Loud sounds, electrical stimulation, puffs of air, stimulation of the face, etc., are all capable of eliciting a middle ear muscle response. It is not just an auditory reflex, but is part of a set of facial reflexes. So in the case of intense gunfire, there are a number of unconditioned stimuli (facial stimuli) that can accompany the sound, each of which could elicit a middle ear muscle response. Unconditioned stimuli abound.

Furthermore, several investigators with a variety of paradigms have demonstrated that the human middle ear is also conditionable (Brasher *et al.*, 1969; Djupesland, 1964, 1965; Yonovitz, 1976). Beyond that, the cognitive capacity of the human adds to the probability. For instance, it has been shown that middle ear muscles contracted as subjects *contemplated* handling a toy that was thought to be noisy (Marshall *et al.*, 1975). It is reasonable to expect the middle ear muscles can contract in advance of the auditory stimulus; therefore the model includes the possibility (but does not require it).

The model also includes no recovery processes. It is not much of a problem at high intensities where recovery begins only very slowly (Hamernik *et al.*, 1988); but for much lower intensities, it could be a shortcoming when one contemplates long exposures, especially with breaks in stimulation.

Where the model might be used as a DRC, it should be noted that the level of risk deemed acceptable, 500 ARUs, should be accepted for *occasional* exposures only. The slowed recovery processes associated with high levels of exposure are consistent with the contention that the threshold shifts are underlain by processes analogous to a sprain in the musculo-skeletal system. While one might tolerate occasional exposures to such stimuli, it would clearly be unwise to do so on a daily basis. An exposure of 200 ARUs would be predicted to produce no threshold shift 1/2 h after exposure and should probably be considered the maximum exposure acceptable on a daily basis. More research is needed here.

Finally, when it comes time to make detailed predictions regarding a novel exposure, it is only too apparent how little data there are regarding hearing loss in the human ear. The result is that there are almost never enough data to make any certain assertion. Yet one cannot wait until "all the data are in" before attempting to produce an exposure standard. When considering a standard for gunfire noise exposure, the AHAAH model is based on an immense amount of data compared to earlier formulations. It reproduces the major transfer functions of the ear, which is an assurance that the conductive path to the cochlea is essentially correct. The basic structure of the model, when in a form appropriate to the cat ear, explained the threshold shift and cell loss data very well. And in the form appropriate to the human ear it has been tested with essentially all the available human data. In that role it has achieved a 95% success rate, which is far in excess of the alternate systems. Coupled with its analytical features and basis in theory, we conclude that it is suitable for use as a DRC now, a finding shared by the Society of Automotive Engineers and the US Army.

[1]The model presumes that susceptibility is normally distributed with a 6-dB standard deviation. The model adjusts for susceptibility by asserting that the susceptible ear behaves like a median ear that is being driven by a more intense wave form. Therefore, during the calculation of risk, to achieve a prediction for the 95 percentile ear the pressure of the wave form is automatically increased by 10 dB (1.64 SDs). It follows that if one wished to calculate the risk for the 50 percentile ear, the peak pressure of the wave form should be reduced by 10 dB (using algorithms within the program) and run through the model. In a similar fashion the risk for other centiles can be calculated.

Army Research Laboratory. (**2007**). "AHAAH model, ⟨www.arl.army.mil/ARL-Directorates/HRED/AHAAH/⟩ Last viewed 8/9/2007).

Brasher, P. F., Coles, R. R. A., Elwood, M. A., and Ferres, H. M. (**1969**). "Middle-ear muscle activity and temporary threshold shift," Int. Audiol. **8**, 579–584.

Brinkmann, H. (**2000**). "Supplementary investigation of the German damage

risk criterion with the Belgian NATO small arms rifle FNC," in Report from NATO Research Study Group RSG 29(Panel 8 - AC/243) Reconsideration of effects of impulse noise, TNO-Report No. TM-00-I008, pp. 6–8.
Buck, K. (2000). "Performance of hearing protectors in impulse noise," RTO-ENP-11 AC/323(HFM)TP/31, pp. 3-1–3-10.
Buckley, G., Setchfield, N., and Frampton, R. (**1999**). "Two case reports of possible noise trauma after inflation of airbags in low speed car crashes," Br. Med. J. (Clin Res. Ed) **318**, 499–500.
Chan, P. C., Ho, K. C., Kan, K. K., Stuhmiller, J. H., and Mayorga, M. M. (**2001**). "Evaluation of impulse noise criteria using human volunteer data," J. Acoust. Soc. Am. **110**, 1967–1975.
Dancer, A. (**2000**). "Proposal for a new damage risk criterion," in Report from NATO Research Study Group RSG 29(Panel 8 - AC/243) Reconsideration of effects of impulse noise, TNO-Report No. TM-00-I008, pp. 11–15.
Dancer, A. (**2003**). "The $L_A eq^8$ an effective DRC for weapon noises," NIOSH/NHCA Impulsive noise: A NORA Hearing Loss Team Best Practice Workshop, Cincinnati, OH, ⟨http://www.cdc.gov/niosh/topics/noise/research/impulse_presentations.html⟩ Last viewed 8/9/2007.
Djupesland, G. (**1964**). "Middle ear muscle reflexes elicited by acoustic and nonacoustic stimulation," Acta Oto-Laryngol., Suppl. **188**, 287–292.
Djupesland, G. (**1965**). "Electromyography of the tympanic muscles in man," Intern. Audiol **4**, 34–41.
Fleischer, G., Muller, R., Heppelmann, G., and Bache, T. (**2002**). "Effects of acoustic impulses on hearing," J. Acoust. Soc. Am. **111**, 2335.
Garinther, G. R., and Hodge, D. C. (**1971**). "Small-rocket noise: Hazards to hearing," Tech Mem No. 7-71, U.S. Army Human Engineering Laboratory, Aberdeen Proving Ground, MD.
Garman, J. (**1998**). "The auditory effects of an airbag deployment—An individual's experience," J. Acoust. Soc. Am. **104**(3), 23.
Hamerey, P., Dancer, A., and Evrard, G. (**1997**). "Etude et realization de bouchons d'orielles perfores non lineares," Rapport No. R 128/97, Institut Franco-Allemand de Recherche de St. Louis, 18 Novembre 1997.
Hamernik, R. P., Ahroon, W. A., and Patterson, J. H., Jr. (**1988**). "Threshold recovery functions following impulse noise trauma," J. Acoust. Soc. Am. **84**, 941–950.
Hodge, D. C., and Garinther, G. R. (**1970**). "Validation of the single-impulse correction factor of the CHABA impulse-noise damage-risk criterion," J. Acoust. Soc. Am. **48**, 1429–1430.
Hodge, D. C., Gates, H. W., Solderholm, R. B., Helm, C. P. Jr., and Blackmer, R. F. (**1964**). "Preliminary studies on the impulse-noise effects on human hearing (Project HUMIN)," Tech Mem. No. 15-64, U.S. Army Human Engineering Laboratory, Aberdeen Proving Ground, MD.
Hodge, D. C., and McCommons, R. B. (**1966**). "Further studies of the reliability of temporary threshold shift from impulse-noise exposure," Tech. Mem. No. 3-66, U.S. Army Human Engineering Laboratory, Aberdeen Proving Ground, MD.
Hodge, D. C., McCommons, R. B., and Blackmer, R. F. (**1965**). "Reliability of temporary threshold shifts caused by repeated impulse-noise exposures," Tech. Mem. No. 3-65, U.S. Army Human Engineering Laboratory, Aberdeen Proving Ground, MD.
Huelke, D. F., Moore, J. L., Compton, T. W., Rouhana, S. W., and Kileny, P. R. (**1999**). "Short communication: Hearing loss and airbag deployments," Accid. Anal Prev. **31**(6), 789–792.
Johnson, D. (**1994**). "Blast overpressure studies with animals and men: A walk-up study," USAARL Contract Report No. 94-2, U.S. Army Aeromedical Research Lab, Ft. Rucker, AL.
Johnson, D. L. (**1998**). "Blast overpressure studies," USAARL Contract Report No. CR-98-03, U.S. Army Aeromedical Research Laboratory, Ft. Rucker, AL.
Johnson, D. L., Patterson, J. D., Nelson, W. R., Ripple, G., Mundie, T. G., Christensen, W. I., and Bova, C. M. (**1990**). "Direct determination of occupational exposure limits for freefield impulse noise," *Protocol for study*, US Army Med R&D Command, Washington, D.C., Vol. **III**.
Kalb, J. T., and Price, G. R. (**1987**). "Mathematical model of the ear's response to weapons impulses," Proceedings of the 3rd Conference on Weapon Launch Noise Blast Overpressure, U.S. Army Ballistics Research Lab, Aberdeen Proving Ground, MD.
Kardous, C. A., Franks, J. R., and Davis, R. R. (**2005**). "NIOSH/NHCA Best-Practices Workshop on impulsive noise," Noise Control Eng. J. **53**, 5353–5361.
Loeb, M., and Fletcher, J. L. (**1968**). "Impulse duration and temporary threshold shift," J. Acoust. Soc. Am. **44**, 1524–1528.
Marshall, L., Brandt, J. F., and Marston, L. E. (**1975**). "Anticipatory middle-ear reflex activity from noisy toys," J. Speech Hear Disord. **40**(3), 320–326.
MIL-STD-1474D. (**2007**). "Department of defense design criteria. Noise limits," ⟨http://hfetag.dtic.mil/docs-hfs/mil-std-1474d.pdf⟩ Last viewed 8/9/2007.
NATO Research Study Group RSG6/PANEL8. (1987). "The effects of impulse noise," Document No. AC/243/(PANEL8/RSG.6)D/9, NATO, 1110 Brussels, Belgium.
NATO Research Study Group RSG.29(Panel 8 - AC/243). (**2000**). "Reconsideration of effects of impulse noise," TNO-Report No. TM-00-I008.
NATO RTO-TR-017. (**2003**). "Reconsideration of the effects of impulse noise," RTO Technical Report No. TR-017/HFM-022, ISBN 92-837-1105x.
OSHA (**1995**). 1910.95 Occupational Noise Exposure.
Patterson, J. D., and Ahroon, W. A. (**2005**). "Evaluation of an auditory hazard model using data from human volunteer studies," USAARL Report No. 2005-01, U. S. Army Aeromedical Research Laboratory, Fort Rucker, AL.
Patterson, J. J. Jr., and Johnson, D. L. (**1994**). "Temporary threshold shifts produced by high intensity freefield impulse noise in humans wearing hearing protection," USAARL Rept. No. 94-46, U. S. Army Aeromedical Research Laboratory, Ft. Rucker, AL.
Patterson, J. J. Jr., Mozo, B. T., Gordon, E., Canales, J. R., and Johnson, D. L. (**1997**). "Pressures measured under earmuffs worn by human volunteers during exposure to freefield blast overpressures," USAARL Report No. 98-01, U.S. Army Aeromedical Research Laboratory, P.O. Box 620577, Ft. Rucker, AL.
Pfander, F. (**1975**). "*Das Knalltrauma*," ("Impulse noise damage") (Springer-Verlag, Berlin).
Price, G. R. (**1976**). "Age as a factor in susceptibility to hearing loss: Young versus adult ears," J. Acoust. Soc. Am. **60**, 886–892.
Price, G. R. (**1998a**). "Airbag noise hazard: from theory toward validation," J. Acoust. Soc. Am. **104**, 1769.
Price, G. R. (**1998b**). "Standards for damage risk for impact/impulse noise," Proceedings of 23rd Annual NHCA Hearing Conservation Conference, Albuquerque, N.M.
Price, G. R. (**2003a**). "An examination of and response to "auditory standard issues" by Dr. James Stuhmiller," ⟨www.arl.army.mil/ARL-Directorates/HRED/AHAAH/⟩, Army Research Lab. Last viewed 8/9/2007.
Price, G. R. (**2003b**). "Impulse noise and the cat cochlea," ⟨http://www.arl.army.mil/ARL-Directorates/HRED/AHAAH⟩, Army Research Lab. Last viewed 8/9/2007.
Price, G. R. (**2005**). "A new method for rating hazard from intense sounds: Implications for hearing protection, speech intelligibility and situation awareness," Keynote 2 in NATO RTO-MP-HFM-123 Symposium "New directions for improving audio effectiveness," April 11–13, 2004, Amersfoort, The Netherlands, ISBN 92-837-1147-5 Available from ⟨http://www.rto.nato.int⟩ Last viewed 8/9/2007.
Price, G. R. (**2006**). "Insights into hazard from airbag noise gained through the AHAAH model," Paper No. 2005-01-2397, Proceedings SAE 2005 Transactions Journal of Passenger Cars: Mechanical Systems, February, Book V114-6, ISBN 0-7680-1692-4.
Price, G. R. (**2007**). "Predicting mechanical damage to the organ of Corti," In *Pharmacological Strategies for Prevention and Treatment of Hearing Loss and Tinnitus Hearing Res.*, Vol. 226, 5–13. ⟨http://dx.doi.org/10.1016/j.heares.2006.08.005⟩ Last viewed 8/9/2007.
Price, G. R., and Kalb, J. T. (**1991**). "Insights into hazard from intense impulses from a mathematical model of the ear," J. Acoust. Soc. Am. **90**, 219–227.
Price, G. R., and Kalb, J. T. (**1998**). "A new approach: The auditory hazard assessment algorithm (AHAA)," Talk to International Conference on Biological Effects of Noise—Australia, Conference Programme and Abstract Book, p. 127.
Price, G. R., and Kalb, J. T. (**1999**). "Auditory hazard from airbag noise exposure," J. Acoust. Soc. Am. **106**, 2629–2637.
Royster, J. D., Royster, L. H., Price, G. R., McMillan, P. M., and Kileny, P. R. (**1999**). "Hazard analysis for a bicycle horn which produced acoustic trauma," Proceedings 24th Annual NHCA Hearing Conservation Conference Proceedings, Atlanta, GA, 25–27 February.
Society of Automotive Engineers (SAE) Standards. (**2003**). "J2531: Impulse noise from automotive inflatable devices," ⟨www.sae.org⟩ Last viewed 8/9/2007.
Saunders, J. E., Slattery, W. H. III, and Luxford, W. M. (**1998**). "Automobile airbag impulse noise: Otologic symptoms in six patients," J. Acoust. Soc.

Am. **104**, 1769.
Schori, T. R. (**1976**). "Evaluation of safe exposure guidelines for moderate and high intensity continuous noise," Aerospace Medical Research Lab Rep. No. AMRL-TR-76-97.
Sommer, H. C., and Nixon, C. W. (**1973**). "Primary components of simulated air bag noise and their relative effects on human hearing," Final Report, DOT Contract No. IA-0-1-2160, Washington, D.C.
Vause, N. C., and LaRue, A. A. (**2001**). "Creating our own casualties - Auditory effects of anti-tank weapons fire without hearing protection - A clinical case study," Proceedings International Military Noise Conference, Baltimore, MD, April 24–26.
Yaremchuk, K. (**1998**). "Otologic effects of airbag deployment," J. Acoust. Soc. Am. **104**, 1770.
Yaremchuk, K., and Dobie, R. A. (**1999**). "The otologic effects of airbag deployment," J. Occup. Hear. Loss **2**(2–3), 67–73.
Yonovitz, A. (**1976**). "Classical conditioning of the stapedius muscle," Acta Oto-Laryngol. **82**, 11–15.

Evidence specifically favoring the equalization-cancellation theory of binaural unmasking

John F. Culling[a)]
School of Psychology, Cardiff University, Tower Building, Park Place Cardiff, CF10 3AT, United Kingdom

(Received 14 November 2006; revised 23 August 2007; accepted 23 August 2007)

Three experiments investigated the roles of interaural correlation (ρ) and of the monaural power spectrum in the detection and discrimination of narrow-band-noise signals (462–539 Hz) in broadband maskers (0–3 kHz). The power and ρ of the target band were independently controlled, while the flanking noise was fixed and diotic. Experiments 1 and 2 involved ρ and power values that would be produced by specific values of signal-to-noise ratio (SNR) in the NoSπ binaural configuration. Listeners were required to discriminate different SNRs via a 2I-FC loudness-discrimination task. At low reference SNRs, changes in ρ fully accounted for listeners' performance, but as reference SNR increased, additional energy in the target band played an increasing role. Experiment 2 showed that at these higher SNRs the combination of information from the power spectrum and ρ was superadditive and could not be explained by simple signal-detection models. The equalization-cancellation (EC) theory would explain these data using the output from interaural cancellation, Y, rather than ρ. Experiment 3 attempted to foil binaural processing, by fixing either ρ or Y across intervals. Consistent with EC theory, when Y was fixed, the contribution of the binaural system appeared negligible, while fixing ρ did not have this effect. [DOI: 10.1121/1.2785035]

PACS number(s): 43.66.Pn, 43.66.Ba, 43.66.Dc [RLF] Pages: 2803–2813

I. INTRODUCTION

Binaural unmasking occurs when a signal is presented in noise with an interaural configuration that differs from that of the noise. The largest binaural unmasking effect occurs in the condition known as NoSπ, where a signal is presented interaurally out of phase against a noise which is interaurally in phase (Hirsh, 1948). At low signal frequencies (200–500 Hz) in broadband masking noise, NoSπ will yield masked thresholds approximately 15 dB lower than NoSo, where both signal and noise are in phase. This article considers two theoretical frameworks that have been used to explain this effect.

Durlach (1963, 1972) proposed the equalization-cancellation (EC) theory, in which the binaural system can apply a range of transformations to the noise at one ear in order to match it to the noise at the other ear, before subtracting one from the other.[1] More recent versions of the EC theory (Durlach, 1972; Culling and Summerfield, 1995; Bareebaart *et al.* 2001a, b, c) have used a restricted set of transformations that include only delay and attenuation. EC theory predicts that NoSπ is the optimal configuration because no transformation is needed to optimally equalize (and so cancel) the diotic noise, while the subtraction process will quadruple the signal power.

In the second framework, the waveforms at the two ears are effectively cross-correlated (Osman, 1971; Colburn, 1973, 1977; Durlach *et al.*, 1986; Jain *et al.*, 1991; Culling *et al.*, 2001, 2006). Signals are detected when the cross-correlation maximum (the "coherence") is reduced by the presence of the signal. Once again, where the masking noise is diotic, internal delays are not needed, so the maximum in the interaural cross-correlation is equal to the correlation of the left- and right-ear waveforms. Listeners are very sensitive to small reductions in interaural correlation (ρ) from unity, but less so when the reference correlation is lower (Culling *et al.*, 2001; Boehnke *et al.*, 2002).

The series of experiments reported here yield two main findings. First, when one examines the idea that ρ is used by the brain as the key decision variable in the detection of masked signals, it can fail at suprathreshold ratios of signal-to-noise power (SNRs). Second, if one directly compares the predictions of EC theory to the predictions of correlation models at these moderate SNRs, EC theory is correct in predicting when stimuli will be indistinguishable by the binaural system, while correlation models are not. Taken together, these results constitute relatively strong psychophysical evidence for a cancellation mechanism in binaural unmasking.

The experiments were directly motivated by two previous studies. Jain *et al.* (1991) provided evidence in support of the idea that the brain uses changes in ρ as a cue for detecting signals in noise and that the phenomenon of binaural unmasking is mediated by this cue. Among other experiments, Jain *et al.* manipulated ρ in a narrow band of noise embedded within a broadband diotic white noise. They compared the resulting detection of deviations in ρ from unity with detection thresholds for narrow-band noises added to broadband noise in the NoSπ configuration. The two were related from theory using Eq. (1), which gives ρ for the addition of an antiphasic target signal to diotic noise (i.e., NoSπ) at any SNR,

[a)]Electronic mail: cullingj@cf.ac.uk

$$\rho(\mathrm{NoS}\pi) = \frac{1 - \mathrm{SNR}}{1 + \mathrm{SNR}}. \quad (1)$$

By use of Eq. (1) they were able to plot detection rates for the two types of stimuli against a common axis. Since the psychometric functions were very similar, they concluded that binaural detection and discrimination of ρ were most likely produced by the same mechanism. Rearranging the equation gives the SNR equivalent to a given value of ρ:

$$\mathrm{SNR}(\mathrm{NoS}\pi) = \frac{1 - \rho}{1 + \rho}. \quad (2)$$

The second study that led to this one was largely inspired by the first. Culling *et al.* (2006) took the idea that interaural correlation underpins signal detection and extended it to the recognition of connected speech, which also displays a binaural unmasking effect (Licklider, 1948). They reasoned that if detection of signals is facilitated by detection of changes in ρ, then discrimination of different speech sounds, and hence speech understanding, would also be. While speech understanding in noise is improved by favorable binaural configurations, there is something of a mismatch between the frequencies that produce the largest binaural unmasking effect in broadband noise and the frequencies most important to speech understanding. Hirsh and Burgeat (1958) found that binaural unmasking declines almost linearly with frequency before asymptoting to about 3 dB (for NoSo versus NoSπ) at 1500 Hz, whereas Fletcher and Galt (1950) found the most important frequencies for speech understanding to be centered around 2000 Hz. Although speech in quiet can be understood when low-pass filtered at 1500 Hz, it seemed likely that listeners would not be able to rely solely on binaural cues to understand speech. Consequently, Culling *et al.* (2006) examined the roles of both monaural spectral information and the binaural correlation changes. In each case, speech in noise was simulated by modulating the interaural correlation and spectral power. These two cues were investigated independently by analyzing the fluctuations in intensity and ρ within each frequency band of stimuli created by adding speech to noise[2] in NoSπ. The resulting measurements were then used to modulate the intensity and/or ρ of an independent noise, creating simulations of the original speech with orthogonally controlled cues. The results showed that neither individual cue was very effective at facilitating speech understanding at SNRs where the original NoSπ stimuli were 50% intelligible. However, when both were combined a large increase in performance was observed.

Among a number of alternative explanations for the data of Culling *et al.*, one possibility was that there is an interaction between monaural and binaural information. The present experiments explore this possibility using much simpler stimuli.

II. EXPERIMENT 1

Experiment 1 measured listeners' ability to discriminate different levels of correlation for a target band of noise embedded within a broadband noise (0–3 kHz) compared to a reference level of correlation, ρ_R. For $\rho_R = 1$, the experiment was a direct replication of conditions "WB Det NoSo," "WB Det NoSπ," and "Fringed Corr-No" (termed here the NoSo, NoSπ, and Corr conditions) from experiment 1 in the article of Jain *et al.* (1991). The only difference is that, here, target bands of noise with a bandwidth of 1 ERB (Moore and Glasberg, 1983) were used rather than 1/10 octave centered on 500 Hz (i.e., 77 Hz rather than 35 Hz wide). The NoSo and NoSπ conditions involved the addition of a 1-ERB-wide target band of noise with appropriate binaural phase to a broadband noise (0–3 kHz). The Corr condition involved the manipulation of ρ within the target band. The other levels of ρ_R extended this detection paradigm into a discrimination paradigm, where the reference stimuli contained cues that were similar, but of lesser magnitude than the cues in the target stimuli. That is, in NoSo and NoSπ, the reference condition has an added band of noise of appropriate interaural phase whose SNR was related to ρ_R by Eq. (2).

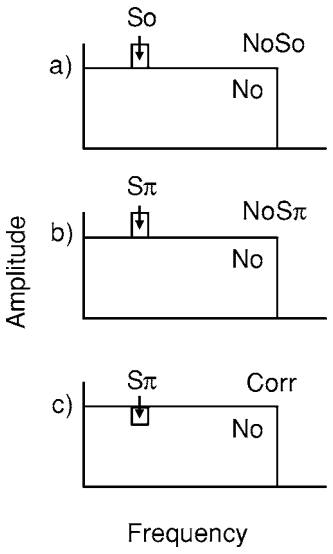


FIG. 1. Schematic diagram of the spectral composition of the stimuli used in Experiments 1 and 2.

It is noteworthy that the manipulation of ρ within a target band in the Corr condition can be achieved by adding a noise at each ear with non-unity interaural correlation to a correlated noise with a complementary dip in its power spectrum. Figure 1(c) illustrates this way of viewing the stimuli for the addition of anticorrelated noise (i.e., with a ρ of -1). If the added noise is interaurally uncorrelated, the scale factor, u, that must be applied to the uncorrelated portion of the noise within the target band to achieve a target correlation, ρ, is given by Eq. (3), while the scale factor for the correlated noise within that band, c, is given by Eq. (4):

$$u = \sqrt{1 - \rho}, \quad (3)$$

$$c = \sqrt{\rho}. \quad (4)$$

Equivalent stimuli can be created by adding interaurally anticorrelated noise. Here, the scale factor for the anticorrelated noise, a, is given by Eq. (5) and that for the correlated noise, c, by Eq. (6):

TABLE I. Values of target interaural correlation, ρ_T, used in Experiment 1 with the given values of reference interaural correlation, ρ_R. Numbers in parentheses are the equivalent SNRs derived using Eq. (2).

ρ_R (SNR dB)	ρ_T (SNR dB)					
1 (−∞)	0.98 (−20)	0.95 (−15.9)	0.9 (−12.8)	0.8 (−9.5)	0.5 (−4.8)	0 (0)
0.9 (−12.8)		0.8 (−9.5)	0.7 (−7.5)	0.5 (−4.8)	0.3 (−2.7)	0 (0)
0.8 (−9.5)		0.7 (−7.5)	0.6 (−6)	0.5 (−4.8)	0.3 (−2.7)	0 (0)
0.7 (−7.5)		0.6 (−6)	0.5 (−4.8)	0.4 (−3.7)	0.3 (−2.7)	0 (0)
0.6 (−6)		0.5 (−4.8)	0.4 (−3.7)	0.3 (−2.7)	0.2 (−1.8)	0 (0)

$$a = \sqrt{\frac{1-\rho}{2}}, \tag{5}$$

$$c = \sqrt{\frac{1+\rho}{2}}. \tag{6}$$

Effectively, therefore, the NoSo and NoSπ conditions are formed by adding an additional band of noise, producing a prominence in the spectum [Figs. 1(a) and 1(b)], while the Corr condition is formed by replacing part of the masking noise, leaving the spectrum flat.

A. Stimuli

The stimuli were generated using MATLAB and played to the listeners at a sampling rate of 44.1 kHz with 16 bit quantization. Gaussian noise samples of 500 ms duration were generated, filtered, scaled, and mixed digitally. Filtering was performed by applying a discrete Fourier transform, setting the amplitude of unwanted frequency bins to zero and then applying an inverse Fourier transformation.

The NoSo stimuli were formed by the addition to a broadband diotic noise of a band of interaurally correlated (i.e., diotic) noise, bandpass filtered from 462 to 539 Hz. The NoSπ stimuli were formed by the addition of an anticorrelated noise of the same bandwidth. The target band of the Corr condition was formed by mixing correlated and uncorrelated bands of noise, according to Eqs. (3) and (4), and inserting them into a complementary notch in the correlated flanking noise. All stimuli were low-pass filtered at 3 kHz and gated with 10 ms raised-cosine onset-offset ramps.

B. Procedure

Three listeners participated in the study. These were the author and two naive listeners who were paid for their participation. In each condition, psychometric functions were measured for a range of target values of ρ, ρ_T, against a higher reference value of ρ, ρ_R. The values of ρ_R were 1.0, 0.9, 0.8, 0.7, and 0.6. These values are equivalent to progressively higher SNRs of −∞, −12.8, −9.5, −7.5, and −6 dB. The set of target values of ρ, ρ_T was adjusted such that the target always had a lower correlation (Table I).

Each listener attended two 1 h experimental sessions for each of the five values of ρ_R. During each session the listener completed either 18 or 15 50-trial blocks. Each block was used to measure one point on a psychometric function of one of the three conditions. At the beginning of each block, listeners were presented with the "target sound," which was an isolated diotic noise, bandpass filtered to the same bandwidth as the target band of the stimuli to follow. The sequence of blocks was selected at random. The complete psychometric function for all three conditions was collected during each session, but the results were averaged over the two sessions.

Each trial consisted of two randomly ordered presentation intervals. In one interval, a reference stimulus was presented using ρ_R in the Corr condition or its equivalent SNR in the NoSo and NoSπ conditions. In the other interval a stimulus with lower value of ρ (or higher value of SNR) was presented. Listeners were required to select the interval containing the target signal, or the interval in which the target signal was loudest. In practice, the perceptual salience of the reference stimuli was such that the former instruction applied for ρ_R=1 or 0.9, while the latter applied for lower values. Listeners responded using the mouse by clicking on a button that appeared on the computer monitor at the onset of the corresponding interval. The presentation level was 57 dB (A). Trial-by-trial feedback was given.

C. Results

The results of Experiment 1 are shown in Figs. 2–6. Each figure shows the psychometric functions for the three individual listeners and their mean on separate panels for one specific value of ρ_R. In order to facilitate comparisons across figures, axis limits have not been adjusted from one figure to the next. The axis in decibels at the bottom of each figure is derived using Eq. (2) from the correlation values that were used to make the stimuli. Corresponding interaural correlation values are marked on the top axis. Increasing interaural correlation corresponds to decreasing SNR.

The results for ρ_R=1 (Fig. 2) replicate those found by Jain *et al.* (1991), in that the psychometric functions for NoSπ and NoSo differ by 16 dB, while those for Corr and NoSπ are roughly the same. It seems likely, given the results for other values of ρ_R, that the apparent difference in slope for the fitted logistic functions is a product of experimental noise.

The results for ρ_R=0.9, 0.8, 0.7, and 0.6 (Figs. 3–6) extend those of Jain *et al.*. Across the four figures, the data are progressively clustered toward the right–hand side of the figure because the higher values of reference SNR makes it inappropriate to present target stimuli of lower SNR (see Table I). Figure 3 shows that, with ρ_R=0.9 (reference SNR =−12.8 dB), the data follow a similar pattern as for ρ_R=1, except that the difference between the NoSo and NoSπ con-

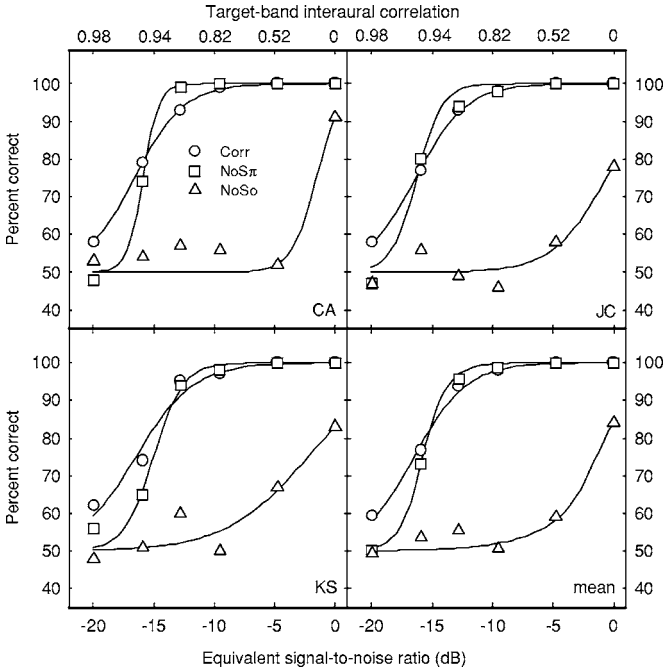


FIG. 2. Results from Experiment 1 using a reference interaural correlation of 1.0. The first three panels each show psychometric functions for the three conditions for one of the three listeners, and the fourth (bottom right) shows the mean functions across all three listeners. Symbols are scores in each condition at each level of target-band interaural correlation. Lines are logistic curves fitted to these data.

ditions (the binaural unmasking effect) has reduced to about 10 dB. The Corr and NoSπ conditions coincide almost exactly with no apparent difference in slope.

Figure 4 shows that, with ρ_R=0.8, a difference in pattern begins to emerge. The NoSo and NoSπ conditions still differ by about 7 dB, but there is a small difference between the NoSπ and Corr conditions at each value of ρ_T, arising from

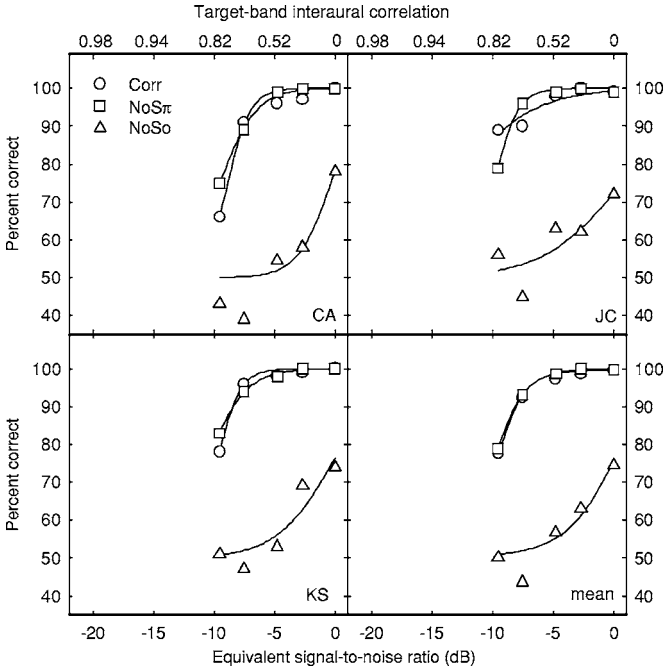


FIG. 3. Similar to Fig. 2, but using a reference interaural correlation of 0.9 (corresponding to −12.8 dB SNR in the NoSπ condition).

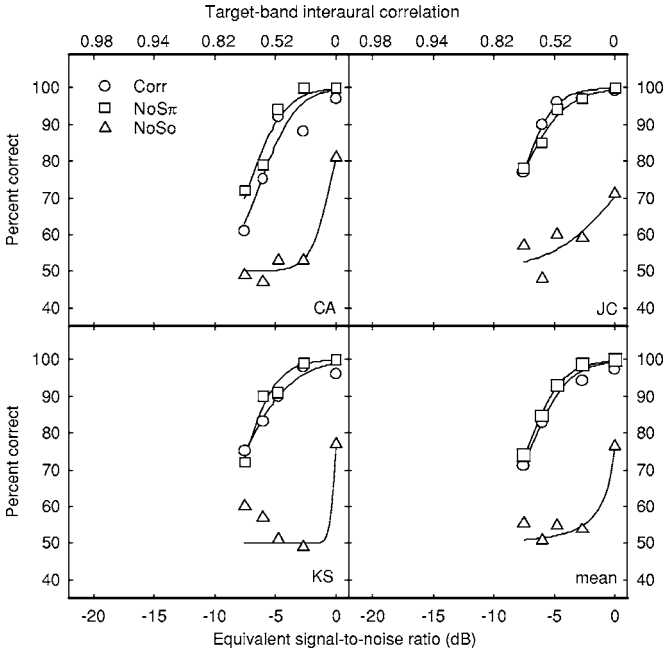


FIG. 4. Similar to Fig. 2, but using a reference interaural correlation of 0.8 (corresponding to −9.5 dB SNR in the NoSπ condition).

the data of listeners CA and KS. From the fitted logistic functions, these differences are equivalent to 0.3 dB at the 75% point.

Figure 5 shows the same effect at ρ_R=0.7, but more consistently across all three listeners; the Corr score is lower than the NoSπ score for each listener at each value of ρ_T (bar one where they are equal). These differences are therefore significant by a sign test (p=0.0001). From the fitted logistic functions, the differences are equivalent to 1 dB at the 75%

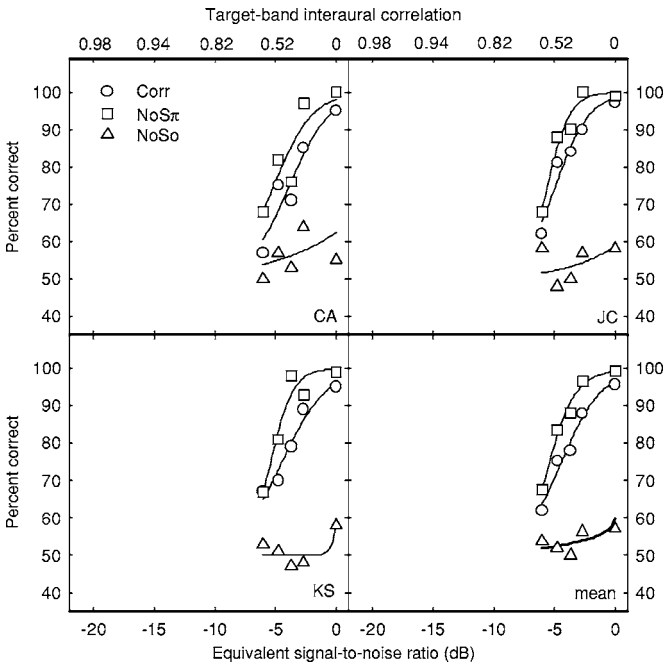


FIG. 5. Similar to Fig. 2, but using a reference interaural correlation of 0.7 (corresponding to −7.5 dB SNR in the NoSπ condition).

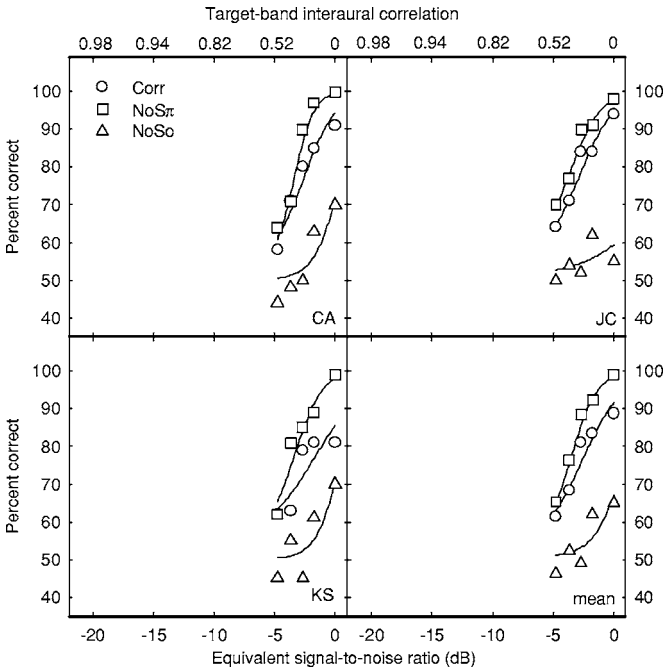


FIG. 6. Similar to Fig. 2, but using a reference interaural correlation of 0.6 (corresponding to −6.0 dB SNR in the NoSπ condition).

point. The binaural unmasking effect is difficult to evaluate, because the scores in NoSo became too low, though clearly it still exists.

Finally, Fig. 6 shows a perhaps larger difference between the Corr and NoSπ conditions for ρ_R=0.6, which again occurs for each listener at each value of ρ_T (bar two where they are equal). These differences are also significant by a sign test (p=0.0002). From the fitted logistic functions, these differences are again equivalent to 1 dB at the 75% point, although the difference now appears to increase with target SNR for two of the listeners and so may be larger at higher threshold levels.

D. Discussion

The data show that the equivalence of the NoSπ and Corr conditions is dependent upon reference SNR. At low reference SNRs (high ρ_R) the results of these two conditions are identical, whereas at higher reference SNRs their results progressively diverge, such that NoSπ discriminations are always easier. Given that for ρ_R=1, the reference stimuli are identical in these two conditions (both are diotic noise), and assuming that the listeners solve the task principally on the basis of loudness of the antiphasic signal, this must mean that the NoSπ and Corr stimuli are equally loud at low SNRs, but that the NoSπ stimuli increase in loudness more rapidly with increasing SNR than the Corr stimuli.

This outcome might be expected given that there are two cues contributing to the loudness of the NoSπ condition (both ρ and the intensity of the target band), both of which would be expected to increase loudness with SNR. However, it is noteworthy that the listeners showed no tendency to improve their performance in the NoSo condition at higher reference SNRs; their ability to detect changes in intensity within the target band remains very poor compared to their ability to discriminate the different SNRs in the other conditions. It may therefore be difficult to account for the increasing difference between the NoSπ and Corr conditions at higher reference SNRs on the basis of increased effectiveness of the intensity-change cue, apparently indicating an interaction between monaural and binaural cues.

Culling *et al.* (2006) also found an apparent interaction between monaural and binaural cues to speech understanding in noise, whereby the combination of both cues gave rise to much higher word recognition rates than either cue in isolation. The interaction seems to occur only at relatively high SNRs, but since speech understanding requires not just the detection of speech energy, but relative judgment of energy across time and frequency, discrimination of different moderate SNRs is probably a central aspect of the task (Culling and Colburn, 2000; Culling *et al.*, 2001). The plausibility of such an interaction was explored further in Experiment 2 using stimuli similar to Experiment 1, but concentrating on the difference between the NoSπ and Corr conditions at relatively high SNRs.

However, before moving on to Experiment 2, it is worth noting that the observed outcome of Experiment 1 is also quite consistent with EC theory. Using the conceptual model in Fig. 1, the (effective) amount of anticorrelated noise which has been added within the target band in the NoSπ condition is greater than that in the equivalent Corr condition. Consequently, when a subtractive operation is applied, a larger residue from cancellation will occur in the NoSπ condition than in the Corr condition. Moreover, the difference in residue will increase with (equivalent) SNR, predicting, without recourse to an interactive process, that the difference between NoSπ and Corr should become greater at suprathreshold levels.

With a few simplifying assumptions, it is possible to further evaluate this interpretation of the results. EC theory predicts that decisions are made on the basis of a decision variable, Y, that is derived from an interaural cancellation process [a waveform $Y(t)$ is derived by the cancellation of the left- and right-ear signals, and the power, Y, of this waveform acts as the decision variable]. We will assume here that, for the stimuli in question, Y is simply proportional to the power of the difference signal, $R(t)-L(t)$, when the left- and right-ear waveforms are subtracted (an assumption that holds true if no equalization operation is required and internal noise is negligible). Y is thus proportional to the variable a^2 from Eq. (5). Further, we will assume that discriminations on the basis of this decision variable are subject to Weber's law, implying that changes in Y that are equal in decibels are equally detectable. Finally, we will also assume that any direct sensitivity to interaural correlation will also follow Weber's law for the equivalent SNR produced by Eq. (2) (cf. Culling *et al.*, 2001). Now, we can compare the results from the Corr condition of Experiment 1 with the results from the NoSπ condition using either the decibel change in $Y(\delta Y)$ between one presentation interval and the other or the decibel change in equivalent SNR (δSNR).

The results of this comparison are shown in Fig. 7 for four reference correlations. The open circles are a replot of the Corr condition (means) from Figs. 3–6 against δSNR.

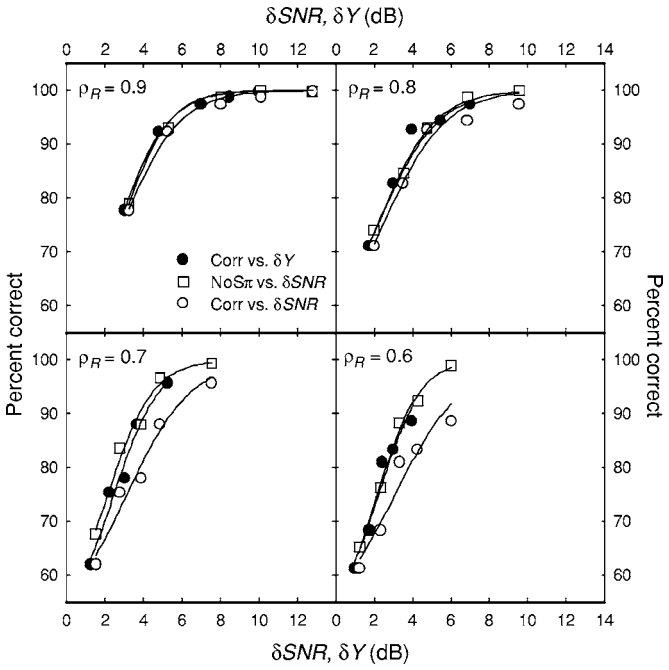


FIG. 7. Results from Experiment 1. NoSπ and Corr conditions replotted as a function of the difference in decibels between the cue in the target and the reference intervals for four levels of ρ_R. Data from the Corr condition is plotted twice, once against the change in equivalent SNR (δSNR) as open circles, and once against the change in $Y(\delta Y)$ as closed circles. Data from the NoSπ condition are plotted as a function of δSNR only. Solid curves are logistic fits similar to Figs. 2–6.

The closed circles are a replot of the same data against δY. For the NoSπ condition (open squares). δSNR and δY are equal (because the diotic noise level within the target band is fixed), and so are only plotted once. For ρ_R=0.8 and 0.9 differences between the three curves are negligible, but for ρ_R=0.6 and 0.7, there is a close agreement between the psychometric functions for Corr and NoSπ when they are plotted against δY, whereas the Corr condition does not agree well when plotted against δSNR. These results suggest that, given the above-mentioned assumptions, EC theory agrees well with the results of Experiment 1, without any separate consideration of a role for monaural detection of intensity change.

III. EXPERIMENT 2

Supposing that Experiment 1 has revealed an interaction between monaural and binaural cues in the discrimination of different signal levels, Experiment 2 set out to characterize that interaction. Signal detection theory offers two typical scenarios for the interaction of multiple cues. In the case of completely independent cues with independent noise (as might typically occur in a cross-modal detection task), d' values derived from each cue in isolation should combine via an orthogonal vector sum (i.e., Pythagoras' theorem) to predict the d' value when both cues are presented simultaneously (Macmillan and Creelman, 1991, p. 257). On the other hand, when two cues are sufficiently related that their detection is limited by one common noise, one would expect the sum of their individual d' values to predict the d' observed when both cues are presented (Macmillan and Creelman, 1991, p. 210). Between these two extremes one may get intermediate results if the two cues are only partially independent.

Experiment 2 set out to characterize the interaction by measuring d' for the monaural and binaural cues independently and comparing with d' for their combined effect. That is, pairs of correlation values were selected using the results of Experiment 1 to produce a large difference between the Corr and NoSπ conditions. The listeners' ability to discriminate these correlation values was then measured in terms of d' and discrimination in the NoSπ condition was compared with the sum of the two d' values in the other two conditions.

A. Stimuli

Two pairs of correlation values were selected for discrimination. These pairs were 0.2 vs 0.6 and 0.3 vs 0.7. The Corr, NoSπ, and NoSo conditions corresponding to these correlations were generated and presented in the same way as those of Experiment 1. That is: in the Corr condition, Eqs. (3) and (4) were used in order to produce these values of correlation in the target band; in the NoSπ condition, SNR was set using Eq. (2) to produce the same target-band correlations as the Corr condition; in the NoSo condition, the SNR was the same as in the NoSπ condition.

B. Procedure

The same three listeners attended two 1 h testing sessions. During each session, they completed 12 blocks of 50 trials. In each session, there were two blocks for each of the two pairs of correlations in each of the three conditions. Thus, there were a total of four blocks for a given combination of condition and correlation pairing. The d' values derived from each of these four blocks were averaged to give the corresponding "raw" d' value for each listener.

C. Results

Figure 8 shows the results for each of the three listeners shown as different symbols. The left-hand side of each panel shows the raw d' values in each of the three conditions, while the right-hand side (separated by the dashed vertical line) shows the summed combinations of the NoSo and Corr d' values. For each pair of correlation values, it is clear that neither form of summation can generate equivalent d' values to the NoSπ condition.

D. Discussion

The experiment set out to characterize how monaural cues (represented by NoSo) and binaural cues (represented by the Corr condition) are combined in the NoSπ condition. However, the results showed that neither form of summation of the monaural and binaural cues was able to account for the NoSπ data. From a strictly numerical point of view, the reason for this is that the d' values from the NoSo condition were so small that, whatever form of addition was used, they

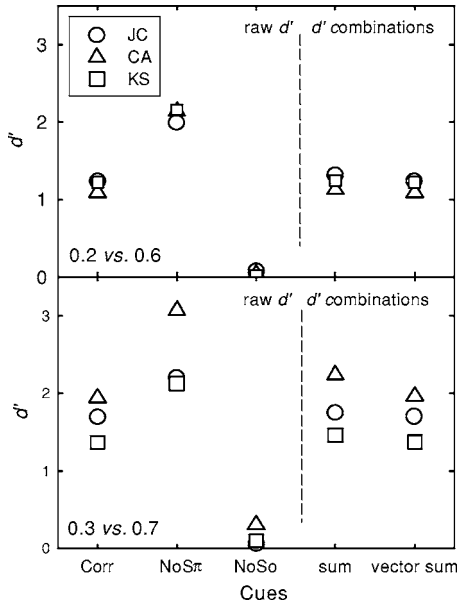


FIG. 8. Results from Experiment 2. Upper panel: d' values for discriminating correlations of 0.2 and 0.6. Lower panel: d' values for discriminating correlations of 0.3 and 0.7. Data to the left of the dashed vertical line are the experimental measurements for each listener (different symbols) in each of the three conditions. Data to the right of the dashed vertical line are derived form the raw data to the left by summing the corresponding d' values from the correlation and NoSo conditions.

produced an almost negligible increment when added to the d' values from the Corr condition. The monaural cues available are clearly very weak.

The fact that the conventional forms of interaction considered by signal detection theory are unable to account for the data does not necessarily rule out the idea that ρ is the critical decision variable in binaural detection. Nonetheless, having to invoke such a powerfully superadditive interaction in order to account for the data is relatively unparsimonious. Given that the equivalence of binaural cues for the Corr and the NoSπ condition was obtained by setting ρ to equal values, one can question whether ρ is in fact the decision variable at work. Again, EC theory may provide a more convenient account of these data, because it does not predict that the Corr and NoSπ conditions should provide equivalent binaural cues. As an alternative, Experiment 3 set out to test whether EC theory can deliver a more internally consistent account.

IV. EXPERIMENT 3

As discussed earlier, Durlach (1963, 1972) proposed a rather different mechanism for binaural detection that relies on canceling out the interfering noise. Here, the internal representation of the stimulus at one ear is subtracted from the internal representation of the stimulus at the other ear, following an equalization process that compensates for interaural differences in the interfering noise. The decision variable is the power, Y, of the resulting residue from cancellation. The final experiment set the predictions of the correlation and cancellation theories against each other. It sought to null out the effect of one or the other candidate binaural cue (interaural correlation or cancellation residue), by keeping it

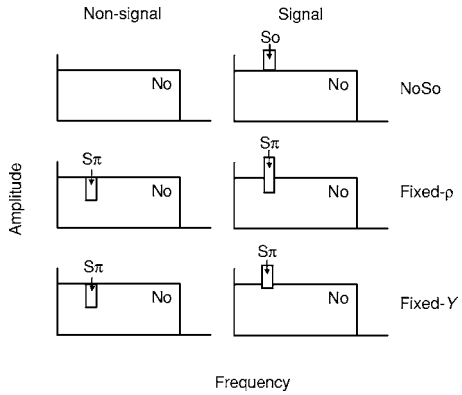


FIG. 9. Schematic illustration of the signal- and nonsignal-interval stimuli of Experiment 3. The spectral prominence in the signal interval is identical for each of the three stimulus types. In the Fixed-ρ condition, the proportion of anticorrelated noise in the target band is the same in the signal and nonsignal intervals; since the intensity of the target band increases, the amount anticorrelated noise increases in proportion. In the Fixed-Y condition, the amount of anticorrelated noise is the same in the signal and nonsignal intervals; since the intensity of the target band increases, the proportion of anticorrelated noise (and the resulting interaural correlation) is reduced in the signal interval. These features have been exaggerated for illustrative purposes.

constant across the two presentation intervals, while changing the spectrum level in the target band across the two intervals. If a candidate binaural cue is fixed in this way it should make no contribution to detection performance and discrimination will be equivalent to a diotic condition where only the spectrum-level change occurs. However, when the spectrum level of the target band changes, keeping one candidate cue constant will always produce changes in the other. For instance, if interaural correlation is kept constant, the cancellation residue will grow in power in proportion with the target band's spectrum level. Consequently, it should be possible for results from only one of the two binaural conditions to match the diotic condition. Using this "nulling" design, the stimuli of the "winning" condition are expected to be indistinguishable binaurally, so there is no need to make the assumptions (however reasonable) about conformance to Weber's law that were employed in the Discussion of Experiment 1.

In designing the stimuli, the residue from interaural cancellation, Y, is operationally defined as the power of the waveform produced by a simple subtraction of the stimulus as presented to the two ears. It is assumed that the actual output from an EC mechanism will be the same when Y is the same, because it seems unlikely that the operation of the EC mechanism will change across intervals. This assumption is considered reasonable in the context of limited variability among other stimulus properties. That is, the stimuli do not vary in any way that might be expected to make the operation of the EC model vary.[3]

A. Stimuli

The stimuli were similar in general to those of the previous experiments, except that the spectrum level of the target band was controlled independently of the two putative binaural cues. The construction of the three conditions is

illustrated in Fig. 9. In the nonsignal interval, the spectrum was always flat. In the signal interval the spectrum level of the target band was elevated above the flanking noise by a power factor, SP, between 0.5 and 2.5 dB. Three different conditions were constructed as follows.

All stimuli were prepared by creating complementary bandpass target and band-stop flanking noises. The target band was manipulated in correlation and power before adding it to the flankers. In the diotic condition, the entire stimulus was diotic in both intervals, with the spectral prominence of the target band controlled by scaling the target band power by SP. In the Fixed-ρ condition, ρ in the target band was manipulated to be 0.7 in both intervals using Eqs. (3) and (4), prior to any scaling. In the Fixed-Y condition, the target-band correlation in the nonsignal interval was also set to 0.7 in the same way, but Y was fixed to be the same in the signal as in the nonsignal interval. For this purpose, the scale factor for the anticorrelated noise in the target band is given by a in Eq. (5), while the scale factor of the correlated portion of the target-band noise, c, is given by Eq. (7), where SP is spectral prominence, applied as a factor to the power (expressed as linear units) of the target band. A value of one (equivalent to 0 dB) implies no spectral prominence,

$$c = \sqrt{\mathrm{SP} - \frac{1-\rho}{2}}. \tag{7}$$

The increased power in the target band for the Fixed-Y condition thus comes purely from additional correlated noise. Consequently, additional power results in a higher correlation in the Fixed-Y condition; in the nonsignal interval it is 0.7, and in the signal interval it varies from 0.73 to 0.83. Three different target bandwidths were employed, centered on 500 Hz. The bandwidths corresponded to 1 ERB (462–539 Hz), 2 ERBs (427–580 Hz), and 4 ERBs (360–670 Hz).

B. Procedure

The same three listeners attended six 1 h testing sessions, two for each of the three bandwidths. During each session, they completed 15 blocks of 50 trials. These 15 blocks covered five levels of spectral prominence (0.5, 1, 1.5, 2, and 2.5 dB) in each of the three conditions. Thus there were two blocks for a given target bandwidth, condition and spectral prominence. The d' values derived from each of these two blocks were averaged for each listener.

C. Results

As in the preceding experiments, the results were very similar for each of the three listeners. Consequently, Fig. 10 shows the mean discrimination rates across the three listeners for each level of spectral prominence. Separate panels show the results with different target bandwidths. Discrimination rates increased with the spectral prominence of the target band in the signal interval in all three conditions, but more rapidly in the Fixed-ρ condition than in the NoSo or Fixed-Y conditions. The results from NoSo and Fixed-Y conditions are essentially indistinguishable. Results for the three different target bandwidths are also very similar.

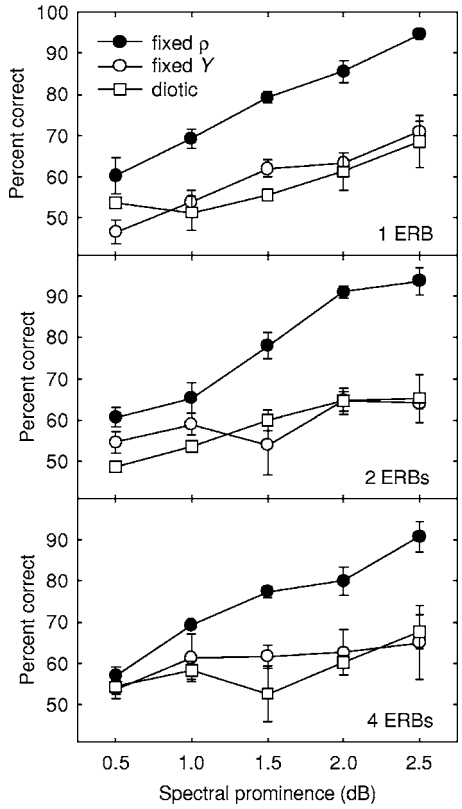


FIG. 10. Results from Experiment 3. Each panel shows data using a different target bandwidth (1, 2, or 4 ERBs). Symbols are mean percent correct discrimination (three listeners) as a function of the spectral prominence of the target band in the signal interval in each of the three conditions. Error bars are one standard error of the mean.

An analysis of variance covering the three bandwidths, three conditions, and five spectral prominences confirmed a significant main effect of spectral prominence [$F(4,8)=65, p<0.0001$], a significant main effect of condition [$F(2,4)=62, p<0.001$], and a significant interaction between the two [$F(8,16)=193, p<0.0001$]. There were no effects related to the bandwidth factor. The effect of spectral prominence reflects the usefulness of this cue in all conditions. The other two effects reflect the differences between the three conditions and their change with spectral prominence and these were analyzed further. Tukey HSD *post-hoc* analysis of the main effect of condition confirmed that the Fixed-ρ condition gave significantly higher discrimination rates than either the diotic or the Fixed-Y conditions ($q>13$, $p<0.01$, in each case), but that these conditions did not differ from each other. The interaction was first analyzed using simple main effects. These revealed that the effect of prominence was significant in each condition [$F(4,8)>10$, $p<0.002$, in each case], while the effect of condition was significant for spectral prominences of 1.5, 2, and 2.5 dB [$F(2,4)>17$, $p<0.02$, in each case]. Tukey HSD analysis of the interaction allowed comparison of the individual conditions at each level of spectral prominence; the Fixed-ρ condition gave significantly higher discrimination rates than the other two conditions at spectral prominences of 1.5, 2, and 2.5 dB ($q>6.5$, $p<0.05$, in each case), and no other differences were significant.

D. Discussion

If ρ were the most pertinent decision variable in binaural detection, one would expect that fixing ρ at 0.7 across the signal and nonsignal intervals would prevent listeners from using that cue. Consequently, listeners should only be able to use the change in spectral prominence of the target band in order to solve the task. In that case, one would expect performance to match the diotic condition where this spectral prominence was the only potential cue. The results instead showed that the Fixed-ρ condition gave consistently higher scores than the diotic condition, the difference increasing with the size of the spectral prominence in the signal interval. This result indicates that fixing ρ is an ineffective way to nullify binaural cues. It also reconfirms that monaural and binaural cues appear to interact strongly when one considers ρ as being the effective decision variable for the binaural system.

Similarly, if Y were the pertinent decision variable, one would expect performance with fixed Y to match the diotic condition. In this case, the results were indeed consistent with this expectation, so fixing Y appears to remove any effect of binaural processing. These results suggest that a far more parsimonious account of the data can be made using the idea of interaural cancellation than using interaural correlation, because the latter requires the intervention of a superadditive cue interaction in order to account for the data. The fact that Fixed-ρ stimuli gave higher scores than the other conditions, with the difference increasing with SP, is also predicted by EC theory; the amount of anticorrelated noise in Fixed-ρ stimuli increases with SP and this provides an increasing binaural cue on top of the increasing monaural cue.

One alternative interpretation of the observed pattern of results is that the salient cue is ρ, but that the differences between the Fixed-ρ and the Fixed-Y conditions are caused by effects occurring at the edges of the target band. An important consideration here is that, while the interaural statistics are under control within the target band, just outside the target band they are not. In particular, consider an auditory filter just above or below the target band. The interaural correlation within this filter is ρ_f. ρ_f will be dominated by a fixed level of correlated flanking noise, but will be influenced by the interaurally decorrelated noise within the target band. In the Fixed-ρ condition, the uncorrelated component of this noise will increase in power with SP, while the flanking noise will not change. Consequently, ρ_f will fall as SP increases. Since low values of ρ are more salient, this effect may facilitate the recruitment of additional frequency channels to the decision process in the Fixed-ρ condition and cause it to be easier for the listeners. Integration of cues across frequency thus offers an explanation of the observed increase in performance with SP in this condition. The experiment addressed this potential confound by exploring the effect of target bandwidth. If the advantage of the Fixed-ρ condition were produced by effects occurring at the edges of the target band, then one would expect the role of these effects to be reduced significantly compared to effects within the target band when the target band is broadened (i.e., the target band is twice or four times as wide, but there are still only two edges). The results, however, show no influence of the width of the target band on the pattern of results, suggesting that performance is determined by cues from the target band and not from the edges.

V. GENERAL DISCUSSION

For narrow-band signals in broadband noise, the effect of adding a signal in the NoSπ configuration was indistinguishable, at threshold, from the change in correlation to which it gave rise. However, at moderate signal-to-noise ratios (>-10 dB), systematic differences in sensitivity were observed using a loudness discrimination task. These differences could not be accounted for by the additional monaural cues that exist in the NoSπ condition, because these cues were found to be too ineffective. On the other hand, the differences were consistent with the operation of an EC mechanism. The operation of an EC mechanism was supported by a nulling experiment, which showed that binaural processing in a loudness discrimination can be foiled by fixing the power of the interaural difference signal, but not by fixing the interaural correlation. These results have clear implications for models of binaural unmasking, and are particularly relevant for discrimination tasks, such as speech perception, where the task is performed above detection threshold.

A number of studies have tested in detail the predictive power of correlation- and cancellation-based models of binaural unmasking. Typically, a broad range of existing data is successfully modeled, but a few problems tend to remain. Durlach (1972) made the first major effort in this direction. Durlach performed a rigorous examination of the analytically derived predictions of EC theory against a wide range of existing findings from the literature. His revised EC theory exhibited a small number of unresolved problems. For instance, EC theory predicted that the binaural unmasking effect should asymptote to zero with increasing signal frequency while a range of empirical studies show that it in fact asymptotes to about 3 dB. Similarly, Colburn (1973, 1977) examined the effectiveness of a correlational model, based on the statistics of auditory nerve responses and was able to predict the vast majority of the data, including additional binaural capacities such as interaural time and intensity discrimination. The most recent comprehensive study of this sort was performed by Breebaart *et al.* (2001a, b, c) using a computational EC model, which again captured a large quantity of empirical data. This analysis included a selection of the empirical studies examined by Durlach and by Colburn and went on to look at studies that used maskers with frequency- and time-dependent interaural parameters.

The general success of such modeling efforts led Colburn and Durlach (1978, p. 509) to remark that "Almost all models of binaural detection appear to be capable of describing the data on how the detection threshold varies with interaural parameters." It has since been widely presumed that different models of binaural unmasking cannot be differentiated psychoacoustically. However, Colburn and Durlach were careful to include the caveat that this observation held

"under the assumption of small signal-to-noise ratio." The above-reported series of experiments show some advantages for equalization-cancellation theory over correlation-based models. In particular, Experiments 2 and 3 show that EC theory offers a more parsimonious account of binaural signal discrimination at moderate SNRs (i.e., better than about −10 dB). The result supports the arguments of van de Par *et al.* (2001). They argued that employing correlation as a decision variable would involve an improbable precision in the stimulus level normalization needed for calculating the normalized correlation when the overall stimulus level was varied. They also suggested that an EC mechanism would provide a less problematic explanatory framework.

The experiment of van de Par *et al.* and Experiment 3 of the current study are representative of a different approach to binaural modeling; rather than attempting to model a large body of archival data, they set out to determine the appropriate parameters for a model through specifically designed empirical studies. Recent research on high-frequency unmasking has displayed a similar approach. This work suggests that the reduction in the binaural unmasking with increasing frequency is most likely related to a change in the neural encoding of the stimulus waveform with increasing frequency; at low frequencies, features of the carrier and envelope are both encoded, but with increasing frequency, the carrier information is lost, leaving only information about the envelope (e.g., Bernstein and Trahiotis, 1996; van de Par and Kohlraush, 1997). This account differs from that of classical EC theory (which relies on inaccurate selection of the ideal equalization delay to generate the effect of frequency) although it has tended to be included in models of binaural unmasking since Colburn's work.

The moderate SNRs examined here have rarely been the topic of psychoacoustic research, yet must be common in many practical situations where signals are detected or discriminated at levels above their detection threshold. In particular, speech perception in noise involves the simultaneous analysis of energy levels in multiple frequency channels at various sensation levels. The problem of binaural unmasking for speech led Culling *et al.* (2001) to make one of the few previous investigations of supra-threshold binaural processing. In pursuit of the idea that correlation was the effective decision variable, Culling *et al.* (2001) measured d' for discrimination of different correlation values. As in the present study, the correlation of a narrow subband of otherwise diotic noise was manipulated. On the assumption that decreases in ρ result in increases in loudness for such stimuli, the experiment was conducted as a loudness discrimination task and the resulting d' values were treated as unidimensional, such that they may be expected to sum (Macmillan and Creelman, 1991, p. 210). This assumption permitted the fitting of continuous functions to the data, which represent cumulative d' as a function of correlation (for instance, the ability to discriminate any positive correlation from unity was termed $d'_{(1,\rho)}$). One of the main findings of that study was that if these functions were rescaled using Eq. (2) in order to present them against the equivalent SNR (in decibels), they became much more linear, and not unlike the growth in loudness with stimulus intensity (e.g., Stevens, 1957).[4] The finding also echoed the work of Egan *et al.* (1969), who found that d' was often a more-or-less linear function of threshold SNR in decibels for both monaural and binaural detection. The present data would make the slightly modified prediction that for binaural detection d' should be a linear function of signal level in decibels, rather than signal-to-noise ratio in decibels.

Finally, the outcome of the present study appears to offer a ready explanation for the pattern of results observed by Culling *et al.* (2006) using speech stimuli. Like the results presented here with narrow-band-noise targets, Culling *et al.* found an apparent interaction between the roles of monaural patterns of power fluctuation and binaural patterns of correlation fluctuation in the intelligibility of speech in noise; neither cue alone supported measurable speech recognition, but when combined, substantial improvements occurred. It now seems likely that this "interaction" in fact reflects to some extent the operation of a cancellation mechanism, rather than a decorrelation detection mechanism. A cancellation mechanism is influenced by both the correlation and the power of the combined stimulus, and so would predict better performance in the combined case. It is not clear, however, whether this effect is the only one involved.

VI. CONCLUSIONS

When one considers correlation as the decision variable in binaural detection, there appears to be a superadditive combination of monaural and binaural information. This superadditive combination occurs more reliably when the stimulus is above threshold. The effect has now been observed for narrow-band target signals as well as speech. The results can be more parsimoniously understood using an EC mechanism as the binaural processor.

ACKNOWLEDGMENTS

I am grateful to Steven van de Par for helpful discussions, and Richard Freyman and the three reviewers for helping improve the paper so much. Work supported by UK EPSRC.

[1]One problem with the EC model is that in order to select the appropriate equalization it seems to require precognition of the interaural configuration of the noise. Two solutions to this problem have been offered. Culling and Summerfield (1995) and Breebaart *et al.* (2001a) each assumed that all equalization operations occur in parallel, an assumption that has some empirical support (Kohlrausch, 1990). Culling and Summerfield then used the rule that the equalization resulting in the lowest power output should be selected. This rule is equivalent to assuming that the noise is more intense than the signal and that by canceling as much of the stimulus as possible, the noise will be disproportionately removed. Breebaart *et al.*, on the other hand, suggest that the output of all transformations can be processed in parallel as well, thus maximizing the recovered data, although in their model verifications they also only used a single output.

[2]The signal-to-noise ratio was nominally −20 dB, although the use of Brown noise, which has a spectral roll off of −6 dB/oct. (i.e., with most of its energy at the lowest frequencies) means that the SNR was frequency dependent, varying between −10 and −20 dB for frequencies above 200 Hz.

[3]In assuming that the internal residue from cancellation is effectively the same as the difference between the left- and right-ear waveforms, the equalization and internal noise elements of the model are ignored. This omission was considered reasonable given that, (1) for these stimuli the

optimum equalization would be to leave the signals unaltered and (2) the internal noise should be constant. Equalization is not needed because, in the present stimuli, the maximum cross-correlation of the noise component of the stimulus is at zero delay (i.e., it is correlated noise), and there is no interaural level difference (note from Fig. 9 that the signal can be considered as purely anticorrelated and the noise as purely correlated). The internal noise should have no influence, because it should be almost identical in both intervals; the only consideration here is that the sound level in the target band may vary by up to 2.5 dB, a change that may have a modest influence on the relative level of the internal noise.

[4]The measurement of loudness as an accumulation of just noticeable differences in intensity is known to give results that deviate systematically from other monaural loudness measures when measured over a wide dynamic range (Hartmann, 1997, pp. 69–71). It is not known whether similar deviations occur in the binaural domain. The present analysis relies on the assumption that any such effects would be negligible across the small changes in loudness produced by interaural correlation.

Bernstein, L. R., and Trahiotis, C. (**1996**). "On the use of the normalized correlation as an index of interaural envelope correlation," J. Acoust. Soc. Am. **100**, 1754–1763.

Boehnke, S. E., Hall, S. E., and Marquardt, T. (**2002**). "Detection of static and dynamic changes in interaural correlation," J. Acoust. Soc. Am. **112**, 1617–1626.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1089.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1105.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition, III. Dependence on temporal parameters." J. Acoust. Soc. Am. **110**, 1105–1118.

Colburn, H. S. (**1973**). "Theory of binaural interaction based on auditory nerve data. I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. **54**, 1458–1470.

Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S., and Durlach, N. I. (**1978**). "Models of binaural interaction," in *Handbook of Perception*, edited by E. C. Carterette (Academic, New York), Vol. **IV**.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**). "Interaural correlation sensitivity," J. Acoust. Soc. Am. **110**, 1020–1029.

Culling, J. F., Edmonds, B. A., and Hodder, K. (**2006**). "Speech perception from monaural and binaural information," J. Acoust. Soc. Am. **119**, 559–565.

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.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **II**.

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.

Fletcher, H., and Galt, R. H. J. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

Hartmann, W. M. (**1997**). *Signals, Sound and Sensation* (Springer, New York).

Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Hirsh, I. J., and Burgeat, M. (**1958**). "Binaural effects in remote masking," J. Acoust. Soc. Am. **30**, 827–832.

Jain, M., Gallagher, D. T., Koehnke, J., and Colburn, H. S. (**1991**). "Fringed correlation discrimination and binaural detection," J. Acoust. Soc. Am. **90**, 1918–1926.

Kohlrausch, A. (**1990**). "Binaural masking experiments using noise maskers with frequency-dependent interaural phase differences. II. Influence of frequency and interaural-phase uncertainty," J. Acoust. Soc. Am. **88**, 1749–1756.

Licklider, J. C. R. (**1948**). "The influence of interaural phase relations on the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Signal Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **50**, 1494–1495.

Stevens, S. S. (**1957**). "Concerning the form of the loudness function," J. Acoust. Soc. Am. **29**, 603–606.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

van de Par, S., Trahiotis, C., and Bernstein, L. R. (**2001**). "A consideration of the normalization that is typically included in correlation-based models of binaural detection," J. Acoust. Soc. Am. **109**, 830–833.

The ability to listen with independent ears[a)]

Frederick J. Gallun,[b)] Christine R. Mason, and Gerald Kidd, Jr.
Department of Speech, Language and Hearing Sciences, and Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 9 February 2007; revised 30 July 2007; accepted 6 August 2007)

In three experiments, listeners identified speech processed into narrow bands and presented to the right ("target") ear. The ability of listeners to ignore (or even use) conflicting contralateral stimulation was examined by presenting various maskers to the target ear ("ipsilateral") and nontarget ear ("contralateral"). Theoretically, an absence of contralateral interference would imply selectively attending to only the target ear; the presence of interference from the contralateral stimulus would imply that listeners were unable to treat the stimuli at the two ears independently; and improved performance in the presence of informative contralateral stimulation would imply that listeners can process the signals at both ears and keep them separate rather than combining them. Experiments showed evidence of the ability to selectively process (or respond to) only the target ear in some, but not all, conditions. No evidence was found for improved performance due to contralateral stimulation. The pattern of interference found across experiments supports an interaction of stimulus-based factors (auditory grouping) and task-based factors (demand for processing resources) and suggests that listeners may not always be able to listen to the "better" ear even when it would be beneficial to do so. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2780143]

PACS number(s): 43.66.Pn, 43.66.Ba, 43.66.Dc [RLF] Pages: 2814–2825

I. INTRODUCTION

A number of influential models of binaural processing contain the assumption that listeners have access to both a monaural and a binaural mode of processing incoming auditory stimulation (e.g., Colburn and Durlach, 1978). In a recent example, Breebaart *et al.* (2001) modeled a wide variety of signal detection experiments by explicitly including both a binaural and a monaural processing path and allowing the model to choose the pathway that leads to the highest signal-to-noise ratio (SNR). While this makes conceptual sense from a bioengineering perspective and can explain a substantial amount of the data on binaural release from masking (see also Zurek, 1993), the evidence does not entirely support the assumption that listeners voluntarily switch between these two fundamentally different modes of processing. Strong evidence in favor of the availability of a monaural listening strategy comes from work on selective listening (e.g., Broadbent, 1958; Cherry, 1953; Moray, 1970; Treisman, 1964, 1969; Wood and Cowan, 1995) in which listeners often show substantial success in reporting the stimuli presented to one ear and ignoring competing stimuli presented to the opposite ear. On the other hand, several more recent studies (e.g., Heller and Trahiotis, 1995; Brungart and Simpson, 2002; 2004; Brungart *et al.*, 2005; Kidd *et al.*, 2003; Gallun *et al.*, 2005; Shub, 2006) have presented evidence suggesting that listeners cannot simply choose the preferred ear and ignore the other.

These recent examples of "contralateral interference" are also problematic because most models of binaural release from masking (e.g., Zurek, 1993) suggest that interaural differences in level (ILD) result in binaural release through the action of the "better-ear" effect, by which it is meant that listeners selectively attend to the ear with the better SNR. In realistic listening situations, this ability to choose a monaural better-ear listening strategy is fundamentally confounded with the use of perceived differences in location generated by ILDs. Whenever the SNR differs at the two ears there is by definition an ILD in the target, the masker, or both. Most models assume that it is the better SNR at one of the ears that drives performance. However, Gallun *et al.* (2005) showed that listeners can exploit target or masker ILDs in order to improve performance even in the absence of a better-ear effect. In that study, the better-ear hypothesis was excluded by manipulations that involved a monaural target and a fixed-level masker at the target ear. Results showed that increasing the masker level at the nontarget ear actually led to improved performance. These findings suggest that at least some of the results that have previously been attributed to better-ear listening may be the result of either (1) perceived differences in location (or some other perceptual cue) generated by the presence of ILDs or (2) a cancellation process similar to that proposed by Durlach (1960, 1963) in his equalization and cancellation (EC) model of binaural unmasking.

Another example of a situation where a change in performance occurred with no change in the SNR at the target ear is found in the results of Kidd *et al.* (2005) in which

[a)]Portions of this research were presented at the 2006 Midwinter Meeting of the Association for Research in Otolaryngology.

[b)]Current address: National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road (NCRAR), Portland, OR 97239. Electronic mail: Frederick.Gallun@va.gov

listeners benefited from the addition of a second masker to the nontarget ear. Targets were narrow-band speech stimuli (described in the following) presented monaurally and the first masker was narrow-band speech in nonoverlapping frequency ranges ("different-band speech" or DBS; also described in the following) presented to the target ("ipsilateral") ear. When a second masker consisting of noise matched in frequency to the speech masker ("different-band noise" or DBN; described in the following) was presented to the contralateral ear, performance was usually better than when no contralateral masker was presented. This result presents problems for a better-ear explanation because there was a change in performance even though the stimulus at the "better" ear (the one that contained the target) had not changed. Three distinct hypotheses were initially proposed, all of which focused on how information in nontarget frequency regions at the nontarget ear could have been used to improve performance.

The first potential explanation for the Kidd *et al.* (2005) result is based on "cueing." If listeners are able to process the two ears independently, they might be able to use the information at the nontarget ear to tell them which frequencies to ignore at the target ear. This strategy, which is similar to what Wang and Brown (2006) and Brungart *et al.* (2006) would call an "ideal time-frequency mask," does not imply a binaural combination of information in the usual sense of the term. This type of independent processing has also been referred to in the past as a "contralateral cue" (e.g., Sorkin, 1965; Taylor and Forbes, 1969; Koehnke and Besing, 1992), although in that literature the task was always detection of an exact copy of the signal presented to the contralateral ear. As the data presented in the following show no evidence of this strategy, it will not be discussed further.

The next two "binaural combination" hypotheses are based on the assumption that listeners must choose when separating signals to use the frequency dimension or the dimension of spatial position (in this case, ear of presentation). This need to choose between strategies is similar to an explanation given by Brungart and Simpson (2007) for why speech from a different-sex talker (which resulted in less interference than a same-sex talker when presented alone) caused more interference than a same-sex talker when presented with a contralateral masker. For the stimuli presented in Kidd *et al.* (2005), the obvious strategy is to separate by frequency, since this dimension distinguishes the target from both maskers, which should then result in a binaural combination of the two maskers. The question to be answered is why such a strategy would lead to improved performance over the case where there was only a single speech masker. The first hypothesis is that if the listener separates target from masker on the basis of frequency, then the result is a monaural target composed of one collection of frequencies and a binaural masker composed of another set of nonoverlapping frequencies. If so, then this binaural difference would have acted as did the binaural differences in Gallun *et al.* (2005), allowing listeners to perceptually segregate the target from the masker and/or to use an EC-type operation to improve intelligibility of the target.

The second binaural combination hypothesis is that when the frequency-aligned speech and noise maskers were combined, it was as if they had been presented monaurally and that the resulting combined masker was simply less intelligible due to energetic rather than informational masking. Evidence against this second binaural combination hypothesis comes from the finding that performance was better in the condition where the speech and noise maskers were in opposite ears than in a control condition in which both maskers were presented monaurally to the target ear. While reduced intelligibility may certainly play a role, the difference between the masking exerted by the contralateral and monaural combinations suggests that intelligibility cannot explain the entire effect.

Because it is not possible to distinguish among these various potential mechanisms on the basis of the data of Kidd *et al.* (2005) alone, the initial goal of this study was to consider some conditions in which identifying the optimal strategy becomes more difficult. The first experiment was carried out as an extension of the work presented in Kidd *et al.* (2005) and Gallun *et al.* (2005), involving the same listeners, the same equipment and response style, and the same types of stimuli. The new conditions tested contralateral presentation of noise matched in frequency to the target rather than to the speech masker. As the first experiment primarily provided evidence of contralateral interference, the second and third experiments were designed to further examine which aspects of the first experiment were responsible for this interference. The overall pattern of results suggests that the difficulties listeners experience can be attributed to (1) task-based factors involving a general limit on the number of source segregation strategies that can be employed in any given listening situation (demand for "processing resources"), and (2) stimulus-based factors involving a tendency for the auditory system to treat similar stimuli as if they were generated by a single source ("auditory grouping").

II. METHODS

A. Listeners

Seven listeners participated, six females and one male, all between the ages of 21 and 40 years and all with pure-tone thresholds within 15 dB of audiometric norms (ANSI 2004) at octave frequencies from 250 to 8000 Hz. All were paid listeners who were familiar with psychometric testing and with the type of processed sentences used. All had a minimum of three weeks experience participating 4–6 h per week in similar tasks. Each experiment involved a subset of three to four of these listeners and each took approximately four weeks of listening to complete.

B. Stimuli

Stimuli were sentences processed into narrow frequency bands as described in Arbogast *et al.* (2002). The original speech was taken from the coordinate response measure corpus (Bolia *et al.*, 2000) with the structure: "Ready [callsign] go to [color] [number] now," with eight callsigns, four colors (white, red, green, and blue) and eight numbers (1–8). To

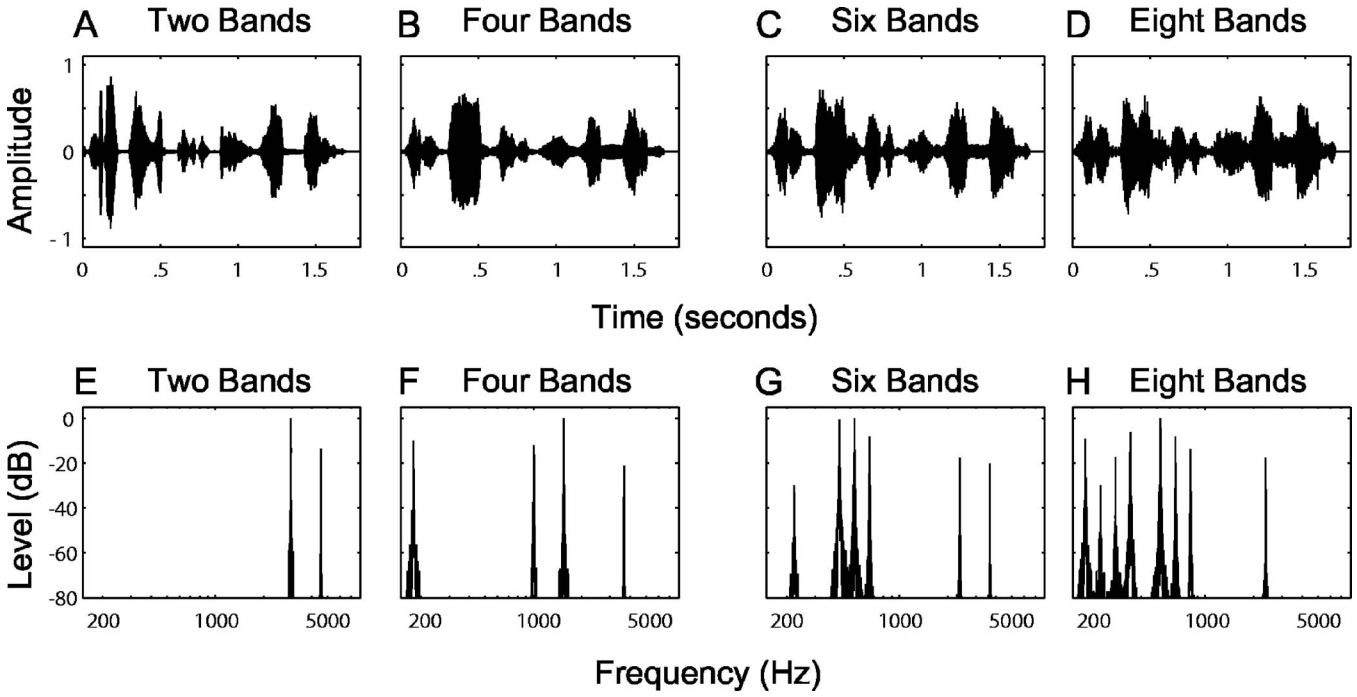


FIG. 1. Examples of wave forms and magnitude spectra for two-band, four-band, six-band, and eight-band target stimuli. Wave form amplitude is shown in arbitrary linear units (e.g., volts). Sound pressure level is displayed in decibels relative to the maximum level for any one band for each target. On each trial the particular bands to be used were chosen randomly from a possible 15. See the text for complete details of the stimulus generation.

restrict the frequency content, sentences were passed through a first-order high-pass Butterworth filter with a cut-off of 1200 Hz to roughly equate energy across the spectrum, after which 15, approximately one-third octave, fourth-order Butterworth filters evenly spaced on a logarithmic scale from 215 to 4891 Hz were used to divide the sentence into 15 narrow bands. Half-wave rectification and low-pass filtering at 50 Hz by a fourth-order Butterworth filter extracted the amplitude envelope within each band, reducing the speech wave forms to a set of 15 amplitude envelopes, each associated with one of the bands.

Processed speech targets and maskers were generated by randomly choosing two, four, six, or eight of the envelopes (depending on the experiment) and using them to modulate a set of pure tones with frequencies equal to the center frequencies of the chosen bands. Presenting the envelope-modulated tones together resulted in a sentence with restricted frequency content and reduced harmonic structure but with the amplitude variations over time that had occurred in those bands in the original sentence. Examples of the four types of processed speech are shown in Fig. 1. Note that on every trial new bands were randomly chosen to compose the target.

The types of processed speech and noise that were used as maskers appear in Fig. 2. In order to distinguish maskers that shared frequency content with the target ("same band") from those that did not ("different band"), the letters SB and DB will be used. For example, "different-band speech" (DBS) was composed only of bands not contained in the target speech and "different-band speech, reversed" (DBSr) was constructed by simply playing DBS wave forms backwards. In all three experiments, the talker, callsign, color, and number used for the masker sentence were all different from those used for the target sentence, but all of the talkers were male. In the first experiment, the target callsign was always "Baron," but in the second and third experiments the target callsign was randomized. To facilitate target sentence identification when the callsign was randomized, the callsign associated with the target sentence was indicated before each trial on the display in front of the listener.

Two types of noise maskers were used. Broadband noise (BBN) maskers were generated in the frequency domain and extended from 20 to 8000 Hz. Narrow-band noise maskers were generated by multiplying processed speech in the frequency domain by BBN. This resulted in noise maskers with the same number of bands as the processed sentences but none of the amplitude modulations necessary for interpreting them as speech. "Same-band noise" (SBN) had the same bands as the target speech, while "different-band noise" (DBN) did not. Four-band example wave forms and magnitude-frequency spectra for these five masker types appear in Fig. 2. Note that in order to ensure that the masker bands were the same or different from the target required a new draw of masker bands on each trial as well. A new draw of BBN was also generated on each trial. For all signals, level in dB SPL was calculated based on the duration of the entire signal.

C. Procedure

All of the stimuli were stored on a computer and played through Tucker-Davis Technology (TDT System II) 16 bit digital-to-analog converters at a rate of 50 kHz, then low-pass filtered at 7.5 kHz. For experiment one, target and masker levels were controlled by independent programmable attenuators (TDT PA4), whereas for experiments two and

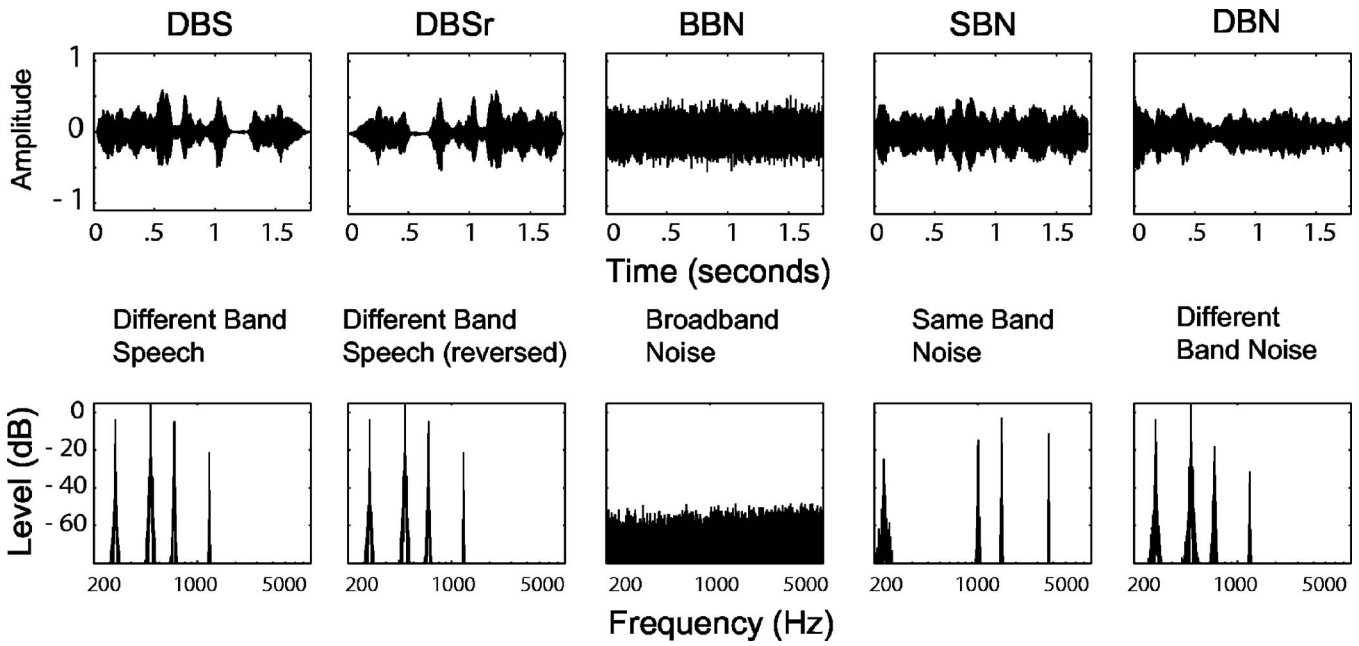


FIG. 2. Examples of wave forms and magnitude spectra of the various maskers used in the three experiments. Each is matched to the four-band target shown in Fig. 1. Wave form amplitude is shown in arbitrary linear units. Sound pressure level is displayed in decibels relative to the maximum level for the four-band target shown in Fig. 1. See the text for complete details of the stimulus generation.

three, relative target and masker levels were set digitally and overall level was controlled by two independent programmable attenuators. Listeners were seated in individual double-walled IAC booths. The stimuli were presented through matched and calibrated TDH-50 earphones. The task of the listener was to identify the color and number from the sentence having the callsign "Baron." Both the color and the number had to be reported accurately for a listener to be correct on each trial. Chance performance was thus about 3% (four colors by eight numbers). Responses and after-trial feedback were given via a handheld device with an LCD screen and buttons corresponding to the possible colors and numbers.

III. EXPERIMENT ONE: CONTRALATERAL SAME-BAND NOISE

The purpose of the first experiment was to examine the degree to which listener performance in a speech-on-speech masking task was affected by the presence of contralaterally presented SBN. Evidence for the influence of an independent masker presented contralaterally came from the improvements found by Kidd *et al.* (2005) when different-band speech (DBS) was presented ipsilaterally and the contralateral masker was DBN. One hypothesis proposed for those improvements was that listeners combined the DBN and DBS into a single binaural signal and made use of the binaural differences generated by a monaural target and a binaural masker. Such a hypothesis would predict that similar improvements should occur for a contralateral SBN noise masker because now the target would be binaural and the DBN masker would be monaural. An additional, but not mutually independent, hypothesis was that the DBN also reduced the intelligibility of the DBS speech masker in the results of Kidd *et al.* (2005). Were this additional hypothesis correct, it would imply that performance should be *reduced* by adding SBN (rather than improved) since now the *target* intelligibility would decrease. The first experiment was primarily designed to give an initial answer as to whether performance improved or declined.

A. Design

The narrow-band processing described in Sec. II was used to generate eight-band targets and spectrally matched eight-band SBN maskers as well as six-band DBS maskers. On each trial, the target and DBS masker were presented to the right ear while the SBN masker was presented to the left ear. The target was presented at 60 dB SPL on all trials. Three levels of DBS were used: 50, 60, and 70 dB SPL. Six levels of SBN were used: 30, 40, 50, 60, and 70 dB SPL and a no-noise control. Within each block of 50 trials, all masker levels were varied randomly from trial to trial and sorted for analysis afterwards. Each listener participated in a minimum of 150 trials for each combination of levels in each condition. Three listeners (L1, L2, and L3) participated immediately after participating in the conditions described in Kidd *et al.* (2005) and Gallun *et al.* (2005).

B. Results

Average performance for the three listeners appears in Fig. 3, with error bars indicating ±1 s.d. across listeners. The large error bars indicate the large differences in the performance of individual subjects that are typical in tasks involving maskers that are separated from the targets in frequency (e.g., Durlach *et al.*, 2005). These same subjects produced similarly divergent levels of performance in the conditions reported in Kidd *et al.* (2005) and Gallun *et al.* (2005), but as with those data sets the patterns of performance were consis-

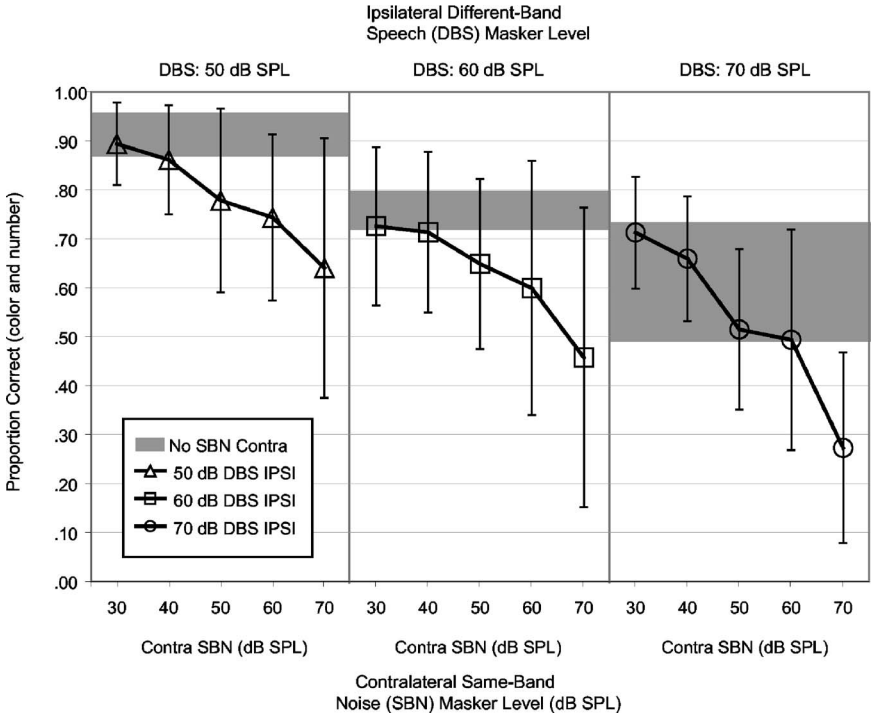


FIG. 3. Average performance for three listeners in experiment one. The listener's task was to identify the color and number keywords contained in an eight-band processed sentence presented to the right ear at a level of 60 dB SPL. Two maskers were always present: an *ipsilateral* (right-ear) different-band sentence (DBS) masker and a *contralateral* (left-ear) same-band noise (SBN) masker. The three panels show performance as a function of contralateral noise level for an ipsilateral sentence masker presented at a level of 50, 60, or 70 dB SPL. The shaded area in each panel indicates ±1 s.d. deviation around mean performance across listeners with no contralateral masker present. Error bars on the symbols show ±1 s.d. across listeners.

tent across listeners. The shaded areas represent the region of performance obtained when no contralateral masker was presented (mean ±1 s.d.).

As can be seen in Fig. 3, the overall trend is that raising either DBS level (successive panels) or SBN level (successive points within a panel) resulted in decreased performance. A repeated-measures analysis of variance (ANOVA) was performed on the percent correct data and DBS level (three values) and SBN level (six values, which includes no SBN masker) were entered as independent factors. The effect of DBS level was statistically significant [$F(2,4)=65.85$, $p<0.001$] as was the effect of SBN level [$F(5,10)=8.95$, $p<0.002$]. The interaction did not reach statistical significance ($p=0.065$).

C. Discussion

The observed decline in performance with increasing contralateral stimulation provides strong evidence that listeners were unable to simply listen to the ear with the higher target-to-masker ratio, as essentially all binaural models would predict. If the matched frequency content of the target and contralateral masker did cause an obligatory binaural combination ("auditory grouping"), then the resulting signal was so degraded by the combination of speech and noise that the improvement due to the binaural differences between target and DBS masker was offset by the reduced intelligibility of the target.

An alternative interpretation, similar to that suggested by Brungart and Simpson (2007) for a finding with unprocessed speech, is that listeners were unable to allocate sufficient processing resources to apply both an ear-based listening strategy and a frequency-based strategy simultaneously and were forced to choose one or the other. This implies that if the ipsilateral different-band speech masker had not been present, then listeners would have been able to segregate the signals at the two ears and performance would not have suffered. It is this hypothesis that is tested in experiment two.

One potential contaminating factor is that of sound conduction through and around the head. At the higher monaural stimulation levels for the contralateral masker there is the chance that the contralateral masker was in fact acting as an ipsilateral masker due to "cross-hearing." If interaural attenuation were only 40 dB, for example, the 50, 60, and 70 dB contralateral same-band noise maskers would have been equivalent to 10, 20, and 30 dB ipsilateral same-band noise maskers. This possibility will also be investigated further in experiment two.

IV. EXPERIMENT TWO: VARIATIONS IN NOISE BAND LOCATION, NUMBER, AND SPECTRUM

Based on the results of the first experiment, it was hypothesized that listeners were unable to process the information at the two ears independently due to the processing demand of also segregating the ipsilateral target and masker by frequency. The implication of this hypothesis is that since the listeners chose to perform the frequency segregation, the frequencies that were similar at the two ears were combined binaurally (despite their dissimilarity in the temporal domain) resulting in a "target," which was actually a mixture of target and SBN masker. It is not difficult to see why this would have led to reduced performance. One possible interpretation of the mixed processing is that a single binaural image was created in which the speech and the noise envelopes were combined, resulting in a less intelligible signal. Another interpretation is that listeners were "distracted" by the noise, although that interpretation seems unlikely given the improvement in performance found by Kidd *et al.* (2005) when the contralateral noise bands corresponded to the ipsilateral speech masker bands. The present experiment is intended to determine whether the apparently obligatory com-

bination of corresponding bands across ears depends on the presence of an ipsilateral segregation task. While the alternative explanation, that the contralateral interference effect only depends on a high degree of similarity between the stimuli in the two ears, is possible, it seems unlikely given the temporal dissimilarity between the processed speech and the same-band noise.

In order to examine this issue more carefully, the ipsilateral DBS masker was removed, and two manipulations of the target and remaining noise masker were introduced. First, the number of bands in the target was varied by presenting two, four, six, or eight bands of speech and, second, either a long-term spectrally-matched noise (SBN) or a BBN was used as a contralateral masker. These two manipulations examined the effects of increasing both the temporal and the spectral differences between the signals at the two ears. Conceptually, these additional differences were regarded as information suggesting that the target and contralateral masker should be treated as independent sources. Under this framework, for every additional frequency band added to the target and to the masking noise, the listener is given additional evidence supporting the conclusion that the amplitude fluctuations in the target and the noise are independent. Thus, the greatest effects of binaural combination should be seen with the fewest numbers of bands. Similarly, since increasing the bandwidth of the noise increases the range of modulation frequencies present in the noise [the high-frequency cut-off is equal to the bandwidth, see Ewert and Dau 2000], the dissimilarity of the modulation spectra at the two ears should be much greater for BBN than for narrow-band noise (SBN).

In order to compare the energetic masking effects of SBN and BBN directly, control conditions were included in which both maskers were presented ipsilaterally instead of contralaterally. These control conditions were also used to provide data relevant to the issue (mentioned in the discussion of experiment one) of possible ipsilateral masking caused by sound conduction through and around the head from the contralateral masker. For this reason, lower noise levels were used ipsilaterally than contralaterally, although the average levels of the broadband and narrow-band noise were equated as described in the following.

A. Design

The narrow-band processing described in Sec. II was used to generate two, four, six, and eight band targets (shown in Fig. 2) and spectrally matched noise (SBN) as well as independent BBN. On each trial, the target was presented to the right ear at an overall level of 50 dB SPL. On each block of 50 trials, the location (ear of presentation), level, and frequency content (SBN or BBN) of the masking noise was kept constant, but the order in which the blocks was run was mixed and the listener was not told in advance precisely which combination to expect. Four new listeners (L4, L5, L6, and L7) participated. None had the experience of the listeners in experiment one, but all had experience with psychophysical testing and had participated in at least one other experiment. In the conditions in which the noise was presented to the nontarget ear (contralateral noise presentation), the noise could be either BBN or SBN (with the number of bands matched to that of the target) and the possible noise levels were 30, 40, 50, 60, or 70 dB SPL. In the conditions in which the noise was presented to the target ear ("ipsilateral" noise presentation), the noise could have the same frequency configurations, but the possible noise levels were 10, 20, 30, 40, and 50 dB SPL. These levels represent 20 dB of simulated interaural attenuation, which is an overly conservative estimate (for these headphones, which are used in clinical audiometry, the conservative estimate of interaural attenuation commonly used is 40 dB). Each listener completed a minimum of 150 trials for each noise level in each condition. The contralateral noise conditions were completed before the conditions in which the noise was presented to the target ear. For comparison, each listener also completed a set of trials in which no noise was presented ("target-alone" condition).

B. Results

The average percent correct data are shown in Fig. 4. The shaded areas represent the region of performance obtained when no masker was present (mean ±1 s.d.). The main effect of number of bands was analyzed through a repeated-measures ANOVA performed on the proportion correct data obtained in the target-alone condition. The effect of number of bands was significant [$F(3,9)=114.327$, $p<0.001$] and paired t-tests showed that performance was the same for the six-band and eight-band targets ($p=0.58$) but that the two-band and four-band targets differed from each other ($p<0.001$) and from the six-band and eight-band targets ($p<0.02$ or lower in all cases).

Because levels were different for the contralateral and ipsilateral maskers, two sets of repeated-measures ANOVAs were performed. Analyses of simple main effects were performed on the individual condition of interest as a general ANOVA showed that there were significant interactions between all the factors. The most important result can be seen by comparing performance as a function of contralateral SBN level for the two-band and eight-band targets. While there is a significant reduction in performance for the two-band target [$F(4,12)=32.93$, $p<0.001$], there is no significant reduction for the eight-band target ($p=0.207$). The four- and six-band targets fall intermediate between these two extremes [$F(4,12)=7.94$, $p<0.01$ and $F(4,12)=3.57$, $p<0.04$]. On the other hand, the effect of level was not significant for the contralateral BBN maskers, regardless of the number of bands in the targets with which they were paired. For both the BBN and SBN presented to the target ear (ipsilaterally), the effect of noise level was significant for all of the numbers of bands ($p<0.01$).

Because so many of the performance values were near ceiling, a four-way repeated-measures ANOVA was performed on the original scores and on a transformed version of the scores, using the Rationalized Arcsine transform (Studebaker, 1985). In order to equate for the different noise levels used, the noise levels were rank-ordered before being entered into the analysis. The results mirrored those for the original scores. Nonetheless, while those conditions that generated scores that were immeasurably high can certainly be

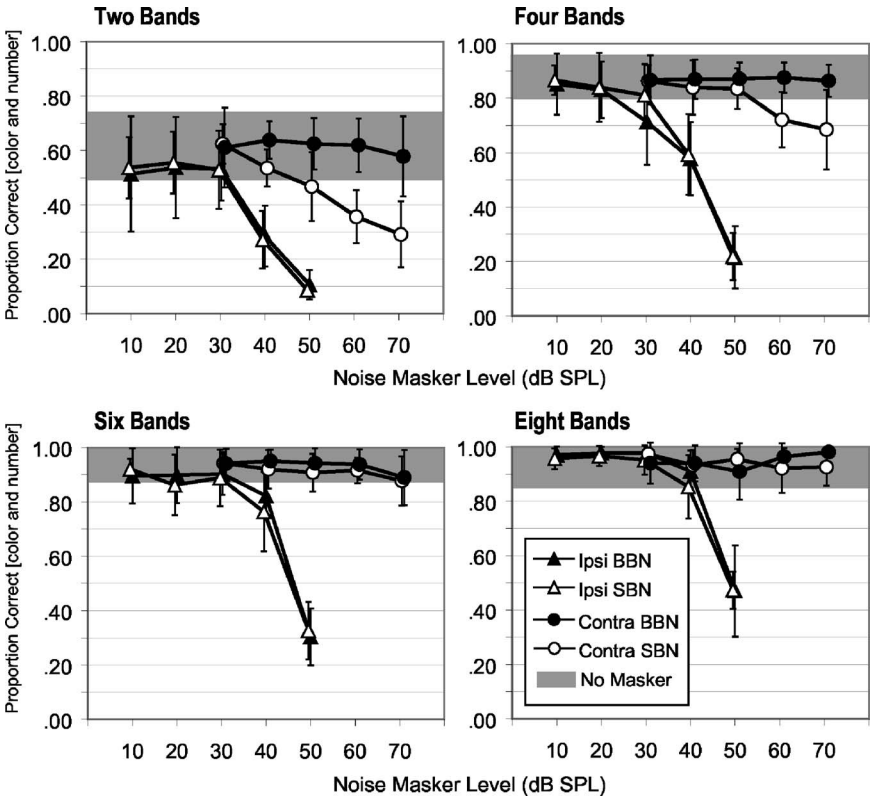


FIG. 4. Average keyword identification performance for four listeners in experiment two as a function of masking noise level, where maskers could be broadband noise (BBN; closed symbols) or same-band noise (SBN; open symbols) and presented to the same ear as the target (ipsilaterally; triangles) or the opposite ear (contralaterally; circles). Targets were presented to the right ear at a level of 50 dB SPL and the four panels show average subject performance for two-, four-, six-, and eight-band targets. The shaded area in each panel indicates ±1 s. d. around mean performance across listeners for that number of target bands with no masker present. Error bars on the symbols show ±1 s. d. across listeners.

regarded as immune to contralateral interference with these stimuli and levels, any comparisons across such conditions should be regarded as specific to these parameters. For example, the fact that no BBN maskers resulted in contralateral interference in this experiment should not be taken to mean that BBN could not potentially exert a contralateral masking effect under different circumstances. Similar care should be taken in interpreting the nonsignificant differences between six- and eight-band stimuli.

C. Discussion

The most important finding obtained from experiment two is that there was an interaction between the number of bands in the target and the amount of contralateral masking that is observed, but only when the long-term spectrum of the masker was matched to that of the target. This has implications for both of the questions that motivated the second experiment. Foremost, this result supports the interpretation that as the temporal similarity between target and masker decreases with increasing numbers of bands, the ability to segregate target from masker increases. This can be considered as support for an auditory grouping explanation of contralateral interference, which refers to the idea that the auditory system is sensitive to the amount of information available indicating that the target and masker are from independent sources. This interpretation is also consistent with the absence of any effect of presenting the broadband noise to the contralateral ear. One possible difficulty with this conclusion is that as the number of bands increases, there is an increase in the number of distinct spectral regions in which both ears are being stimulated. This similarity in spectral information at the two ears could be an indication that the two ears are being stimulated by the same sound source, in which case more bands should lead to increased grouping across the ears. Future work should examine the relative contributions of temporal and spectral similarity in conditions where there is ambiguity as to whether or not various signals should be combined or segregated. It is possible that temporal similarity (or dissimilarity) is simply a more powerful grouping cue than spectral similarity.

In terms of the influence of the ipsilateral speech masker in experiment one, the fact that so little contralateral masking was obtained with the six- and eight-band targets in the second experiment supports the interpretation that listeners are more capable of attending to the ear with the best signal-to-noise ratio if there is only one masker present. Such an interaction of task demands and stimulus configurations suggests that in order to understand the ability of listeners to process sounds independently at the two ears, it is necessary to consider carefully both the task demands and the nature of the stimuli being presented. For example, while BBN clearly can be an effective masker when presented ipsilaterally (it was just as effective as the SBN when presented to the target ear), when presented to the nontarget ear, there was no evidence of contralateral masking for even the two-band target. It would be interesting to determine whether or not this result holds in the presence of an ipsilateral masker that was spectrotemporally dissimilar to the noise as well as dissimilar to the target.

The data from experiment two also show that the results of the first experiment cannot be dismissed as an example of cross-ear masking. There are three reasons for this. The first is that the BBN and SBN stimuli produced indistinguishable masking patterns when presented to the target ear but signifi-

cantly different patterns of masking when presented contralaterally. In particular, notice the difference in the slopes of the masking functions for ipsilateral and contralateral presentation. This suggests that fundamentally different mechanisms of masking are active. The second reason is that when presented ipsilaterally, there is no evidence of masking at levels below 30 dB SPL, which is the greatest conceivable level that even a very conservative estimate of interaural attenuation could provide. The third reason that acoustic crossover is ruled out by these results is that while there is clear evidence of ipsilateral masking for the six-band and eight-band targets, there is no evidence of contralateral masking at levels that produce contralateral masking in the two-band and four-band stimuli. These three findings provide clear evidence that the contralateral masking observed in experiments one and two was a result of specific interactions between the SBN masker and the target that were not facilitated by acoustic crossover but by binaural processing at higher auditory centers.

A final point regarding the results of experiment two concerns the difference between the masking exerted by the contralaterally presented BBN and SBN. Since the two maskers were equated for overall level, rather than level within the target band (as is shown clearly in Fig. 2), it is possible to argue that the different results for the contralateral maskers resulted from different amounts of energy within the target bands. This suggestion is incompatible with the fact that (as can be seen in all panels of Fig. 4) when the BBN and SBN were presented ipsilaterally they resulted in nearly identical performance (and there was substantial masking, so ceiling effects cannot be thought to have influenced the data). This suggests that the difference in effectiveness in the contralateral presentation was due not to within-channel energy but rather to spectro-temporal similarity to the target.

V. EXPERIMENT THREE: CONTRALATERAL MASKING IN THE PRESENCE OF REVERSED SPEECH

In experiment two, the six- and eight-band targets suffered no measurable interference even when the contralateral noise was at the same levels as for the conditions in experiment one where substantial interference was observed. If the ability to hold the two ears separate depends on task demands, and specifically on the availability of cognitive processing resources, then it is possible the decreased interference resulted from the fact that there were only two tasks to be performed concurrently (interpret the speech target and keep the stimuli at the two ears separate) as opposed to the three tasks that had to be performed in experiment one (interpret the speech, separate the ears, and keep ipsilateral target and masker separate). Note that this argument could also be framed in terms of allocating attention, where attention (or processing resources) must be divided among three tasks in experiment one and between two tasks in experiment two. If a single limited resource is required for both the speech recognition and the source segregation tasks (as described in the model of Navon and Gopher, 1979) then there should be a distinct improvement when listeners go from three tasks to two.

This is similar to saying, as suggested by Brungart and Simpson (2007), that if listeners cannot apply multiple source-segregation strategies simultaneously, then introducing a masker will reduce performance proportional to the degree to which the second masker requires an orthogonal method of source segregation. By this argument, the contralateral interference conditions in experiment two allowed much improved performance over similar conditions in experiment one because now listeners were able to devote a greater proportion of their processing resources to segregating target from masker on the basis of ear of presentation.

In order to examine why there was more contralateral interference for the eight-band target in the first experiment than for the six- or eight-band targets in the second experiment, a subset of the stimuli from experiment two were combined with an additional ipsilateral masker. This manipulation was expected to increase task complexity by requiring listeners to perform both an ear-based segregation and a frequency-based segregation. In addition, the stimuli were chosen in a manner that would allow examination of the interactions between the number of tasks to be performed and the spectrotemporal similarity of the target and the masker without the confounding issue of confusions between two intelligible speech stimuli. Different-band speech that had been reversed in time (DBSr, see Fig. 2) was chosen as the ipsilateral masker and six-band speech was chosen as the target. DBSr resembles the DBS masker of experiment one (and thus the target) in terms of its distribution of long-term temporal fluctuations and it matches the DBS masker exactly in terms of long-term average spectral composition. It differs from DBS, however, in that it is unintelligible and as such is less likely to be confused with the target. Thus, the increase in the number of tasks is more controlled because while the listener still must distinguish between temporal variations that form target words and those that do not, the additional task of discarding intelligible words that are not the target is removed. This is worth considering because, indeed, over 90% of all responses in every condition in experiment one included at least one word that had been presented either as a target or a masker. If the entire cause of contralateral masking in experiment one was confusion between the target words and intelligible words from the DBS masker, then the results should resemble those of experiment two instead, and there should be no masking for a six-band target.

A. Design

The processed six-band speech targets from experiment two were presented to the right ear, always at a level of 50 dB SPL. Six-band DBSr maskers were also presented to the right ear at a level set individually for each listener such that performance with no contralateral stimuli was roughly 85% correct identifications of color and number. The same four listeners from experiment two participated, and the DBSr levels were fixed for the remainder of the experiment as follows: L4:35 dB; L5:50 dB; L6:30 dB; L7:50 dB SPL. Once these levels had been established from psychometric functions obtained in a pilot test, data collection was started in the contralateral masking conditions. Two con-

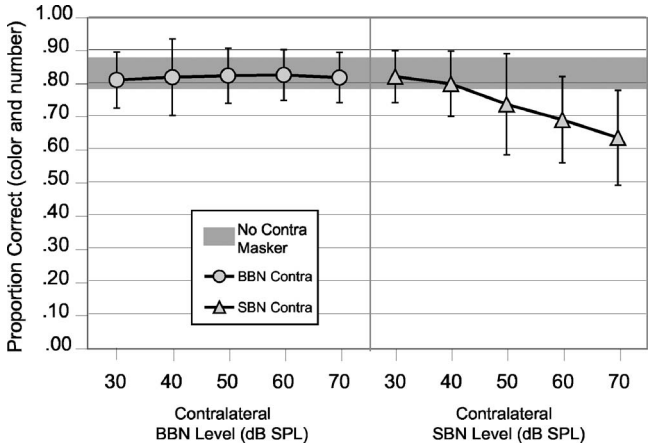


FIG. 5. Average keyword identification performance for four listeners in experiment three. The target was always an eight-band processed sentence presented to the right ear at a level of 50 dB SPL. Six-band different-band speech, reversed (DBSr) was also presented to the right ear (ipsilaterally) in all conditions. DBSr level was set individually for each listener in order to produce roughly equivalent performance when no masker was present (shaded area in each panel). The two panels represent performance as a function of contralateral masker level. Contralateral maskers were either broadband noise (BBN, left panel) or same-band noise (SBN, right panel). The shaded area in each panel indicates ±1 s.d. around mean performance across listeners with no contralateral masker present. Error bars on the symbols show ±1 s. d. across listeners.

tralateral conditions were tested, BBN and SBN (with six matched bands), each at five levels ranging from 30 to 70 dB SPL in 10 dB steps. Masker type and level was fixed for a block of 50 trials but the order of the blocks was randomized. All listeners completed 150 trials at each level for both conditions.

B. Results

Average results are plotted in Fig. 5. The shaded areas represent the region of performance obtained when no contralateral masker was presented (mean ±1 s.d.). A repeated-measures ANOVA was conducted on the data (including performance with no masker, resulting in six contralateral masker levels) and the results showed that there was an interaction between contralateral masker type (the two panels; BBN versus SBN) and contralateral masker level [$F(5,15)=8.92$, $p<0.001$]. Consequently, analyses of the effects of contralateral masker level were conducted separately for BBN and SBN. For BBN, the effect of contralateral masker level did not reach statistical significance [$F(5,15)=0.08$, $p=0.994$], while for SBN, the effect of level was significant [$F(5,15)=7.03$, $p<0.001$]. In order to determine which masker levels were responsible for the significant effect, an analysis was conducted examining the effect of masker type at each contralateral masker level. The results of that analysis indicated that the only significant difference was at a masker level of 70 dB [$F(1,3)=20.785$, $p<0.02$]. This statistical result was undoubtedly due to the wide variability in performance across listeners, as can be seen in Fig. 5.

C. Discussion

Despite the wide range of DBSr levels necessary to obtain sufficiently accurate proportion correct in the baseline task (a range of 20 dB) and the substantial variability across listeners, every listener showed a distinct downward trend in performance with increasing masker level for the SBN masker. This third example of contralateral masking demonstrates that even with an unintelligible masker that should not be easily mistaken for the target, the intelligibility of a monaurally presented six-band target is reduced by the presence of a frequency-matched noise masker presented to the opposite ear but not by a broadband noise. The results of experiment two argue strongly against a cross-ear masking explanation for such an effect. Similarly, the close match between the ipsilateral masking for BBN and SBN seen in experiment two argues against a difference in the within-band effectiveness of the two masker types. Currently, no computational models of binaural processing exist that can account for this pattern of results. While there is much that seems to be under voluntary control, these findings are quite difficult to reconcile with the view that the auditory system always makes use of the information that has the best signal-to-noise ratio. Were that the case, the complete indifference of the system to the stimulus at the left ear that occurred for the contralateral BBN masker would surely have been expected for the SBN masker as well. This was not the case.

The results of experiment three are also important because they support the mounting evidence that there can be substantial contralateral interference in speech identification even when the contralateral masker cannot be confused with the target. Evidence of such interference was also obtained by Brungart and Simpson (2002) with unprocessed reversed speech maskers. They concluded that "a listener's ability to ignore a masker in the contralateral ear is directly related to the difficulty of the within-ear segregation task in the target ear (p. 2993)." Their results (and those reported here) also agree with the conclusion reached by Brungart *et al.* (2005), who suggest that speech-on-speech masking is strongly influenced by "speech-like fluctuations in the spectral envelope" of the masker. This conclusion is supported by the analysis in Kidd *et al.* (2005) that showed that processed-speech maskers with modulation spectra matched to the processed-speech target were more effective than same-band noise maskers that had shallower dips in the envelope but a dissimilar modulation spectrum.

VI. GENERAL DISCUSSION

Taken together, the results of these three experiments provide strong evidence against any simple model of binaural processing in which listeners are always able to base decisions entirely on selecting the ear with the best signal-to-noise ratio. The most important influence on the ability to listen with independent ears is task-based demands on processing resources. In these experiments the task-based factor was the presence of multiple maskers that required more than one segregation strategy to distinguish the target from the maskers. Also important are stimulus-based auditory grouping factors, which can be thought of as the presence of information suggesting that target and masker arise from independent sources. Here, it was suggested that temporal similarity and spectral similarity are both important but that

temporal dissimilarity can be more potent than spectral similarity when no additional task is required (although ceiling effects reduced the strength of this conclusion).

The evidence that listeners are not always capable of "turning off" an ear in order to avoid interference has important implications for the interpretation of results in which performance is improved by presenting different target-to-masker ratios at the two ears. In particular, it suggests that any model that assumes that listeners always maximize performance by comparing the output of a binaural and a monaural strategy will be wrong under some listening conditions. This problem has also been noted by Shub and his colleagues (Shub and Colburn, 2004; Shub *et al.*, 2005; Shub, 2006). What remains to be determined is the extent to which observing such nonoptimal performance depends on stimuli that represent unusual deviations from what listeners normally experience in real-world environments. For one very important group of listeners at least, those with cochlear implants, the sort of narrow-band processing and monaural presentation used in these experiments are actually quite close to what they experience. This is because the processing is based on simulations of the signal transformations that allow the cochlear implant to directly stimulate the auditory nerve with sound information (Shannon *et al.*, 1995; Arbogast *et al.*, 2002). Consequently, these results may contain important information about the sorts of interference that might occur for listeners with two cochlear implants. Even for normal hearing listeners, the factors that were hypothesized to have led to the use of the nonoptimal binaural listening mode are quite common. In particular, the correlation of frequency content across ears is going to be present for most binaural stimuli. It should also be remembered that Brungart and Simpson (2002, 2004), who used unprocessed speech, reported more interference from speech signals than from noise.

The results of these three experiments seem to indicate the presence of two significant factors that give rise to contralateral interference. The first involves the complexity of the listener's task, while the second is based on the auditory system's propensity to treat similar stimuli as if they were generated by a single source in the environment. Task complexity, in this context, refers to what Norman and Bobrow (1975), Navon and Gopher (1979), and Wickens (1984) have all called demands on task-specific processing resources. The suggestion is that the human perceptual system is composed of multiple types of processing resources and that whether or not interference is observed depends upon the degree to which multiple tasks require access to the same resources. In this conceptualization, processing resources refers to everything from input channels or feature analyzers to system processes like retrieval of representations from short-term or long-term memory. Certainly task complexity or multiple demands for processing resources is a factor that would be present at many levels in a real-world environment.

The second factor that seems to have influenced listener performance in these experiments is the spectrotemporal similarity between the targets and the maskers. The aspects of similarity that were varied by the choice of stimuli included spectral similarity (in the use of on- and off-frequency maskers and narrow-band and wideband maskers) and ongoing temporal envelope similarity (as affected by the use of speech or noise envelopes for the maskers and narrow-band or wideband noise maskers).

In experiment one, listeners were provided with an eight-band target, but a different-band ipsilateral speech masker and a contralateral same-band noise masker were also present. To be identified, the target still had to be segregated from the spectrally different bands presented ipsilaterally and the spectrally similar bands presented contralaterally. If the same resources were required for all three tasks, performance would have been reduced by the need to perform all three. Even if the speech identification task required different resources, there would still have been a dual-task to be performed with the two segregation tasks.

In experiment two, the same eight-band target was used (among others), but without the ipsilateral different-band speech masker. Now the same-band noise had no effect when presented contralaterally. The same result was obtained with the six-band target, but significant contralateral masking was seen for the two- and four-band targets. In addition, there was no contralateral masking for any of the targets in the case of the broadband masker. This release from masking due to increasing numbers of bands can be explained if it is first assumed that the demands on processing resources were reduced by removing the ipsilateral speech masker. This allowed near-baseline performance to be achieved with the six- and eight-band targets. Because the number of target bands was reduced, however, demands on resources were increased, as demonstrated by performance levels obtained with no masker present (grey bands in Fig. 4). For the two- and four-band targets, the masking observed in the presence of the narrow-band maskers may be evidence that listeners were less able to achieve the segregation of the signals at the two ears. If some of the resources necessary to perform the ear-based segregation were being used to understand the two- and four-band speech, then it makes sense that performance declined in the presence of the contralateral maskers. What still needs to be explained is the difference between the BBN and the SBN, with no masking whatsoever occurring for the BBN.

While all of the stimuli were similar in their temporal onsets, which would have encouraged grouping by common onset (cf., Bregman, 1990), it was only those maskers that were spectrally similar to the targets (SBN as opposed to BBN) that led to contralateral masking. If it takes processing resources to overcome the auditory system's tendency to combine similar stimuli into a single auditory object, it seems plausible that the decreased intelligibility of the two-band target (compare grey bands in Fig. 4) may have drawn resources from the segregation task and thus increased the degree to which the ipsilateral and contralateral bands were combined into a single binaural object. Consequently, listeners may have experienced a binaural percept as they did with the perfectly correlated DBS maskers presented binaurally to subjects by Gallun *et al.* (2005). Because listeners had fewer cues indicating that the SBN and the target were from different auditory sources than they did for BBN and the target, it is possible that a combined binaural signal was created for the two- and four-band targets, but only for the SBN stimuli.

The ongoing changes in envelope would have provided the strongest source of information indicating that this binaural combination should not occur and as this information was increased (by increasing the number of bands), so the amount of contralateral masking decreased. Clearly, more research is required to understand the interaction between the availability of processing resources and degree to which listeners erroneously combine targets and maskers into single objects, thus removing ongoing differences in pitch, location, and intensity between target and masker.

In experiment three, a six-band target was significantly less intelligible in the presence of a masker with speech-like modulations (DBSr) at the ipsilateral ear and a SBN at the contralateral ear. This decrease in performance did not occur with the same ipsilateral DBSr masker and BBN at the contralateral ear. This finding supports the idea that listener performance is influenced both by the processing load imposed, which explains why a six-band target suffered in experiment three but not experiment two, and by the spectrotemporal similarity between target and masker, which explains why the similar SBN hurt performance but the dissimilar BBN did not.

As a final note, it is worth examining the question of whether the large individual differences seen here (and in many other studies of informational masking) should be considered as evidence that different listeners were employing different strategies to solve the tasks with which they were presented. While this is probably an appropriate description of some studies (see, for example, Oxenham *et al.*, 2003), what characterized the use of different strategies in that case was a difference in the *pattern* of results, with some listeners performing similarly on two tasks and others performing quite differently. The listeners in these experiments, however, differed only in the level of performance that they achieved or, in the case of the third experiment, the level of masking noise that was needed to achieve a particular level of performance. Thus, there is evidence that some listeners were better at applying the segregation strategies than others, but not that the actual strategies differed among listeners. It should be noted, however, that these experiments were not designed with the goal of examining individual differences. Future work in this area would benefit from an emphasis on designing experiments and computational models that would allow differences in performance that can be obtained with a single strategy to be clearly differentiated from the use of different strategies by different listeners.

VII. SUMMARY

This series of three experiments provides further support for the conclusion that listeners are not always able to choose a monaural better-ear listening strategy when it would be helpful to do so.

(i) In experiment one, performance decreased relative to performance in the presence of a DBS masker presented ipsilaterally when a SBN masker matched in long-term spectrum to the target was presented contralaterally.

(ii) In experiment two, interference from both ipsilateral and contralateral SBN maskers was observed for two- and four-band targets but only ipsilateral interference occurred for six- and eight-band targets. Only ipsilateral interference was observed for BBN maskers, but it occurred for all numbers of target bands.

(iii) In experiment three, the presence of different-band speech presented ipsilaterally and reversed in time (DBSr) resulted in contralateral interference for SBN maskers but not BBN maskers when the target had six bands and the DBSr was presented at a level that resulted in a minimum amount of masking for all listeners.

These results support the conclusion that the findings of Kidd *et al.* (2005) in which release from DBS masking occurred when a DBN masker was presented contralaterally represented a true case of contralateral interference and not simply cross-ear masking. Nor are those results entirely attributable to the generation of a binaural masker image, although creation of such an image cannot be ruled out. The results of experiment two, in particular, suggest that listeners may indeed combine information across ears, especially when the spectrotemporal similarity of the information at the two ears is high and the processing demands of the task are also high. The results of experiment three provide additional support for the importance of both task-based factors such as processing demands and stimulus-based auditory grouping factors, such as spectrotemporal similarity, in determining whether or not contralateral interference will occur.

ACKNOWLEDGMENTS

The first experiment evolved out of a suggestion by Virginia Richards, to whom we are gratefully indebted. We also would like to thank Antje Ihlefeld and Douglas Brungart for comments on an earlier draft of this manuscript. We are also grateful for the assistance of our listeners as well as to Kelly Egan and Jackie Stachel for assistance with data collection. This work was supported by the Department of Veterans Affairs, Veterans Health Administration, Rehabilitation Research and Development Service through Associate Investigator Award No. C4855H to Frederick Gallun at the National Center for Rehabilitative Auditory Research as well as AFOSR Award No. FA9550-05-1-2005 and by Grant Nos. DC00100, DC04545, and DC04663 from NIH/NIDCD. F.G. was also supported by F32 DC006526 from NIDCD.

ANSI (**2004**). ANSI S3.6-2004, "American National Standard Specification for Audiometers," (American National Standards Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

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.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Broadbent, D. E. (**1958**). *Perception and Communication* (Pergamon, New York).

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Iso-

lating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.
Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.
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.
Brungart, D., and Simpson, B. (**2007**). "The effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **122**, 1724–1734.
Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd, G., Jr. (**2005**). "Across-ear interference from parametrically-degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **118**, 292–304.
Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.
Colburn, H. S., and Durlach, N. I. (**1978**). "Models of binaural interaction," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. **IV**.
Durlach, N. I. (**1960**). "Note on the equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **32**, 1075–1076.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G., Jr. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.
Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.
Gallun, F. J., Mason, C. R., and Kidd, G., Jr. (**2005**). "Binaural release from informational masking in a speech identification task," J. Acoust. Soc. Am. **118**, 1614–1625.
Heller, L. M., and Trahiotis, C. (**1995**). "Interference in detection of interaural delay in a sinusoidally amplitude-modulated tone produced by a second, spectrally remote sinusoidally amplitude-modulated tone," J. Acoust. Soc. Am. **97**, 1808–1816.
Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.
Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.
Koehnke, J., and Besing, J. M. (**1992**). "Effects of roving level variation on monaural detection with a contralateral cue" J. Acoust. Soc. Am. **92**, 2625–2629.
Moray, N. (**1970**). "Introductory experiments in auditory time sharing: Detection of intensity and frequency increments," J. Acoust. Soc. Am. **47**, 1071–1073.
Navon, D., and Gopher, D. (**1979**). "On the economy of the human-processing system," Psychol. Rev. **86**, 214–255.
Norman, D. A., and Bobrow, D. G. (**1975**). "On data-limited and resource-limited processes," Cognit Psychol. **7**, 44–64.
Oxenham, A., Fligor, B. J., Mason, C. R., and Kidd, G., Jr. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shub, D. E. (**2006**). "Monaural perception under dichotic conditions," Ph.D. thesis, Health Sciences and Technology, MIT, Cambridge.
Shub, D. E., and Colburn, H. S. (**2004**). "Monaural intensity discrimination under dichotic conditions," J. Assoc. Res. Otolaryngol. (Abstracts).
Shub, D. E., Pogal-Sussman, T. E., and Colburn, H. S. (**2005**). "The effects of distractor frequency on monaural intensity discrimination under monotic and dichotic conditions," J. Assoc. Res. Otolaryngol. (Abstracts).
Sorkin, R. D. (**1965**). "Uncertain signal detection with simultaneous contralateral cues," J. Acoust. Soc. Am. **38**, 207–212.
Studebaker, G. A. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Taylor, M. M., and Forbes, S. M. (**1969**). "Monaural detection with a contralateral cue (MDCC). I. Better than energy detector performance by human observers," J. Acoust. Soc. Am. **46**, 1519–1526.
Treisman, A. (**1964**). "Verbal cues, language and meaning in selective attention," Am. J. Psychol. **77**, 206–219.
Treisman, A. M. (**1969**). "Strategies and models of selective attention," Psychol. Rev. **76**, 282–299.
Wang, D. and Brown, G. J., eds. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms and Applications* (IEEE Press/Wiley-Interscience, Hoboken, NJ).
Wickens, C. D. (**1984**). "Processing resources in attention," in *Varieties of attention*, edited by R. Parasuraman and D. R. Davies (Academic Press, New York), pp. 63–102.
Wood, N., and Cowan, N. (**1995**). "The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953)," J. Exp. Psychol. Gen. **124**, 243–262.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by Studebaker, G. and Hochberg, I. (College-Hill Press, Boston).

Perception of across-frequency interaural level differences

Tom Francart[a] and Jan Wouters[b]
Exp ORL, Dept. Neurosciences, Katholieke Universiteit Leuven, O&N2, Herestraat 49 Bus 721, B-3000 Leuven, Belgium

(Received 27 June 2006; revised 16 August 2007; accepted 17 August 2007)

The interaural level difference (ILD) is an important cue for the localization of sound sources. Just noticeable differences (JND) in ILD were measured in 12 normal hearing subjects for uncorrelated noise bands with a bandwidth of 1/3 octave and a different center frequency in both ears. In one ear the center frequency was either 250, 500, 1000, or 4000 Hz. In the other ear, a frequency shift of 0, 1/6, 1/3, or 1 octave was introduced. JNDs in ILD for unshifted, uncorrelated noise bands of 1/3 octave width were 2.6, 2.6, 2.5, and 1.4 dB for 250, 500, 1000, and 4000 Hz, respectively. Averaged over all shifts, JNDs decreased significantly with increasing frequency. For the shifted conditions, JNDs increased significantly with increasing shift. Performance on average worsened by 0.5, 0.9, and 1.5 dB for shifts of 1/6, 1/3, and 1 octave. Though performance decreases, the just noticeable ILDs for the shifted conditions were still in a range usable for lateralization. This has implications for signal processing algorithms for bilateral bimodal hearing instruments and the fitting of bilateral cochlear implants. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2783130]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Ts, 43.64.Me [AK] Pages: 2826–2831

I. INTRODUCTION

An important cue for localization of sound sources is the difference in sound pressure level between both ears, the so-called interaural level difference (ILD). The ILD is considered especially useful for localization at higher frequencies (>1000 Hz). As the wavelength of low-frequency sounds is long compared with the size of the head, low-frequency sounds "bend" around the head, resulting in a small ILD. Although ILDs are very small below about 500 Hz, they may be as large as 20 dB at high frequencies (Moore, 2003).

Nevertheless, the human auditory system is able to perceive ILDs in the low frequencies with just noticeable differences (JNDs) as small as ±1 dB, measured with pure tones (Mills, 1960; Yost and Dye, 1988). Low frequency ILD cues are used for localizing nearby sources (Brungart, 1999; Brungart *et al.* 1999), in the so-called "proximal region," the region within 1 m of the center of the head.

Mills (1960) presented five subjects with a reference stimulus with no ILD, followed by a stimulus with an ILD. The stimuli were pure tones. Using the method of constant stimuli, the JND in ILD was determined from half the interquartile separation of the psychometric curves for each subject. JNDs were around 1dB for 1000 Hz, somewhat smaller for lower frequencies and around 0.5 dB for frequencies higher than 1000 Hz.

Yost and Dye (1988) measured JNDs in ILD for pure tones and different reference signals at 75% correct, using a linear fit of the psychometric curve. For the reference at ILD=0 dB they found JNDs of approximately 0.75, 0.85, 1.20, 0.70 and 0.73 dB for 200, 500, 1000, 2000, and 5000 Hz. In the 2AFC procedure, subjects heard one stimulus on the right-hand side and one on the left-hand side and had to respond which one was on the right.

Hartmann and Constan (2002) tested the hypothesis of the level meter model: Can the ILD be seen as an integrated measure of stimulus energy, independent of stimulus details? Differences between correlated and uncorrelated stimuli were assessed using white noise and low pass filtered noise (<1000 Hz). A 2AFC, 1 up/3 down adaptive procedure, targeting the 79% correct point was used and subjects had to determine the direction of change (right-to-left or left-to-right). They conclude that the level meter model is sound within half a decibel, i.e., the thresholds for each of the four tested correlation conditions were within 0.5 dB of each other. JNDs for uncorrelated white noise are in the order of 0.6 dB. For the low pass noise condition they are in the order of 0.9 dB.

Users of a combination of a cochlear implant (CI) in the one ear and a hearing aid in the other (a so-called bilateral bimodal system), are in the situation where normal (high-frequency) ILD cues are absent, because their residual hearing in the acoustically aided ear in many cases does not extend beyond 500 Hz. Because of the different nature of stimulation in both ears and current signal processing schemes for cochlear implants, they also do not have access to precise interaural time difference (ITD) cues for localization (Ching *et al.*, 2005; Seeber *et al.*, 2002).

As in most practical situations sound sources are not in the proximal region, bimodal hearing aid users are extremely bad at localizing sound sources (Ching *et al.*, 2004, 2001; Seeber *et al.*, 2004; Tyler *et al.*, 2002). This could be improved by artificially introducing or amplifying ILDs in the low frequencies, even if the latter are not present in the incoming auditory signal.

For many bimodal hearing instrument users, it is probably not possible to stimulate the same frequency regions in

[a]Electronic mail: tom.francart@med.kuleuven.be
[b]Electronic mail: jan.wouters@med.kuleuven.be

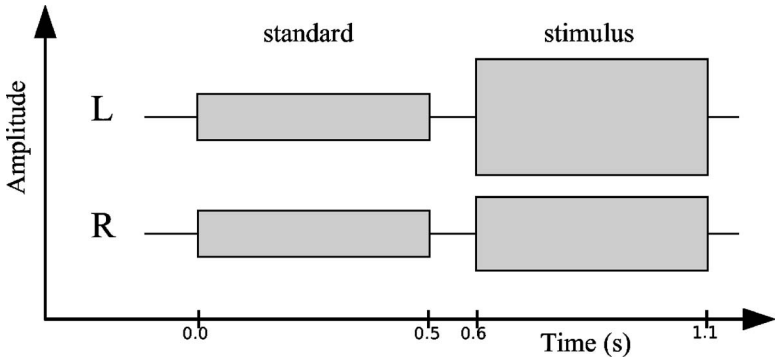


FIG. 1. Example of a standard-stimulus sequence with a positive rove. For this trial, the correct answer would be "The stimulus sounded on the left-hand side of the standard."

both cochleas, because on the one hand the tip of the electrode array of the cochlear implant does not reach the apex of the cochlea, so the lowest frequencies cannot be stimulated. On the other hand, the usable residual hearing of the other ear does in many cases not extend beyond 500 Hz. This is illustrated by the pitch-matching data of Boex *et al.* (2006). They measured the acoustic pitch corresponding to the place pitch elicited by stimulation of certain electrodes of the cochlear implant in six users of bimodal systems. For the most apical electrode of each subject, they found pitches of 460, 100, 290, 260, 570, and 300 Hz. Though these pitches are lower than could be expected based on Greenwood's function (1990), they are in most cases still high enough to make it difficult to stimulate the same acoustical frequencies in the other ear.

For users of bimodal systems, due to current clinical fitting procedures, the cochlear implant in the one ear and the hearing aid in the other will most certainly not stimulate the same frequency ranges. The same is true for users of bilateral CIs: Currently both CIs are fitted more or less independently and the electrode positions along the left and right basilar membrane are not tuned to the same frequencies.

The object of the work described in this article was to assess whether it is possible for normal hearing subjects to perceive ILDs for different degrees of frequency mismatch between the signals in both ears. Therefore, JNDs in ILD were determined for different frequency shifts in one of the two ears. This was done for different base frequencies, using uncorrelated noise band stimuli to simulate the difference in stimulation between the acoustic and electric part of a bimodal system and to eliminate potentially confusing ITD cues. Note that uncorrelated stimuli result in a diffuse sound image that is not externalized, i.e., it is perceived inside the head. This makes the task harder (Hartmann and Constan, 2002), but also more realistic when considering binaural bimodal hearing systems, where subjects are presented with largely uncorrelated signals.

Similar work for ITDs was done by Nuetzel and Hafter (1981) and Saberi (1998). They tested subject sensitivity to interaural delay in the envelope of respectively high-frequency amplitude modulated sinusoids and frequency modulated sinusoids and found that as the carrier frequency difference increases, time differences are still detected, but performance drops rapidly. Given that critical bands in binaural experiments have a bandwidth similar to estimates in monaural experiments (Breebaart *et al.*, 2001; Holube *et al.*, 1998), we expect performance detecting ILDs to deteriorate when large frequency shifts are introduced.

II. METHODS

A. Procedure

1. General procedure

The JND in ILD was determined for each condition using several runs of an adaptive 1 up/2 down procedure, targeting the 71% correct point. The procedure determined the ILD of the stimulus that was presented. The start value was 10 dB and the initial stepsize was 2 dB. After 2 reversals, the stepsize was decreased to 0.4 dB and after 10 reversals to 0.2 dB. The procedure continued until 12 reversals were obtained. No feedback was given. The mean of the last 6 reversals was taken as the JND for a certain run. If the procedure saturated, i.e., the parameter was 10 or 0 dB, the run was discarded and repeated.

In each trial, a standard was presented first, which contained no ILD, followed by a short pause of 0.1 s, followed by the stimulus that contained a certain ILD. The ILD pointed with equal probability to left or right and the magnitude was selected according to the parameter adapted by the adaptive procedure. The subjects had to respond whether they heard the stimulus on the left or right side of the standard. One specific case is illustrated in Fig. 1.

Two experiments were done. In the first experiment, to avoid subjects using monaural cues, the overall stimulus level was roved uniformly over ±5 dB. It can be shown that in this case, a JND of 4.2 dB could theoretically be attained by only attending to one ear. Because some of the found JNDs were larger than 4.2 dB, a second experiment was done with a level rove of ±10 dB.

Subjects were instructed to respond whether they heard the stimulus on the left or right side of the standard. If they were not able to lateralize, they were encouraged to compare the left and right loudness levels. They were also asked to close their eyes during the runs to avoid visual disturbances (there are indications that visual cues can influence responses on localization tests (Lewald and Getzmann, 2006)). They responded using the left and right arrow keys of a computer keyboard. The experiments were unattended by the experimenter, except for the introduction to the task and regular checks. One run took, depending on the subject, between 78

and 388 s, with a median of 160 s. This resulted in an average total time of 3.5 h or more of testing per subject in experiment 1, excluding any breaks or short pauses between different runs. The subjects participating in experiment 2 were tested for an additional 1.5 h.

2. Conditions

JNDs in ILD were determined for four base frequencies: 250, 500, 1000, and 4000 Hz. The most relevant base frequencies for the application on bimodal hearing are 250 and 500 Hz, because the residual hearing of most subjects that use a bimodal hearing system is restricted to the low frequencies. The 1000 and 4000 Hz base frequencies were added as higher frequency reference conditions.

For one adaptive run, the center frequency of the one ear was always one of the base frequencies and the center frequency of the other ear was the base frequency shifted with 0, 1/6, 1/3, or 1 oct. As noise bands of 1/3 oct wide were used, this results in, respectively, full overlap, partial overlap, marginally no overlap, and no overlap at all of the shifted noise band with the base noise band. The shifts were performed in upward direction.

Per subject, two of the base frequencies and all of the shifts were presented for each selected base frequency. A condition consists of a certain base frequency combined with a certain shift. In experiment 1 each condition was presented 8 or 10 times and in experiment 2 it was presented 4 times. To minimize the chance of training effects influencing only a single condition, conditions were always interleaved.

B. Experimental setup

1. Stimuli and test setup

The stimuli were noise bands of 1/3 oct wide, filtered with a 50th-order Butterworth filter to ensure a minimal amount of overlap beyond the cutoff frequencies of the noise bands presented to both ears. To avoid confusing interaural time difference cues, noise bands were at all time instants uncorrelated between both ears and new noise bands were generated for each standard and each stimulus. Linear in and out ramping was performed over 0.2 s to avoid clicks and confusing onset cues. The total stimulus duration was 1 s.

For every run, the ear to be presented with the frequency-shifted stimulus was selected at random. On the whole, each ear was presented an equal number of times with the unshifted stimulus.

To obtain an approximately centered reference signal, the left and right channels were equalized in root mean square (RMS) level with respect to the A-weighted dB scale. In this way, the left and right channels sounded approximately equally loud, so the reference signal was centered in the head. Note that as a consequence of the dBA weighting, especially at the lower frequencies, the levels of the channels differ between ears if measured in dB sound pressure level (SPL) in conditons with frequency-shifted noise bands.

The ILD was introduced as follows: If S_L and S_R are the levels of the left and right channels of the standard measured in dSPL, I is the ILD to be introduced, r is the rove level, randomly selected from the interval $[-5,5]$ or $[-10,10]$, and L_L and L_R are the levels of the left and right channels of the stimulus, the stimulus was generated according to the following equations (all in dSPL):

$$L_L = S_L + I/2 + r \tag{1}$$

$$L_R = S_R - I/2 + r \tag{2}$$

If the same center frequency was presented at both ears, when measuring absolute levels in dSPL, S_L was the same as S_R and the ILD was I. If different center frequencies were presented, S_L and S_R differed because of the dBA weighting, and the resulting ILD was $I+S_L-S_R$.

All stimuli were presented in a soundbooth using the APEX (Laneau *et al.*, 2005) program running on a personal computer, driving a LynxOne sound card that was connected via a mixer to a set of Sennheiser HD250 Linear II headphones.

Calibration of the left and right channels was done by setting the mixer such that a 1/3 oct noise band with a center frequency of 1000 Hz had an overall RMS level of 65 dB-SPL. The level of the other stimuli in dBA was equal to the level of the 1000 Hz stimulus in dBA.

2. Subjects

Twelve subjects participated in experiment 1 and came to the lab for 3 or 4 sessions of 1–2 h. Six of those subjects participated in experiment 2 and came to the lab for an additional 1 or 2 sessions.

All subjects were volunteers and were paid for their cooperation. Their hearing was normal, except for one subject who had a threshold of 40 dBHL (hearing level) at 4000 Hz. He was only presented the conditions with base frequencies 250 and 500 Hz and only participated in experiment 1. Two subjects were male, ten female and all between 18 and 28 years of age.

III. RESULTS

A. Experiment 1

JNDs in ILD were repeatedly measured for all base frequencies and all frequency shifts. To assess possible training effects, the sequence of results of runs for each frequency/shift condition of each subject is shown in Fig. 2. Each sequence was normalized by dividing by the mean of the last 6 runs in that sequence. The full line connects the averages at each time instant. No clear average long term training effect is evident from Fig. 2. Also, no clear training effect can be seen for any of the subjects separately. As there seems to be a small effect in the first few runs, the first two measurements of each condition were discarded from further analysis.

A summary of the results of experiment 1 is presented in Fig. 3. Results are shown per base frequency and frequency shift, but averaged over all runs and over all subjects. The error bars are at least partly due to intersubject variance, as opposed to intra-subject variance, as was seen from an ANOVA. The JND in ILD increases per increasing shift (i.e., it is harder to discriminate loudness differences when the

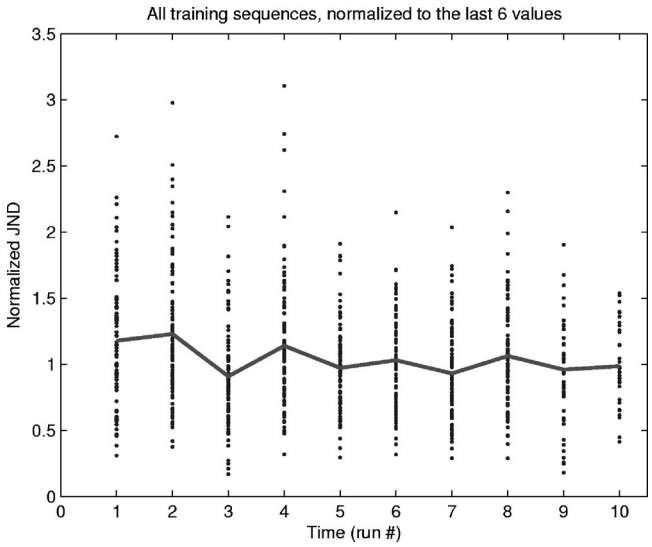


FIG. 2. All normalized sequences of runs of experiment 1. All values of each sequence were divided by the mean of the last 6 runs of that sequence. Each dot represents the result of an adaptive run. The full line connects the averages at each time instant.

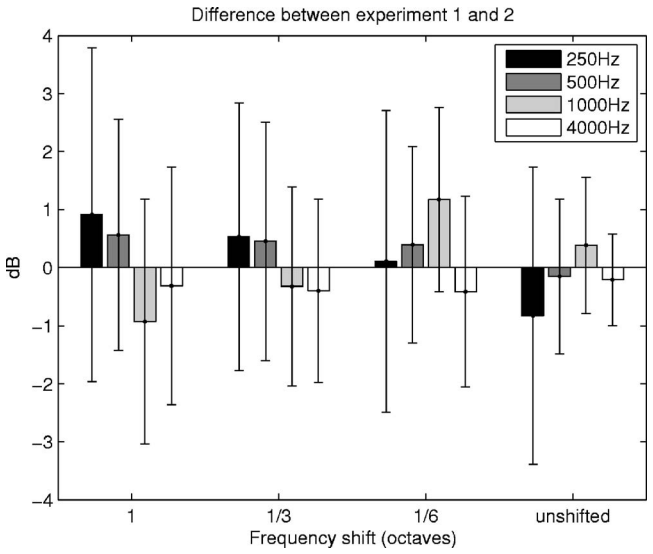


FIG. 4. Differences between experiments 1 and 2. The bars show the difference in JND. The error bars represent the combined error of both experiments. Positive values indicate that the JND in experiment 1 (±5 dB rove) was larger than the JND in experiment 2 (±10 dB rove).

frequencies in both ears are less similar) and the JND decreases per increasing base frequency (i.e., it is easier to discriminate ILDs when the center freqencies in both ears are higher). All frequency conditions differ significantly from each other $[F(3,391)=25.8,\ p<0.00001]$ as well as all shift conditions $[F(3,391)=39.9,\ p<0.00001]$ except for the shifts of 1/3 and 1/6 oct.

As the JNDs for the one octave shift conditions are in the neighborhood of the 4.2 dB value that could theoretically be monaurally attained when using a rove of ±5 dB, this experiment was repeated in experiment 2 with a rove of ±10 dB.

B. Experiment 2

The small training effect in the first few runs of experiment 1 is not observed in the results from experiment 2. This is probably due to the fact that all 6 subjects who participated in experiment 2 also participated in experiment 1 during about 3 h. Therefore no measurements were discarded from experiment 2 based on training effects.

An ANOVA with factors subject, frequency and shift indicates a significant effect for shift $[F(3,161)=24.5,\ p<0.0001]$. Post hoc analysis with Bonferroni correction shows that all shift conditions differ significantly from each other, except for the shifts of 1/3 oct with 0 oct and 1/6 oct with 0 oct.

Figure 4 shows the differences in threshold values between experiment 1 and 2. On average the JND increased by 0.06 dB toward experiment 2. This difference is however not significant in an analysis of variance with extra factor experiment. In what follows, we will therefore focus on the results of experiment 1 because it was performed with more subjects and most results are below the 4.2 dB threshold anyway.

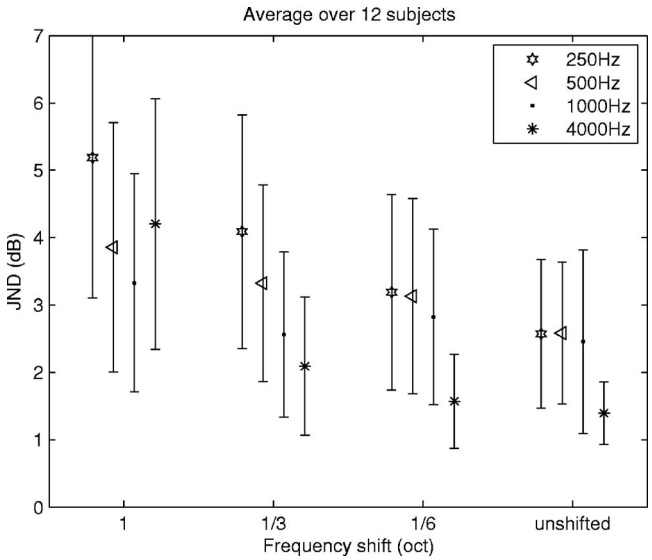


FIG. 3. JNDs in ILD (dB) as a function of base frequency and frequency shift for experiment 1 (±5 dB rove). The total length of the error bar is twice the standard deviation. The data were checked for normality using the Kolmogorov–Smirnov test.

IV. DISCUSSION

Figure 3 shows that the JND in ILD increases per increasing shift and the JND decreases per increasing base frequency. The unshifted conditions yield JNDs of 2.6, 2.6, 2.5, and 1.4 dB for 250, 500, 1000, and 4000 Hz respectively. Hartmann and Constan (2002) report a JND of 0.6 dB for white noise stimuli and 0.9 dB for low pass noise (<1000 Hz). Their procedure is similar to ours, but to compare the results, their values have to be multiplied by a factor of 2 to compensate for the difference in definition of ILD. Translating their results yields ILDs of, respectively, 1.2 and 1.8 dB Further differences are due to the fact that, in our experiments, noise bands of a much smaller bandwidth are used. Hartmann and Constan (2002) observe that, for both bandwidths used, JNDs decrease (i.e., performance increases) when the bandwidth increases. Buus (1990) reports that the JNDs for monaural level discrimination decrease when the bandwidth increases. He however used different

stimuli: both ears were stimulated sequentially, whereas in this study both ears were stimulated simultaneously.

When considering the results in terms of frequency overlap between the ears, it can be seen that as soon as the overlap decreases by 1/6 oct, performance decreases significantly. Further decreasing the overlap by 1/3 oct does not yield a significant change versus the 1/6 oct shift. This can be explained by the fact that although physically the spectra of the unshifted and 1/3 oct shifted noise band are nearly perfectly separated, there is some spread in the excitation patterns in the cochlea, resulting in a certain amount of overlap. The 1 oct shifted noise band yields significantly worse performance than all other shift conditions, caused by even less overlap in the excitation patterns in the cochlea.

Though significantly larger for the shifted conditions, JNDs are still in a range usable for lateralization of sound sources. The results for the shifted conditions partly confirm the simple level meter model proposed in Hartmann and Constan (2002). The results roughly confirm that the auditory system integrates energy over different frequencies, even over critical band boundaries. Performance however worsens on average by 0.5, 0.9, and 1.5 dB for shifts of respectively 1/6, 1/3, and 1 oct, relative to the unshifted condition. This refines the level meter model.

According to Hartmann and Rakerd (1989) the interpretation of our results could be complicated by the fact that the subjects could have ignored the standard that was presented before each stimulus and compared the stimuli to each other, resulting in a larger ILD cue than when comparing the stimulus to the standard. However, this seems unlikely because (1) in contrast to Hartmann and Rakerd (1989), we used level roving, making stimuli with the same ILD sound differently, (2) the subjects were repeatedly encouraged to always listen carefully to the standard, (3) an adaptive procedure was used, resulting in a reduction of the effect described by Hartmann and Rakerd (1989) and (4) the results of our unshifted baseline condition correlate well with the results found in the literature. Moreover, even if the absolute values of our results would not be accurate, this does not influence the main conclusions which are based on comparisons between conditions, unless the subjects would have changed detection strategies between conditions, which seems unlikely.

Though we did not directly measure whether subjects were able to lateralize the stimuli or rather compared level differences between both ears, we did ask them how they did it for each condition. All 12 subjects reported being able to lateralize in all conditions except for the 1 oct shift. In the 1 oct shift condition, they reported to "sometimes" attend to level differences instead of lateralizing. This attending to level differences can indicate a non-fused image which might be part of the cause of the increased JNDs in the 1 oct shift condition versus the other shift conditions.

V. CONCLUSIONS

From our JND in ILD experiments with 12 normal hearing subjects, we can conclude that

- ILDs can be detected for uncorrelated narrowband (1/3 oct) noise, with JNDs in the range 1.4–5.2 dB and
- When a frequency shift is introduced in one of both ears, ILDs can still be detected, albeit with a slightly higher JND.

The fact that ILDs can be detected across frequencies has important implications for localization using bilateral cochlear implants and contralateral bimodal systems.

For bilateral CIs, it means that bilateral matching of electrodes is less important for ILD perception than could be assumed (though performance is still best for the unshifted condition).

For bilateral bimodal systems, it implies that lateralization using ILDs might be improved by introducing or amplifying ILD cues between the acoustical part (the hearing aid) and the low-frequency electrodes of the electrical part. A signal processing system that has access to full band signals of both ears could determine the direction of a prominent sound source and use that direction to calculate a corresponding ILD to introduce in the low frequencies. The patient would then have to be trained to localize sound sources using these artifical ILD cues.

ACKNOWLEDGMENTS

This research was carried out in the frame of the IWT (Institute for the Promotion of Innovation by Science and Technology in Flanders) Project No. 050445 and was partly sponsored by Cochlear Ltd. The authors thank the test subjects for their patient and enthusiastic participation to the experiments. They also thank the editor and three reviewers for their helpful criticism on earlier versions of the manuscript.

Boex, C., Baud, L., Cosendai, G., Sigrist, A., Kos, M. I., and Pelizzone, M. (**2006**). "Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing," J. Assoc. Res. Otolaryngol. **7**(2), 110–124.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.

Brungart, D. S. (**1999**). "Auditory localization of nearby sources. III. Stimulus effects," J. Acoust. Soc. Am. **106**, 3589–3602.

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. II. Localization of a broadband source," J. Acoust. Soc. Am. **106**, 1956–1968.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

Ching, T., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Ching, T. Y., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (**2001**). "Should children who use cochlear implants wear hearing aids in the opposite ear?," Ear Hear. **22**, 365–380.

Ching, T. Y., van Wanrooy, E., Hill, M., and Dillon, H. (**2005**). "Binaural redundancy and inter-aural time difference cues for patients wearing a cochlear implant and a hearing aid in opposite ears," Int. J. Audiol. **44**, 513–521.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hartmann, W., and Constan, Z. (**2002**). "Interaural level differences and the level-meter model," J. Acoust. Soc. Am. **112**, 1037–1045.

Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle–A decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Laneau, J., Boets, B., Moonen, M., van Wieringen, A., and Wouters, J. (**2005**). "A flexible auditory research platform using acoustic or electric

stimuli for adults and young children," J. Neurosci. Methods **142**, 131–136.
Lewald, J., and Getzmann, S. (**2006**). "Horizontal and vertical effects of eye-position on sound localization," Hear. Res. **213**, 99–106.
Mills, A. W. (**1960**). "Lateralization of high-frequency tones," J. Acoust. Soc. Am. **32**, 132–134.
Moore, B. C. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Elsevier Science Amsterdam, The Netherlands), Chap. 7.
Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.
Saberi, K. (**1998**). "Modeling interaural-delay sensitivity to frequency modulation at high frequencies," J. Acoust. Soc. Am. **103**, 2551–2564.
Seeber, B., Baumann, U., and Fastl, H. (**2004**). "Localization ability with bimodal hearing aids and bilateral cochlear implants," J. Acoust. Soc. Am. **116**, 1698–1709.
Tyler, R., Parkinson, A., Wilson, B., Witt, S., Preece, J., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: speech perception and localization," Ear Hear. **23**, 98–105.
Yost, W., and Dye, R., Jr. (**1988**). "Discrimination of interaural differences of level as a function of frequency," J. Acoust. Soc. Am. **83**, 1846–1851.

Estimation of interaural level difference based on anthropometry and its effect on sound localization

Kanji Watanabe[a)]
Faculty of Engineering, University of Yamanashi, 84–4 Ebinokuchi, Tsuchiya, Yori-Honjo, Akita 0150055 Japan

Kenji Ozawa
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 4–4–37 Takeda, Kofu, Yamanashi 400–8510 Japan

Yukio Iwaya and Yôiti Suzuki
Research Institute of Electrical Communication/Graduate School of Information Sciences, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai 9808577 Japan

Kenji Aso
Faculty of Engineering, University of Yamanashi, 4–4–37 Takeda, Kofu, Yamanashi 400–8510 Japan

(Received 17 December 2006; revised 23 August 2007; accepted 24 August 2007)

Individualization of head-related transfer functions (HRTFs) is important for highly accurate sound localization systems such as virtual auditory displays. A method to estimate interaural level differences (ILDs) from a listener's anthropometry is presented in this paper to avoid the burden of direct measurement of HRTFs. The main result presented in this paper is that localization is improved with nonindividualized HRTF if ILD is fitted to the listener. First, the relationship between ILDs and the anthropometric parameters was analyzed using multiple regression analysis. The azimuthal variation of the ILDs in each 1/3-octave band was then estimated from the listener's anthropometric parameters. A psychoacoustical experiment was carried out to evaluate its effectiveness. The experimental results show that the adjustment of the frequency characteristics of ILDs for a listener with the proposed method is effective for localization accuracy.
 [DOI: 10.1121/1.2785039]

PACS number(s): 43.66.Qp, 43.66.Pn [RAL] Pages: 2832–2841

I. INTRODUCTION

We can present a virtual sound environment to listeners with simple devices such as headphones if spatial sound is synthesized with a high degree of accuracy. Head-related transfer functions (HRTFs) are known to include comprehensive auditory cues for sound source positions, including source azimuth, elevation, and distance to a listener (Blauert, 1983). Because the reflection and diffraction around the external ear, head, and torso of the listener influence HRTFs, they vary as a function of source position and show large inter-subject variation. Therefore, implementation of a high-definition sound localization system can be achieved using a set of HRTFs of the listener. Known problems are that the listener cannot localize the sound image accurately (Asano *et al.*, 1990; Bronkhorst, 1995; Kawaura *et al.*, 1991; Martens, 2003; Morimoto and Nomachi, 1982), and front-back confusion occurs (Kistler and Wightman, 1992; Makous and Middlebrooks, 1990; Middlebrooks, 1992; Wenzel *et al.*, 1993) if the HRTFs used are degraded or are not identical to the listener's own. Shaw and Teranishi (1968) used an artificial model to investigate the resonances of the pinna, which corresponded to the features of HRTFs. Haneda *et al.* (1999) proposed the common acoustical pole and zero model for modeling of a set of HRTFs. Using their model, they showed that the resonances of the pinna corresponded to the common acoustical poles, which were independent of source positions. Genuit (1984) first studied what features of the human body and pinna contributed to characterize the HRTFs for the purpose of HRTF estimation. Algazi *et al.* (2001) analyzed the relationship between the spectral features and the size of anthropometric parameters with their HRTF database. Inoue *et al.* (2005) proposed a method of estimating the HRTFs because they are influenced by the listener's anthropometry, and reported that the HRTFs were estimated based on the physical features of the bandwidth of 0–8 kHz.

It would not be necessary to estimate the HRTFs perfectly if localization cues that are included in the HRTFs were synthesized appropriately. The interaural time difference (ITD), interaural level difference (ILD), and the spectral structure are well known localization cues. Both ITD and ILD are extremely important cues for localization in the horizontal plane, whereas the spectral structure is very important for localization in the sagittal plane, in which the variations of ITD and ILD are slight (Hebrank and Wright, 1974; Morimoto, 1999). The roles of ITD and ILD in localization are different. Wightman and Kistler (1992) showed that ITD was dominant when stimuli including low-frequency compo-

[a)]Present address: Faculty of Systems Science and Technology, Akita Prefectural University, Akita 0150055, Japan. Electronic mail: kwatanabe@akita-pu.ac.jp

nents were present, although ILD was dominant if low-frequency components were removed from the stimuli. Because the spectral structure is generally complex, estimation is considered difficult. Both ITD and ILD vary as functions of the difference of paths from the source to the right or left ear. Therefore, those cues are related closely to anthropometric parameters such as the head size and the pinna position. Several studies have investigated a model or estimation of ITD and ILD. Algazi *et al.* (2002) proposed a spherical model to approximate the head and torso for a numerical model of ITD. Phillips and Hall (2005) examined the process in the auditory system for interaural cues. Estimation of ITD based on a HRTF corpus (Watanabe *et al.*, 2005a) has been investigated; moreover, improvement in azimuthal localization by fitting ITD to each listener has been demonstrated (Watanabe *et al.*, 2005b). On the other hand, because ILD is frequency dependent, estimation of ILD is expected to fit a spectral cue, which is not existent in ITD for each listener. For example, if the frequency characteristics of ILD reflect the front-back asymmetry of the human body, that might be a cue for front-back discrimination of the source direction. A recent study analyzed the discrepancy between ITD and ILD (Brungart and Rabinowitz, 1999); results showed that ILD varies with the source position differently than ITD for a nearby point source. Therefore, it is suggested that an investigation for ILD as well as ITD is very important. For that reason, a method based on ILD estimation is expected to be useful for fitting the localization cues to each listener with a low burden in many applications such as auditory displays (Begault, 1994; Takane *et al.*, 2003).

In this study, estimation of ILD based on a listener's anthropometry is investigated. ILD is an important localization cue when sound sources are on the horizontal plane. Therefore, the target of this study is the effect of individualization of ILD on azimuthal localization. In fact, ILD is dependent on frequency. Therefore, it is calculated as the difference in each 1/3-octave band level between the right and left HRTFs. The azimuthal variation of ILDs is then modeled by superposition of sinusoidal functions. The relationship between ILDs and the anthropometric parameters is analyzed using multiple regression analysis. The azimuthal variation of ILDs in each 1/3-octave band is then estimated from the listener's anthropometric parameters. The effect of estimated ILDs is evaluated by a psychoacoustical experiment.

II. DATA USED FOR ANALYSES

A. HRTF data

Measured HRTFs and anthropometric parameters for many listeners are necessary to analyze the relationship between ILD and anthropometry. An estimation equation of ILD is introduced based on the results of analysis. In this paper, ILDs are calculated from measured HRTFs in an HRTF corpus referred from an earlier study (Watanabe *et al.*, 2005a). Because the HRTFs included in this corpus were measured in a blocked-ear-canal condition (Møller, 1992), those data were not affected by the acoustical characteristics of the ear canal. Therefore, the data are suitable for estimation based on anthropometry of the outer profile of a listener.

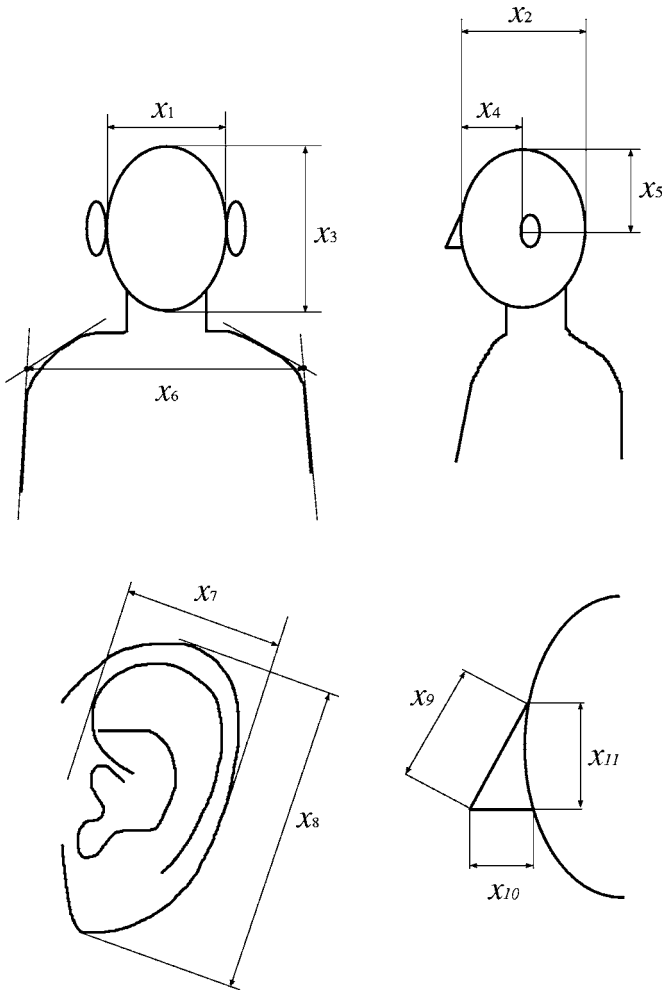


FIG. 1. Anthropometric parameters used to estimate ILD. These values were measured for 59 listeners when HRTFs were measured.

B. Anthropometric parameters

Because ILDs vary as a function of source direction, the auditory system can use them as azimuthal localization cues (Blauert, 1983). The ILDs arise because of acoustic interference of the head and ear such as reflection and diffraction. Because such interference depends on frequency, ILDs can be represented as a function of frequency, as well as source direction. Therefore, for estimation, not only the head size but also the size of the pinna, ear position, and so on are selected as anthropometric parameters. Figure 1 shows the 11 parameters used for estimation in our proposed method.

The HRTF corpus includes HRTFs of 59 listeners, including 43 males, 15 females, and a head and torso simulator (HATS; KOKEN, SAMRAI), and their anthropometric parameters. The human listeners ranged in age from 20 to 30 years (Watanabe *et al.*, 2005a).

III. FREQUENCY CHARACTERISTICS OF ILD AND MODEL OF DIRECTIONAL DEPENDENCE

A. Frequency characteristics of ILD

Actually, ILD is a function of frequency defined as

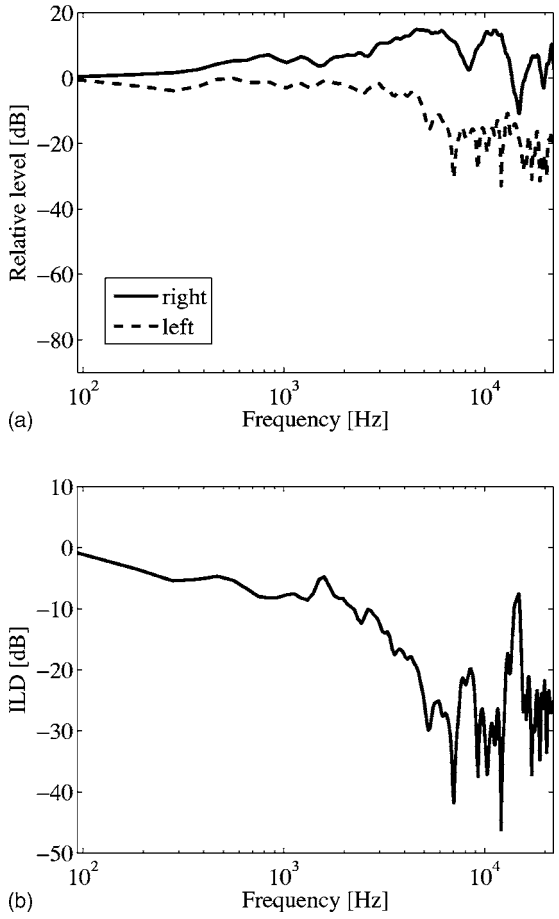


FIG. 2. Frequency characteristics of HRTF and ILD. The abscissa is frequency, and the ordinate is the magnitude responses of (a) HRTFs and (b) ILD. The upper panel shows the magnitude responses of right-ear and left-ear HRTFs. The lower panel shows ILD calculated as the difference between right and left HRTFs.

$$\mathrm{ILD} = 20 \log \frac{|H_L(f)|}{|H_R(f)|}, \tag{1}$$

where f is the frequency, and $H_R(f)$ and $H_L(f)$ respectively denote the right and left HRTFs. Figure 2 shows magnitude responses of HRTFs of the right and left ears and ILD at the source direction of 90°. The ILD varies simply with frequency in the low-frequency region and shows complex characteristics in the high-frequency region. However, it is not necessary to estimate the complexity in the high frequency region because the resolution of the auditory system is restricted in this region. In this study, ILD is calculated as the level difference in each 1/3-octave band between the right and left HRTFs (Fig. 3).

B. Model of ILD

Figure 4 shows the relationship between the source direction and ILD in 1/3-octave bands for three representative bands. Solid lines are the measured ILDs of a listener, dashed lines are the modeled ILDs described in this section, and dotted lines are the estimated ILDs described in Sec. IV. In a higher frequency band, the range of the ILD value is larger and the azimuthal variation is complex. However, the

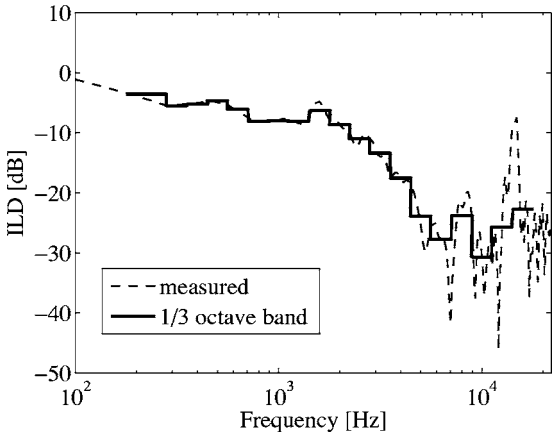


FIG. 3. Frequency characteristics of ILD. The dashed line is the measured ILD that is taken from Fig. 2(b). The solid line is a 1/3-octave band ILD defined as the difference in each 1/3-octave band level between the right and left HRTFs.

overall tendency of the variation seems sinusoidal. In this study, therefore, the variation of ILD in each 1/3-octave band is modeled separately as

$$\hat{y}_m = \sum_{k=1}^{D} C_k \sin \frac{2\pi k m}{M}, \tag{2}$$

where m is the suffix corresponding to a source direction at $360(m-1)/M$ (deg), $\hat{y}_m$ is an ILD of the source direction, M is the number of source directions, C_k is the magnitude of the kth sine function, and D is the order of the model. In Fig. 4, ILDs modeled by Eq. (2) with the order of 10 are shown as dashed lines.

The error of the model is then evaluated to determine the order of the model. The error of the model E_{model} is defined as

$$E_{\mathrm{model}} = \frac{1}{M} \sqrt{\sum_{m=1}^{M} (y_m - \hat{y}_m)^2}, \tag{3}$$

where y_m is a measured ILD, $\hat{y}_m$ is the ILD modeled by Eq. (2), and M is the number of source directions. In this investigation, $M=72$. Figure 5 shows the relationship between the order of the model and the error for the bands with center frequencies of 500, 2000, and 5000 Hz. The variation of ILD is sinusoidal in Fig. 4(a). Therefore, the error is almost constant if the orders of the model are larger than three in Fig. 5. Although the variations are complex in Figs. 4(a) and 4(c), the errors are almost constant if the orders are 10 or higher. Therefore, in this study, ILDs are modeled with the order $D=10$. Figure 6 shows the relationship between the center frequency and the average error calculated for ILDs derived from all HRTFs in the corpus. In Fig. 6, the average errors are less than 3 dB up to 10 kHz. As shown in Fig. 4, the outlines of the modeled and measured ILDs are similar. Therefore, it is expected that the directional dependencies of ILDs are modeled well.

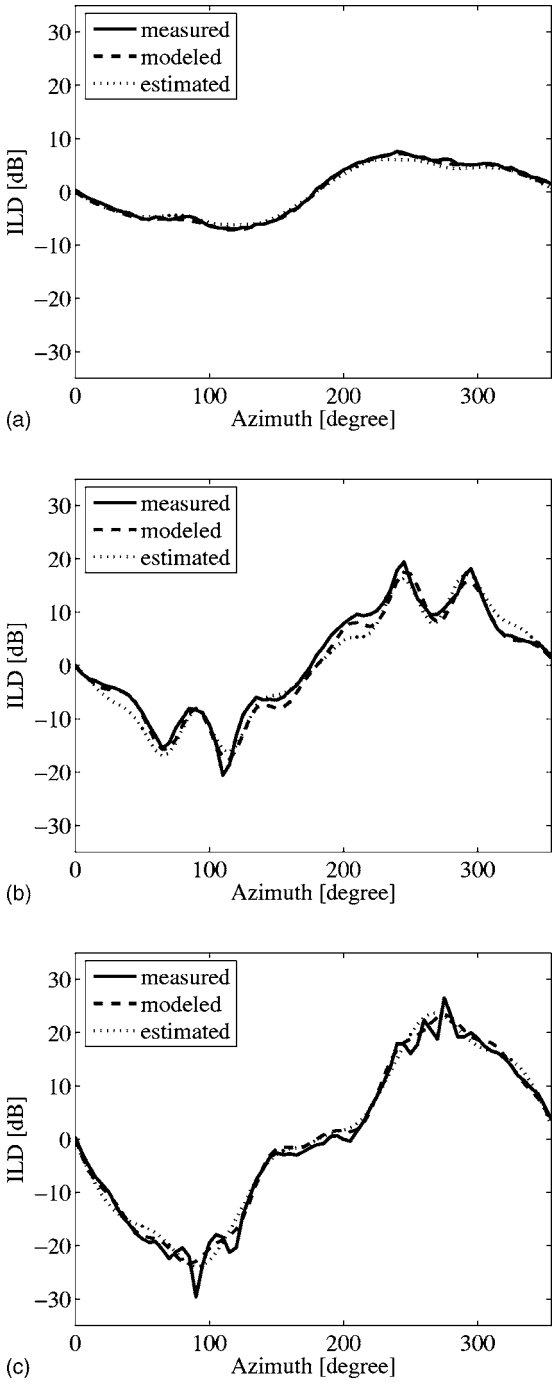


FIG. 4. Relations between source directions and ILD. The solid line is the measured ILD. The dashed line is the modeled ILD calculated using Eq. (2). The dotted line is the estimated ILD by the method proposed in this paper. The abscissa is the source direction, which is 0° in the front, 90° on the right, and 270° on the left. The ordinate is the ILD of a 1/3-octave band. Three panels respectively show examples of bands with center frequencies of 500, 2000, and 5000 Hz.

IV. ESTIMATION OF ILD FROM LISTENER'S ANTHROPOMETRY

Because the parameter of the model C_k reflects the individuality of ILD, an ILD fitted to each listener is obtainable by estimating C_k. Therefore, we decided to estimate C_k from the listener's anthropometry. An ILD fitted to each listener is then obtained by Eq. (2).

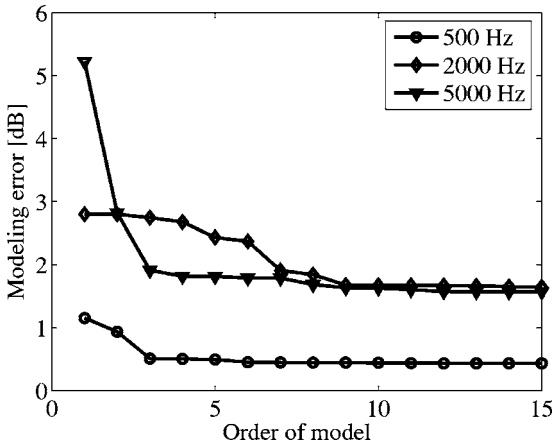


FIG. 5. Relationships between orders of the model and error for the bands with center frequencies of 500, 2000, and 5000 Hz. Each error is calculated using Eq. (3) as the difference between each value of measured and modeled ILD, and averaged in source directions.

In this study, the relationship between C_k and anthropometric parameters is analyzed using multiple regression analysis. The relationship is defined as

$$C_k(f_c) = a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{k11}x_{11} + b_k, \quad (4)$$

where $C_k(f_c)$ is the magnitude of the kth sine function in the 1/3-octave band of the center frequency of f_c, $a_{ki}(i = 1, 2, \ldots, 11)$ is a partial regression coefficient, x_i is the anthropometric parameter shown in Fig. 1, and b_k is a constant term. The estimation of $C_k(f_c)$ is applied for each 1/3-octave band.

The ILDs and the anthropometric parameters from 54 of the 59 listeners in the corpus were used for multiple regression analysis. The ILDs of four of the other five listeners were estimated using Eq. (4), and they participated in a psychoacoustical experiment. Three examples of estimated ILDs are shown in Figs. 4(a)–4(c) as dotted lines. As shown in Fig. 4, each curve is suitable for the outlines of the measured ILDs. Estimation error E_e is defined as

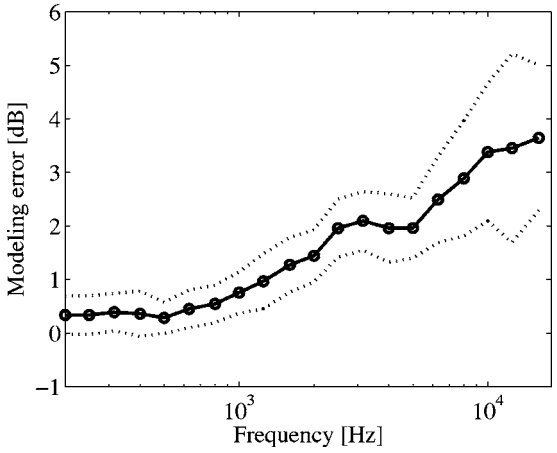


FIG. 6. Relationship between center frequency and average error calculated for ILDs derived from all HRTFs in the corpus. Those ILDs were modeled with the order of 10 for each 1/3-octave band. The solid line shows the average error calculated by all HRTFs in the corpus. The dotted lines show average ±1 s.d. Circles denote the center frequencies of 1/3-octave bands.

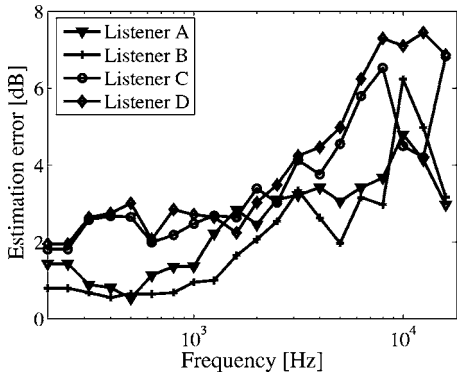


FIG. 7. Estimation errors of ILD calculated as the difference between measured and estimated ILDs. Those errors are calculated for each of the four listeners who participated in the psychoacoustical experiment.

$$E_e = \frac{1}{M}\sqrt{\sum_{m=1}^{M}(y_m - \tilde{y}_m)^2}, \tag{5}$$

where $\tilde{y}_m$ is the estimated ILD. Figure 7 shows the estimation error calculated from each band for the four listeners who participated in the psychoacoustical experiment. As shown in Fig. 7, the overall accuracies are nearly equal among listeners. A psychoacoustical experiment in Sec. V will prove whether the estimation was performed adequately.

V. PSYCHOACOUSTICAL EXPERIMENT FOR EVALUATION OF ESTIMATED ILD

A. Method

A psychoacoustical experiment was performed to evaluate the effectiveness of estimated ILD on localization. A stimulus convolved with an HRTF was presented to the listeners, who were asked to state the perceived direction in the horizontal plane. The conditions of test stimuli were as follows:

(1) Listener's own HRTFs for the right and left ears (own).
(2) HRTFs of someone other than the listener (other).
(3) HRTFs of someone other than the listener, but their ILDs were replaced with estimated ones (ILD_fit).
(4) The same as "ILD_fit," but ITDs were also replaced with estimated ones (ILD+ITD_fit).

The HRTFs in condition "other" were selected from the corpus so that the differences in ILD from each listener's own were the greatest. Difference L was denned as

$$L = \sum_{m=1}^{M}\sum_{n=1}^{N}(y_{mn} - y'_{mn})^2, \tag{6}$$

where y_{mn} is the listener's own ILD in the nth 1/3-octave band at the mth source direction, y'_{mn} is an ILD which is not the listener's own, N is the number of bands (N=20), and M is the number of source directions. The HRTFs in condition "ILD_fit" are obtained using the following process:

(1) Each 1/3-octave band ILD at a certain direction was estimated using the proposed method.
(2) The 1/3-octave band ILD was linearly interpolated to increase sample points to the same points as those of the original HRTF because the ILD values were added to the value of the magnitude response of HRTF in each corresponding frequency bin in the next step.
(3) The interpolated ILD was added to the magnitude response of HRTF on the ipsilateral side (the right ear side from 0° to 180° and the left ear side from 180° to 360°) to obtain the magnitude response of HRTF on the contralateral side.
(4) The phase response was synthesized using the minimum phase component.

The ITDs of condition "ILD+ITD_fit" were obtained using the method described in an earlier study (Watanabe *et al.*, 2005b). Watanabe *et al.* (2005b) defined ITD as the difference in initial delays of head-related impulse responses (HRIRs) between the right and left ears, where an HRIR was obtained using inverse Fourier transformation of an HRTF. The estimation of ITDs was achieved using a similar method to that used for the estimation of ILDs, i.e., the relationship between ITDs and the anthropometric parameters analyzed by multiple regression analysis was used. The individualization of the ITDs was implemented by replacing the ITD of "ILD_fit HRTF" using the estimated ITDs.

Four males with normal hearing acuity participated in this test. Each stimulus was pink noise with a duration of 2 s convolved with an HRTF in a sampling frequency of 44.1 kHz. It was presented through headphones (STAX, SR202). Because the intervals of source directions were 15° in the horizontal plane, the number of source directions was 24. Four HRTF conditions (own, other, ILD_fit, and ILD+ITD_fit) were used, and each stimulus was tested six times in random order. Consequently, the total number of trials for each listener was 576: 4 conditions ×24 directions ×6 repetitions.

B. Results

The results for all listeners in each condition are shown in Fig. 8. In condition "own," the localization seems to be accurate. In contrast, several localization errors occurred in condition "other." In condition "ILD_fit," there are improvements on localization accuracy: There are more responses lying along the positive diagonal than in the condition "other." However, the localizations on or near the median plane (azimuths around 0° or 180°) do not seem to be improved as shown in condition "ILD_fit." For example, Listener B perceived a sound image in front as being behind in both conditions "other" and "ILD_fit," i.e., front-back confusion occurred. Figure 9 shows the front-back confusion rate calculated in conditions "other" and "ILD_fit" for two direction groups: One is the source directions on or near median plane (0°, 15°, 165°, 180°, 195°, and 345°), the other is directions except "on/near median plane," which are abbreviated as "other directions" hereafter. In this analysis, the results for the lateral source directions of 75°, 90°,105°, 255°, 270°, and 285° are excluded because the errors around those directions can be considered as azimuthal errors, not front-back confusions (Zahorik *et al.*, 2006). In Fig. 9, the

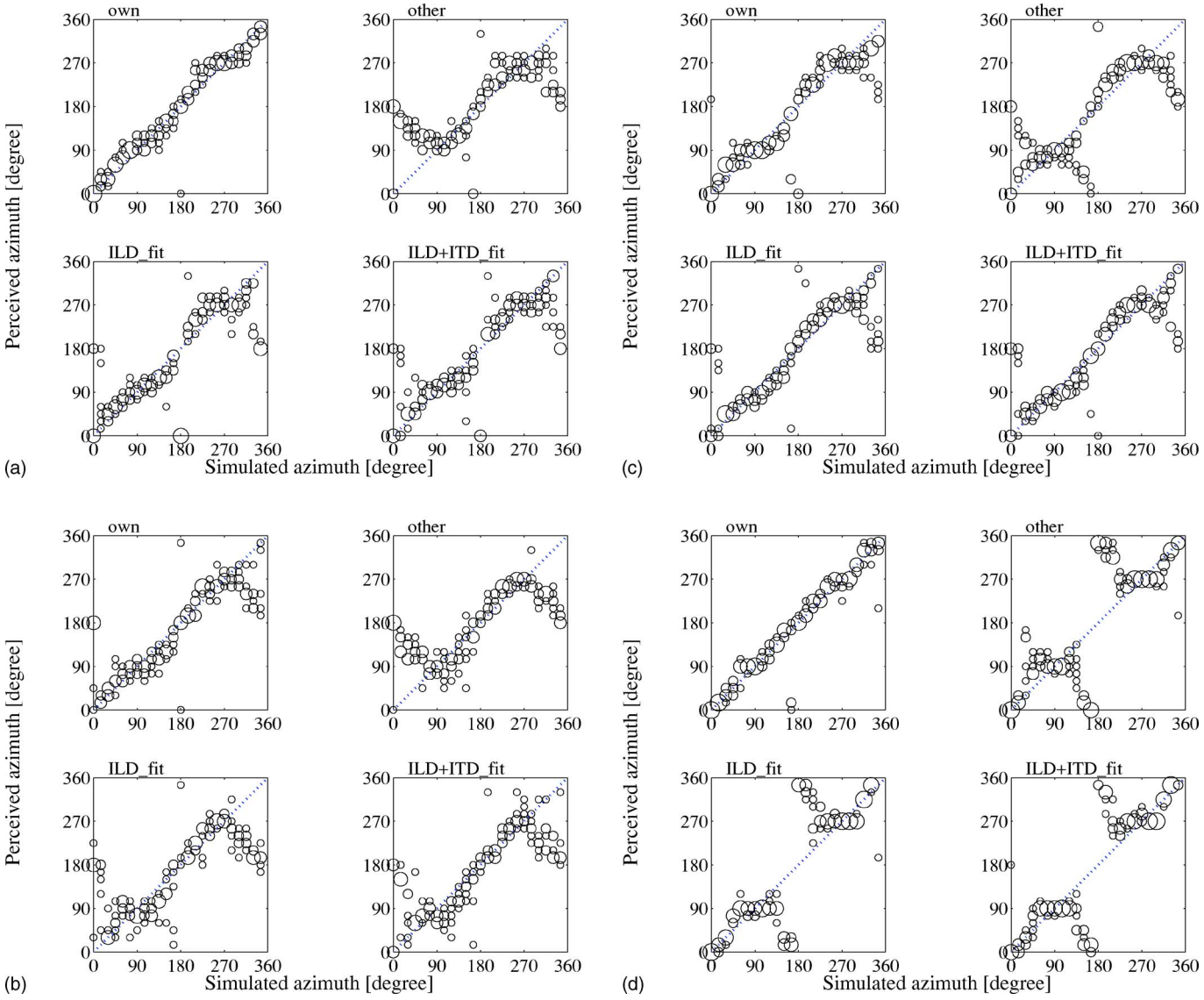


FIG. 8. (Color online) Scatterplots of simulated directions vs a listener's perceived directions. The abscissa is the simulated direction and the ordinate is the listener's perceived direction. A circle is plotted so that its area is proportional to the number of listener's responses. The condition "own" indicates that the listener's own HRTFs are used. The condition "other" indicates that the HRTFs of someone other than the listener are used. The condition "ILD_fit" indicates that the same HRTFs in the condition "other" are used except that their ILDs are replaced with estimated ones. The condition "ILD+ITD_fit" indicates the same HRTFs in condition "ILD_fit" are used except that ITDs are replaced with estimated ones.

front-back error rates on or near the median plane do not differ greatly between the two conditions, whereas the error rates for other directions in condition "ILD_fit" exhibit generally lower rates than those in condition "other." The effects of conditions of ILD ("other" or "ILD_fit"), source direction groups, and listeners were analyzed using three-way analysis of variance. In this analysis, inverse sine transformation was applied to each value of front-back confusion rate because "rates" are analyzed (Rao, 1973; Snedecor and Cochran, 1989). The interaction between conditions and source direction groups was statistically significant $[\chi^2(2)=11.9, p<0.01]$. Therefore, the simple main effects of conditions of ILD for each direction group were analyzed. Results show that the simple main effect was not significant when the source directions were on or near median plane, although the effect was significant when the source directions were other directions $[\chi^2(1)=16.2, p<0.01]$, which means that the effect of the proposed method is dependent on source directions. In addition, improvement of front-back confusion was found in directions, except on or near the median plane, which is quite reasonable because ILDs are almost zero at azimuths around 0° or 180°,resulting in the small effect of fitting the ILD profile to the listener's own.

The tendencies of improvement differ among listeners. For Listeners A and B, improvement is observed at frontal directions except for those near the median plane. As for Listener C, improvement is apparent at frontal and rear directions except those near the median plane. Listener D exhibits a different tendency from the other three. That is, the improvement is found only at 30°–60° and frequent front-back errors occurred for directions of 105°–255°. In other words, this listener tended to perceive a sound image presented in the rear direction as being at frontal one.

The difference between the results in conditions "ILD

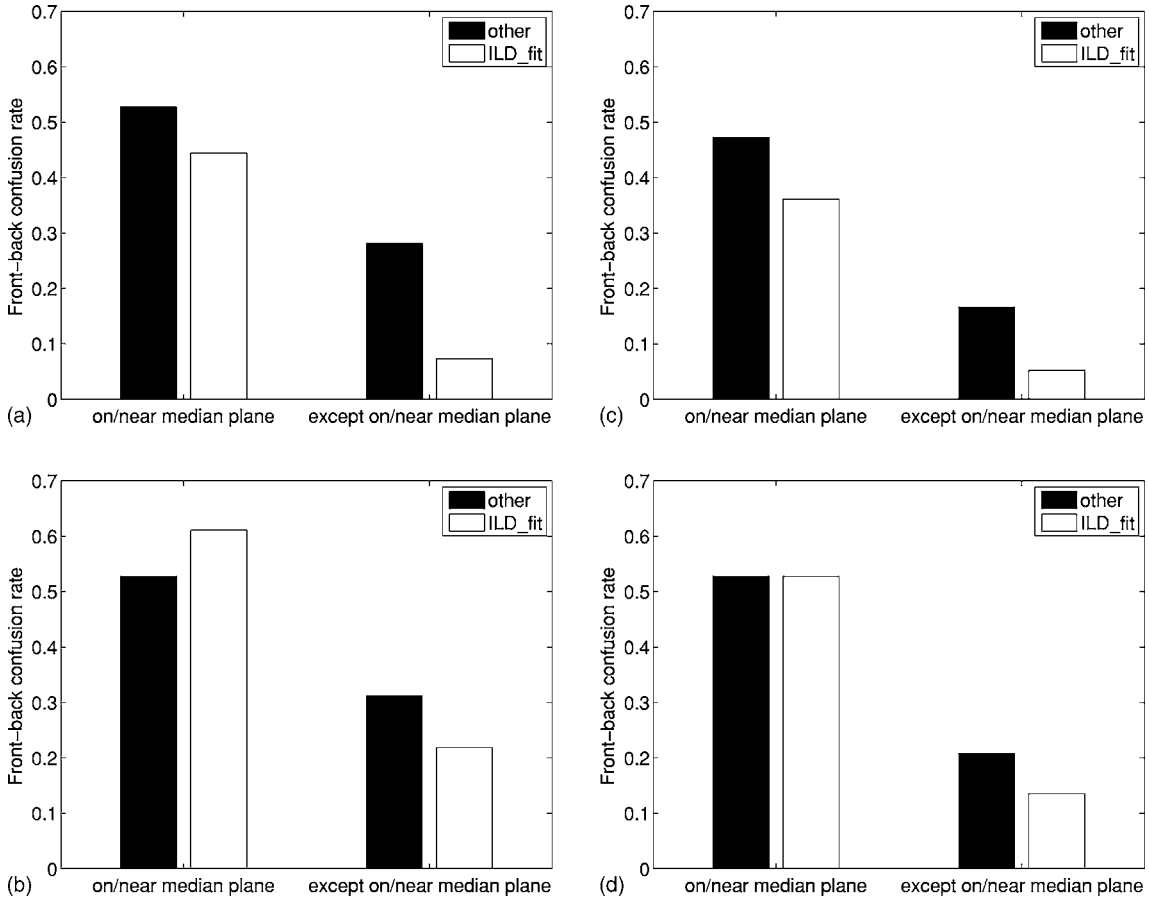


FIG. 9. Front-back confusion rate for each listener in two source direction groups: One is the source direction on or near the median plane, the other is directions except "on/near median plane." Black bars represent rates calculated in condition "other." White bars denote rates calculated in condition "ILD_fit."

_fit" and "ILD+ITD_fit" is small. In other words, the effects of fitting ITD were small because the ITDs of the HRTFs in condition "other" were similar to those of the listeners' own since those HRTFs were selected so that the distance of ILDs, not ITDs, from a listener's own was largest. The difference in ITDs between the HRTFs used in conditions "own" and "other" were calculated. The average value was 0.035 ms, whereas the difference in the ITDs used in Watanabe *et al.* (2005b) was larger than 0.1 ms. Therefore, the ITDs of the HRTFs used in this experiment were near the listener's own and the effects of ITD correction were small.

VI. DISCUSSION

A. Improvement of localization

The degradation of localization in condition "other" is a typical result when using an HRTF which is not the listener's own. Because the front-back confusion is frequently observed, it is considered that listeners cannot use spectral cues (Hebrank and Wright, 1974; Morimoto *et al.*, 2005) included in HRTFs. As described previously, the HRTFs in condition "other" were selected so that the distance of ILDs from a listener's own was the largest. However, the difference in HRTFs was unclear. Therefore, the difference between HRTFs used in conditions "own" and "other" is evaluated for confirmation. Spectral distortion (SD) to evaluate the difference was calculated as

$$\mathrm{SD} = \sqrt{\frac{1}{W}\sum_{w=1}^{W}\left(20\log\frac{|H'(f_w,\theta)|}{|H(f_w,\theta)|}\right)^2}, \tag{7}$$

where $|H(f_m,\theta)|$ and $|H'(f_w,\theta)|$, respectively, denote magnitude responses of the listener's own and not own HRTFs, and W is the sample length of HRTF, f_w is the frequency, and θ is the source direction. Equation (7) indicates that the larger SD is, the larger the difference between the two HRTFs. Table I shows the SD between each listener's own HRTFs and the HRTFs used in condition "other" or HRTFs in the corpus (Watanabe *et al.*, 2005a) averaged for all source directions. As shown in Table I, the difference of HRTFs used in condition "other" was almost equal to that of HRTFs in the corpus. Therefore, the difference between HRTFs in conditions "own" and "other" was large to some extent in terms of the spectral distance of HRTFs. As shown by the results of

TABLE I. Difference between listener's own and not own HRTFs (dB).

Listener	Condition "other"	HRTF corpus (Watanabe *et al.*, 2005a) (average)
A	6.4	6.2
B	6.5	5.7
C	8.1	5.9
D	6.6	6.1

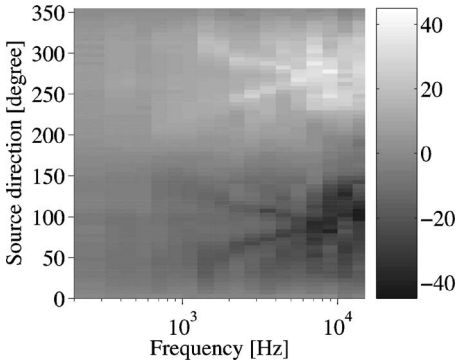


FIG. 10. Directional dependence of frequency characteristics of ILD calculated from Listener A's HRTFs. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the value of each 1/3-octave band ILD in decibels.

experiment and the SD value, listeners can only inadequately use the spectral cues included in HRTFs. On the other hand, in this experiment, the spectrum of HRTFs at the ipsilateral side used in condition "ILD_fit" was the same as that used in condition "other"; furthermore, improvements for localization were observed by fitting ILDs to the listener. Consequently, it is suggested that the front-back localization cue is included in the frequency characteristics of ILD.

B. Analysis of frequency characteristics of ILD

Figure 10 shows the frequency characteristics of ILDs obtained from Listener A's HRTFs. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the value of each 1/3-octave band ILD in decibels. In Fig. 10, although ILDs in the low-frequency region are almost constant, the features of ILDs such as peaks or dips vary as a function of source directions for the frequency region higher than 1 kHz. In regions of 5 kHz or higher, the overall tendency of the variation seems sinusoidal. In the region of about 1–5 kHz, there seem to be peak-like or dip-like bands, which are about 10 dB larger or smaller than neighboring frequency bands. The frequency of the peak or dip shifts according to the source direction; it becomes highest at 90° or 270°, which are positions immediately lateral to the head. In addition, anterior-posterior asymmetry is also observed. Carlile and Pralong (1994) reported the frequency-dependent variation of ILD, as well as monaural HRTF. Peak-like or dip-like features were observed, although its frequency band was broadened by processing of the auditory filter model. These support our results. The magnitude spectra of HRTFs are analyzed to clarify the causes of the peaks and dips of ILDs. Figure 11 shows the magnitude spectra of Listener A's HRTFs. Figures 11(a) and 11(b) respectively show the HRTFs at the ipsilateral and the contralateral sides. In Fig. 11(a), two dips are apparent at around 6–8 kHz and at 10 kHz. In fact, Morimoto *et al.* (2005) reported that these are localization cues for elevation. On the other hand, the HRTFs at the contralateral side show peaks and dips, which vary in a way that is similar to that observedin the frequency characteristics of ILDs. Therefore, the peaks and dips in ILD are derived from the HRTF at the contralateral side. Those features have also been observed in other studies (Mehrgardt

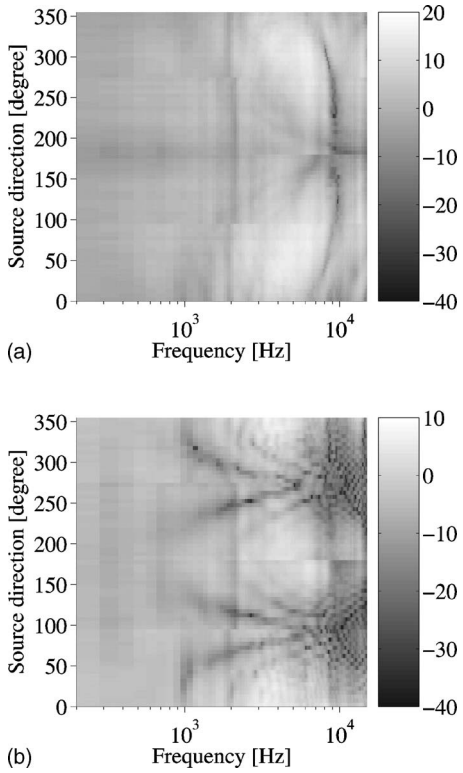


FIG. 11. Magnitude spectrum of the HRTFs used in condition "own" calculated from Listener A's HRTFs. (a) The HRTFs at ipsilateral side and (b) the HRTFs at the contralateral side. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the value of HRTFs in decibels.

and Mellert, 1977; Middlebrooks, 1999; Shaw, 1974; Wightman and Kistler, 1989a, b; Zahorik *et al.*, 2006). Figure 12 shows a spherical model of a head to explain, qualitatively, the above-described characteristics of ILDs. Half of the average head width of the listeners who participated in the experiment was used as the radius *r*. In Fig. 12, both ears are shown to be positioned slightly rearward from the point immediately lateral to the head. Two paths from the source to the ear at the contralateral side are roughly examined: One is the path through the front of the head (path 1); the other is the path through the back of the head (path 2). In this spherical model, the dips in frequency characteristics are assumed to be derived from superposition of two waves in the opposite phase. For that reason, the phase differences are calculated for each frequency, which corresponds to the time dif-

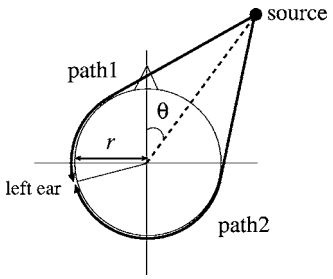


FIG. 12. A spherical model of a head arranged to explain the frequency characteristics of ILDs qualitatively. Half of the average head width of the listeners who participated in the experiment was used as the radius *r*. Two paths from the source to the ear at the contralateral side are also shown.

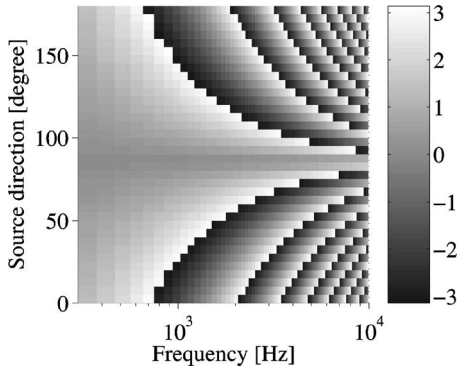


FIG. 13. Phase difference of two sound waves observed at the contralateral side. Each value is calculated using the spherical head model in Fig. 12. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the phase difference in radian.

ference calculated from the difference of two paths shown in Fig. 12. Figure 13 shows the phase difference based on the model in Fig. 12. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the phase difference in radian. In Fig. 13, the position of the phase difference of $-\pi$, which is the opposite phase, shifts into the higher frequency region as the source direction varies from 0° to 90°, whereas that position shifts into the lower frequency region from 90° to 180°. This tendency corresponds to the movement of the dips ofHRTFs at the contralateral side in Fig. 11(b).

Based on the frequency characteristics of the above-described ILD, we attempt to explain individual differences in improvement of localization in this experiment. The difference between the ILD used in conditions "own" and "other" or "ILD_fit" were calculated. For example, the results of Listeners A and D, which were greatly improved and not improved on localization, respectively, are shown in Fig. 14. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the difference of 1/3-octave band ILDs in decibels. In Fig. 14, the differences of ILDs for Listener A between conditions "other" and "own" are large in the region around 1–5 kHz, whereas the differences between "ILD_fit" and "own" in that region are small. The results for Listener D show that the differences of ILDs between conditions "ILD_fit" and "own" are large in the region around 1–5 kHz, although the overall difference is smaller than the difference of ILDs between condition "other" and "own." These findings suggest that the fitting of ILD in the 1- to 5-kHz region contributed to the improvement of localization. From those results, the estimation errors shown in Fig. 5 are sufficiently small at least for Listener A, although they are insufficiently small for Listener D.

On the other hand, the overall difference of the left side (180°–360°) in Figs. 14(a) and 14(b) seems to be symmetrical, which suggests that the improvements in localization errors at the left side are small compared to those at the right side. Such right-left asymmetry should be applied to the model of ILD in the future.

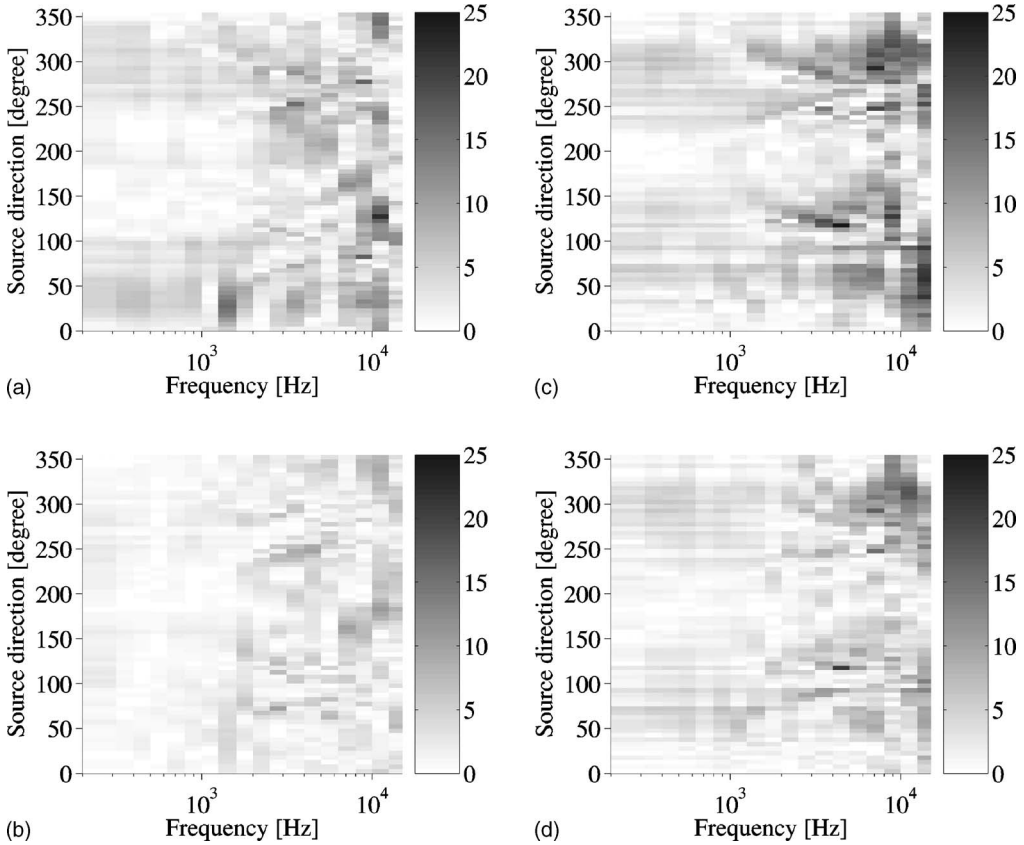


FIG. 14. Difference of ILDs from own ILDs. The abscissa is frequency, the ordinate is the source direction, and the brightness denotes the difference of 1/3-octave band ILDs in decibels. The data are calculated as the difference between the ILDs derived from the HRTFs used in conditions "own" and "other" [(a) Listener A and (c) Listener D], and the difference between those used in conditions "own" and "ILD_fit" [(b) Listener A and (d) Listener D].

VII. CONCLUSIONS

In this study, estimation of ILD based on a listener's anthropometry was investigated. The ILDs were calculated in each 1/3-octave band to take account of its frequency characteristics. The results of the psychoacoustical experiment suggest that fitting ILDs for each listener improves localization when monaural spectral cues are degraded. Analysis of the frequency characteristics indicates that ILDs in the frequency range of 1–5 kHz are important and that the HRTFs in the contralateral side to a sound source affect the frequency characteristics of ILDs in that range. In addition, a simple spherical head model can explain those features of ILD. Therefore, we conclude that the frequency characteristics of ILDs are important for localization. When the fitted ILDs agree with the listener's own, great improvement of localization error is obtainable, even if the monaural HRTFs are not those of the listener. This means that sound localization on the horizontal plane with nonindividualized HRTF can be improved if ILD is fitted to the listener.

Given these findings, further investigation of the spectral cues of ILDs or monaural and interaural spectral cues should be conducted.

ACKNOWLEDGMENTS

The authors wish to thank Professor Masayuki Morimoto and Dr. Kazuhiro Iida for useful comments on ILD spectral cues. The authors also wish to thank Dr. Nobutaka Ono for useful comments related to analysis of the frequency characteristics of ILD. A part of this study was carried out under the Cooperative Research Project of the Research Institute Electrical Communication, Tohoku University (H18/A10).

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2002**). "The use of head-and-torso models for improved spatial sound synthesis," in *Proceedings of the 113th AES Convention*, Los Angeles, CA.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in Proceedings of the 2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 21–24.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Begault, D. R. (**1994**). *3-D Sound for Virtual Reality and Multimedia* (AP professional, Cambridge, CA).

Blauert, J. (**1983**). *Spatial Hearing* (MIT, Cambridge, MA).

Bronkhorst, A. W. (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98**, 2542–2553.

Brungart, D. S., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. I. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

Carlile, S., and Pralong, D. (**1994**). "The location-dependent nature of perceptually salient features of the human head-related transfer functions," J. Acoust. Soc. Am. **95**, 3445–3459.

Genuit, K. (**1984**). "A model for the description of outer-ear transmission characteristics," [translation of "Ein Modell zur Beschreibung von Au βenohrübertragungseigenschaften"], Ph.D. thesis, Rhenish Westphalian Technical University, Aachen.

Haneda, Y., Makino, S., Kaneda, Y., and Kitawaki, N. (**1999**). "Common-acoustical-pole and zero modeling of head-related transfer functions," IEEE Trans. Speech Audio Process. **7**, 188–196.

Hebrank, J., and Wright, D. (**1974**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.

Inoue, N., Kimura, T., Nishino, T., Itou, K., and Takeda, K. (**2005**). "Evaluation of HRTFs estimated using physical features," Acoust. Sci. & Tech. **26**, 453–455.

Kawaura, J., Suzuki, Y., Asano, F., and Sone, T. (**1991**). "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," J. Acoust. Soc. Jpn. (E) **12**, 203–216.

Kistler, D. J., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.

Makous, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188–2200.

Martens, W. L. (**2003**). "Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis," Acoust. Sci. & Tech. **24**, 220–232.

Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

Møller, H. (**1992**). "Fundamentals of binaural technology," Appl. Acoust. **36**, 171–218.

Morimoto, M., (**1999**). "Role of two ears in upper hemisphere localization," J. Acoust. Soc. Am. **105**, 1163.

Morimoto, M., and Nomachi, K. (**1982**). "Binaural disparity cues in median-plane localization," J. Acoust. Soc. Jpn. (E) **3**, 99–103.

Morimoto, M., Yairi, M., Iida, K., and Itoh, M. (**2005**). "The role of low frequency components in median plane localization," Acoust. Sci. & Tech. **24**, 76–82.

Phillips, D. P., and Hall, S. E. (**2005**). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," Hear. Res. **202**, 188–199.

Rao, C. R. (**1973**). *Linear Statistical Inference and Its Applications* (Wiley, New York).

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.

Shaw, E. A. G., and Teranishi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby sound source," J. Acoust. Soc. Am. **44**, 240–249.

Snedecor, G. W., and Cochran, W. G. (**1989**). *Statistical Method* (Iowa State University Press, Ames, IA).

Takane, S., Suzuki, Y., Miyajima, T., and Sone, T. (**2003**). "A new theory for high definition virtual acoustic display named ADVISE," Acoust. Sci. & Tech. **24**, 276–283.

Watanabe, K., Iwaya, Y., Gyoba, J., and Suzuki, Y. (**2005a**). "Head-related transfer function corpus for virtual auditory environment," Information Technology Letters **4**, 237–240 (in Japanese).

Watanabe, K., Iwaya, Y., Gyoba, J., Suzuki, Y., and Takane, S. (**2005b**). "An investigation on the estimation of interaural time difference based on anthropometric parameters," Trans. the Virtual Reality Soc. of Jpn. **10**, 609–617 (in Japanese).

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F.-L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening. II. Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C. (**2006**). "Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," J. Acoust. Soc. Am. **120**, 343–359.

Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration

Paul Iverson[a] and Bronwen G. Evans
Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 5 December 2006; revised 20 August 2007; accepted 20 August 2007)

This study examined whether individuals with a wide range of first-language vowel systems (Spanish, French, German, and Norwegian) differ fundamentally in the cues that they use when they learn the English vowel system (e.g., formant movement and duration). All subjects: (1) identified natural English vowels in quiet; (2) identified English vowels in noise that had been signal processed to flatten formant movement or equate duration; (3) perceptually mapped best exemplars for first- and second-language synthetic vowels in a five-dimensional vowel space that included formant movement and duration; and (4) rated how natural English vowels assimilated into their L1 vowel categories. The results demonstrated that individuals with larger and more complex first-language vowel systems (German and Norwegian) were more accurate at recognizing English vowels than were individuals with smaller first-language systems (Spanish and French). However, there were no fundamental differences in what these individuals learned. That is, all groups used formant movement and duration to recognize English vowels, and learned new aspects of the English vowel system rather than simply assimilating vowels into existing first-language categories. The results suggest that there is a surprising degree of uniformity in the ways that individuals with different language backgrounds perceive second language vowels.
[DOI: 10.1121/1.2783198]

PACS number(s): 43.71.Hw, 43.71.Es [ARB] Pages: 2842–2854

I. INTRODUCTION

It is clear that one's first-language (L1) phonetic categories affect second-language (L2) vowel learning. For example, Spanish listeners have difficulty learning to discern the difference between English /i/ and /ɪ/ (e.g., Escudero and Boersma, 2004; Flege *et al.*, 1997; Morrison, 2002), presumably because they both sound like the same Spanish vowel (/i/). In contrast, German listeners have less difficulty learning English /i/-/ɪ/ (Bohn and Flege, 1990; Flege *et al.*, 1997), presumably because they sound like two different German vowels (/i/ and /ɪ/). These types of L1/L2 interactions have been well established at the level of individual phonetic categories, but there has been little research on whether these interactions have broader implications for how individuals learn entire vowel systems. The present study addresses this issue by comparing how individuals with a range of L1 vowel systems (Spanish, French, German, and Norwegian) learn English vowels.

The task of learning an L2 vowel system may be fundamentally different for individuals whose L1 vowel system is large and complex (e.g., Norwegian) than for individuals whose L1 vowel system is small and simple (e.g., Spanish). One possibility is that individuals could take the cues used in their L1 vowel system and apply them to learning an L2, which could be an advantage to listeners with more complex L1 vowel systems (i.e., those that use more cues). For example, L1 English speakers are better than L1 Spanish speakers at learning vowel length contrasts in Swedish, and this may occur because English vowels vary more systematically in duration than do Spanish vowels (MacAllister *et al.*, 2002). English speakers also use duration when learning French /ɔ/-/o/, even more than do L1 French speakers (Gottfried and Beddor, 1988). Likewise, Japanese has a vowel duration contrast, and individuals appear to apply this to learning English such that they primarily use duration to distinguish English /i/-/ɪ/ (Morrison, 2002) as well as using duration to contrast stressed and unstressed vowels within words (Lee *et al.*, 2006). However, the opposite pattern of results is sometimes found; individuals with no L1 vowel duration contrasts (e.g., Spanish and Catalan) often still use duration to distinguish English /i/-/ɪ/ (Cebrian, 2006; Escudero and Boersma, 2004; Morrison, 2002). A reliance on duration when learning L2 vowels may simply be a strategy that is often used when listeners have difficulty discerning a spectral difference (Bohn, 1995; Bohn and Flege, 1990), regardless of whether their L1 contrasts vowel duration (see Flege *et al.*, 1997).

There has been almost no work on whether the use of formant movement (i.e., both diphthongs and intrinsic formant movement for monophthongs) transfers between one's L1 and L2; Italians are better able to learn formant movement for English /e/ if they begin learning at younger ages (Flege *et al.*, 2003), but it is unknown whether the use of formant movement in one's L1 makes learning formant movement in an L2 easier or harder. Even among English

[a]Corresponding author.

monophthongs, formant movement has been shown to be an important cue for native listeners; recognition accuracy can decline by 13–23 percentage points when formant movement is flattened in synthesized or signal processed speech (e.g., Assmann and Katz, 2005; Hillenbrand and Nearey, 1999; Iverson *et al.*, 2006). The reliance on such acoustic information has important theoretical implications, because it suggests that listeners may have phonetically detailed category representations for vowels (i.e., exemplars; e.g., Goldinger, 1996, 1998; Hawkins and Smith, 2001; Johnson, 1997; Nygaard *et al.*, 1995, Nygaard and Pisoni, 1998; Pisoni, 1997), rather than having more abstract representations based only on the primary acoustic cues.

In addition to the potential effects of L1 cues, the sheer number of vowels in an L1 may have implications for L2 vowel learning. Novice L2 learners are thought to use their existing L1 categories when listening to the L2 (i.e., they assimilate the L2 vowels into L1 categories; Best, 1995; Best *et al.*, 2001; Flege, 1995, 2003). This L1 assimilation strategy could be problematic for individuals with small L1 vowel systems, because it is more likely that there will be cases of multiple L2 vowels assimilating to the same L1 category (e.g., English /i/ and /ɪ/ assimilating to Spanish /i/), making it harder for listeners to discern differences among these L2 vowels. Despite this initial difficulty, the small L1 inventory may make it easier for individuals to learn. That is, Flege (1995) has argued that new categories are easier to learn when they are far away from existing categories, and one could imagine that individuals with smaller vowel systems would have more unused room in the vowel space to learn new categories (although it is not clear that individuals with smaller vowel systems actually have more unoccupied space; see Meunier *et al.*, 2003).

Individuals with larger L1 vowel systems may be more successful in using assimilation (i.e., less chance of multiple L2 vowels assimilating to the same L1 category), but they may have more difficulty learning new categories. If individuals with large vowel systems have less unoccupied space to learn new vowels, they would need to change their existing L1 category representations to better match the L2 vowels, creating merged or compromise categories (Flege, 2003; MacKay *et al.*, 2001). Changing existing categories in this way is thought to be more difficult than learning entirely new categories (Flege, 1995, 2003; Munro *et al.*, 1996). It is thus possible that individuals with larger L1 vowel systems may rely more on L1 assimilation and less on new learning than do individuals with smaller L1 vowel systems.

The present study investigated whether L1 speakers of Spanish, French, German, and Norwegian fundamentally differ in the cues that they use when listening to English vowels. Spanish has five vowels (/i/, /e/, /a/, /o/, and /u/), and duration is not used contrastively (Stockwell and Bowen, 1965; Flege, 1989). The status of formant movement is less clear; Spanish seems to lack true diphthongs (i.e., single phonemes marked by the movement between two vowel positions), but monophthongal vowels can occur consecutively in Spanish and are sometimes considered to be diphthongs (Stockwell and Bowen, 1965; Delattre, 1965). French has eleven oral monophthongal vowels (/i/, /y/, /e/, /ø/, /ɛ/, /œ/, /a/, /ɑ/, /ɔ/, /o/, and /u/), and four nasal vowels (/ɑ̃/, /ɔ̃/, /ɛ̃/, /œ̃/; the present study focuses only on the oral vowels). French has no diphthongs and duration contributes negligibly to vowel distinctions (Delattre, 1965). German has 15 monophthongal vowels that form seven long-short (tense-lax) vowel pairs (/i/-/ɪ/, /e/-/ɛ/, /u/-/ʊ/, /o/-/ɔ/, /y/-/ʏ/, /ø/-/œ/, and /a/-/aː/, plus /ɛː/), and three diphthongal vowels (/aɪ/, /aʊ/ and /ɔɪ/; Delattre, 1965; Strange *et al.*, 2005). Norwegian has 18 monophthongal vowels that form nine long-short (tense-lax) pairs (/iː/-/i/, /yː/-/y/, /eː/-/ɛ/, /øː/-/œ/, /æː/-/æ/, /ɑː/-/ɑ/, /oː/-/ɔ/, /uː/-/u/, and /ʉː/-/ʉ/), and four diphthongs (/ɑi/, /ɛi/, /æʉ/ and /øy/; see Kristoffersen, 2000). These L1s thus vary both in terms of the number of vowels and in terms of the cues that are used.

We gave subjects a battery of tests to evaluate their vowel recognition and perceptual category representations. The baseline recognition ability of subjects was evaluated by having them identify natural recordings of English /b/-V-/t/ words. Two tests evaluated the subjects' representation of target (static) formant frequencies, formant movement, and duration. Listeners identified signal-processed natural vowels in noise, in order to examine whether flattening formant movement or equating duration affected recognition for L2 English speakers (see Iverson *et al.*, 2006). Listeners also mapped their perceptual vowel spaces (best exemplar locations) in both their L1 and L2, within a five-dimensional acoustic space that included F1 and F2 at the onset and offset of each vowel, and duration (see Iverson and Evans, 2003; Iverson *et al.*, 2006). Finally, listeners rated how natural English vowels assimilated into L1 categories. Our aims were to: (1) examine to what extent the representation (measured by best exemplars) of target formant frequencies, formant movement, and duration are able to predict the ability of L2 learners to recognize natural English vowels; (2) whether the representation of these cues varies between language groups (e.g., whether individuals more accurately represent L2 duration when their L1 contrasts duration); and (3) whether these representations are due to category assimilation (i.e., using an existing L1 category when listening to the L2) or new learning.

Our L1 subject groups were selected to be relatively heterogeneous (e.g., in terms of English experience) in order to increase individual variation in vowel recognition accuracy. For example, we examined which aspects of English vowel categorization (e.g., representation of duration) correlated with individual differences in English vowel identification accuracy within each L1 group, rather than focusing solely on between-group differences. Although it is more common in the literature to closely control groups for experience, experience is only one of the many factors that can determine whether an individual is good or poor at recognizing L2 vowels (e.g., motivation, aptitude, and type of experience are also important). Our research strategy was to take advantage of this individual variability in order to understand the vowel recognition process, rather than treat it as a confound that should be removed.

II. METHOD

A. Subjects

A total of 114 subjects were tested. Five of these were omitted because of missing data (computer problems or unable to complete). Subjects were additionally screened based on their L1 tests (see Procedure); three were omitted because their L1 signal-to-noise threshold for vowel recognition was more than 2 SD above the average (indicating a possible hearing problem), and six were omitted because their L1 best exemplars were more than 2 SD different from the average (indicating an inability to reliably perform the task). After screening, the numbers of subjects within each L1 group were 25 for Spanish, 19 for French, 21 for German, 18 for Norwegian, and 17 for English. All subjects reported that they had no hearing or language impairments.

Spanish speakers were all tested in London. We tested speakers from a range of countries (nine Spain, five Colombia, three Peru, two Chile, two Cuba, and one each from Argentina, Mexico, Uruguay, and Venezuela), but all had a standard Spanish five-vowel system. The subjects were 21–45 years old (median 29 years). Subjects began learning English when they were 4–24 years old (median 13 years). The subjects had 0–4 years of experience living in English-speaking countries (median 1 year).

French speakers were all tested in London, and all subjects grew up in France. The subjects were 21–55 years old (median 29 years). Subjects began learning English when they were 2–14 years old (median 11 years). The subjects had 0–17 years of experience living in English-speaking countries (median 3 years).

German speakers were tested in Potsdam, Germany (11 subjects) and London (ten subjects); all subjects grew up in Germany. We did not test any subjects who had nonstandard German vowel systems (e.g., Bavarian accent). The subjects were 19–64 years old (median 28 years). Subjects began learning English when they were 6–15 years old (median 11 years). The subjects had 0–30 years of experience living in English-speaking countries (median 2 years).

Norwegian speakers were tested in Trondheim, Norway (15 subjects) and London (three subjects); all subjects grew up in Norway. The subjects were 20–35 years old (median 22 years), and subjects began learning English when they were 8–10 years old (median 9 years). The subjects had 0–6 years of experience living in English-speaking countries (median 0 years).

English speakers were tested in the United Kingdom. All were monolingual and grew up in England. They were 18–49 years old (median 28 years). The English speakers completed only a subset of the tasks to provide normative data (English speech in noise and English vowel-space mapping).

B. Stimuli and apparatus

All subjects were tested in quiet rooms, with stimuli played over headphones at a user-controlled comfortable level. PCs (desktops, laptops, and pocket PCs) were used to play the stimuli and collect responses. Stimulus recordings were made in an anechoic chamber, with 44,100 16-bit samples/s, and later down sampled to 11,025 samples/s.

1. Natural /b/-V-/t/ recordings

We recorded a single speaker for each language, and used this same speaker for all tasks (except for English speech in noise). This was designed to facilitate across-task comparisons of the results (i.e., eliminate variability due to talker differences). All speakers were male, and were native speakers of their respective L1. A /b/-V-/t/ context was used (/b/-V-/ta/ for Spanish) because this was phonologically legal in all of our languages. This context created nonwords in several of the languages, so speakers were also given common real words with that vowel to help illustrate what should be said. The English speaker recorded the words *beat* /i/, *bit* /ɪ/, *bet* /ɛ/, *Burt* /ɜ/, *bat* /a/, *Bart* /ɑ/, *bot* /ɒ/, *but* /ʌ/, *bought* /ɔ/, *boot* /u/, *bait* /eɪ/, *bite* /aɪ/, *bout* /aʊ/, and *boat* /əʊ/, in the sentence *Say___again*; English vowels that would create nonwords in the /b/-V-/t/ context (e.g., /ʊ/) were not included in the study. The Spanish speaker recorded the words and nonwords *bita* /i/, *beta* /e/, *bata* /a/, *bota* /o/, and *buta* /u/ in the sentence *Digo la palabra___de nuevo*. The French speaker recorded the words and nonwords *bit* /i/, *but* /y/, *béte* /e/, *beute* /ø/, *bête* /ɛ/, *boeute* /œ/, *batte* /a/, *bâte* /ɑ/, *botte* /ɔ/, *bôte* /o/, and *bout* /u/, in the sentence *Je dis___encore*. The German speaker recorded the words and nonwords *biet* /i/, *büüt* /y/, *bitt* /ɪ/, *bütt* /ʏ/, *beet* /e/, *bööt* /ø/, *bett* /ɛ/, *bäät* /ɛː/, *bött* /œ/, *batt* /a/, *bad* /aː/, *bott* /ɔ/, *boot* /o/, *butt* /ʊ/, *buud* /u/, *beit* /aɪ/, *baut* /aʊ/, and *beut* /ɔɪ/ in the sentence *Sag___nochmal*. The Norwegian speaker recorded the words and nonwords *bit* /iː/, *byt* /yː/, *bitt* /ɪ/, *bytt* /y/, *bet* /eː/, *bøt* /øː/, *bett* /ɛ/, *bøtt* /œ/, *bœtt* /æ/, *bœt* /æː/, *batt* /ɑ/, *bat* /ɑː/, *bått* /ɔ/, *båt* /oː/, *bott* /u/, *butt* /ʉ/, *bot* /uː/, *but* /ʉː/, *bait* /ɑi/, *beit* /ɛi/, *baut* /æʉ/, and *bøyt* /øy/ in the sentence *Det var___jeg sa*. The speakers read each word individually off of a computer screen, in random order. Each word was recorded four times, and was screened for intelligibility. Speakers also read a short passage (Aesop's *The north wind and the sun*, translated into each language).

In order to facilitate cross-language comparisons (particularly for vowel-space mapping), signal processing was used to equate the formant frequencies and F0 of each talker in the carrier sentences and in the paragraph (e.g., to eliminate differences in vowel quadrilaterals related to vocal tract length). The signal processing followed the Praat *Change Gender* command (Boersma and Weenink, 2005), except that the processing stages were applied individually so that the pitch pulse analyses could be hand corrected. Specifically, a new sample rate was imposed on the stimuli to scale the formants (e.g., the sample rate could be changed from 22,050 to 24,255 if one needed to raise the formant frequencies by 10%), and then pitch synchronous overlap and add (PSOLA) was applied to scale the F0 and duration to correct for changes introduced by the sample rate change, as well as to equate F0 between talkers. The formant frequencies were scaled so that the F2 of the /i/ produced by each talker was 2290 Hz (an average male value; Peterson and Barney, 1952). The F2 of /i/ was selected because it is consistently produced across talkers and is easy to measure accurately.

The F0 was scaled so that the median (as measured from the short passage) was 112 Hz (an average male value; Hazan and Markham, 2004). Across talkers, there were only small changes made to the formant frequencies (maximum 8% change) and F0 (maximum 15 Hz change), so the effects of signal processing were thus subtle.

2. Vowel-space mapping

A large set of synthesized stimuli were created to map best exemplars. The stimuli were embedded in the natural carrier sentences, including the /b/ burst and the /t/ stop gap from the natural recordings. The stimuli were created using the cascade branch of a Klatt synthesizer (Klatt and Klatt, 1990). For each language, the synthesis parameters were chosen so that the synthesized vowel approximated the original vowel in the natural carrier sentence in terms of F0 and amplitude contours. All other synthesis parameters were the same for each language. The upper-formant frequencies (F4=3500 Hz, F5=4500 Hz), formant bandwidths (100, 180, 250, 300, and 550 Hz for F1–F5), tilt (0 dB slope), and open quotient (60%) were the same for all stimuli in all languages. The stimuli primarily varied F1, F2, and duration. F3 varied as a function of F2; F3 was fixed at 2500 Hz whenever F2 was less than 2300 Hz, but otherwise F3 was raised so that it was always 200 Hz greater than F2.

The F1 and F2 formant frequencies changed linearly from the beginning to the end of the vowel, and there were no additional consonantal formant transitions. F1 frequency was restricted so that it had a lower limit of 5 equal rectangular bandwidth (ERB) (Glasberg and Moore, 1990) and an upper limit of 15 ERB. F2 frequency was restricted so that it had a lower limit of 10 ERB, was always at least 1 ERB higher than F1, and had an upper limit defined by the equation $F2=25-(F1-5)/2$. The stimuli were synthesized in advance with a 1-ERB spacing of the vowel space, and with seven log-spaced levels of duration (54, 75, 104, 144, 200, 277, and 383 ms), for a total of 109,375 individual stimuli for each language. The ERB and log-duration transforms allowed us to efficiently distribute the stimuli with regard to perception, although the goodness optimization procedure does not require this equal perceptual spacing.

As discussed above, different natural carrier sentences had been signal processed to equate formant frequencies and F0 across the talkers of the different languages (see description above of the natural /b/-V-/t/ recordings), with the goal of making sure that the perceptual vowel space maps would not differ across languages due to physiological differences between the talkers' voices (see Ladefoged and Broadbent, 1957). To test whether this was successful, a pilot experiment was run in which six L1 English speakers chose best exemplars for *beat, Burt*, and *bat* in the carrier sentences for every language (subjects were asked to ignore the fact that the carrier sentences were not always in English). The procedure was the same as for the vowel space mapping procedure described below. Repeated-measures analyses of variance (ANOVA) revealed that there was no significant effect of sentence context for any of the acoustic dimensions (F1 and F2 at the onset and offset, and duration), $p>0.05$. This demonstrates that the acoustic normalization was successful, and that the best exemplars chosen in different languages in this experiment could thus be directly compared.

3. English speech in noise

The stimuli were from a previous study (Iverson *et al.*, 2006), recorded from a different British English talker (female) than used in the rest of this study. The speaker was recorded saying *Say hVd again* with 11 words: *heed* (/i/), *hid* (/ɪ/), *head* (/ɛ/), *had* (/a/), *heard* (/ɜ/), *hud* (/ʌ/), *hod* (/ɒ/), *hard* (/ɑ/), *hoard* (/ɔ/), *hood* (/ʊ/), and *who'd* (/u/). Diphthongs were not included because they become unintelligible when formant movement is removed. Two additional versions of the vowels were created that: (1) removed all formant movement, and (2) equated duration. The changes to the stimuli were made using Praat (Boersma and Weenink, 2005). Formant movement was removed using linear predictive coding (LPC) analysis and resynthesis. Specifically, LPC analyzed the signal from the start of voicing after the /h/ to the start of the /d/ closure; the signal was inverse filtered to produce an LPC residual; a time slice of the LPC analysis was identified that represented the vowel's target formant frequencies (defined as the point where F1 reached a peak); and this single LPC slice was used to filter the entire LPC residual. This process created stimuli that retained the natural F0 of the original stimuli, but had formant frequencies that remained fixed at each vowel's target values. Duration was equated using PSOLA, such that the durations of the /h/, the /d/ closure, and the vowel were set to the mean values for the talker.

The speech-shaped noise conformed to CCITT Rec. G227 and was produced by a Wandel and Goltermann RG-1 noise generator. The signal-to-noise ratio (SNR) was calculated for each individual stimulus, by comparing the RMS amplitude of the stimulus and noise.

C. Procedure

1. English vowel identification in quiet

Subjects heard natural recordings of the English speaker and gave a closed-set identification response (all 14 words as response options). To give their response, they mouse clicked on a button which listed the stimulus word (e.g., *bot*) as well as a common English word that had the same vowel (e.g., *hot*). Prior to starting the experiment, they heard the speaker read *The North Wind and the Sun* passage. They were shown the word response alternatives and were able to ask questions if they were unsure which vowels were indicated. Subjects identified four repetitions of each vowel, for a total of 56 trials in a random order.

2. English vowel identification in noise

The task was the same as for identification in quiet, except that speech-shaped noise was added to the stimuli adaptively to find the SNR level that yielded 50% correct responses (i.e., 1-up/1-down Levitt procedure; Levitt, 1971). The adaptive series were blocked by the three stimulus conditions (natural speech, no formant movement, equated duration), and there were three blocks for each stimulus condition (i.e., a total of nine blocks). The task began with a

+10 dB SNR level. At the start of the experiment (prior to three reversals) the level was reduced by 6 dB after each correct response and increased by 6 dB after each incorrect response. Afterwards the level was changed in 2 dB steps. The procedure was stopped after seven reversals. The SNR was calculated for each condition by averaging the last four reversals of each block (when the step size was 2 dB), and taking the median of the values for the three blocks. The SNR was classified as undefined when there was more than one adaptive series that was aborted due to repeatedly reaching the upper limit (+10 dB SNR) of the search (i.e., indicating poor recognition even with minimal noise).

3. L1 vowel identification in noise

The procedure was identical to English vowel identification in noise, except that the words were from the listener's L1 and only natural stimuli were used. This task served as a hearing screening for subjects; subjects were omitted when their SNR thresholds were more than 2 SD above the average for their L1 group.

4. L1 assimilation

Subjects heard the English /b/-V-/t/ words, and identified which of their own L1 /b/-V-/t/ words sounded closest to the word that they heard. They were told to imagine that they were listening to an L1 English speaker who was trying to learn to speak their language. After each identification, they mouse clicked on a graphical continuum to rate whether this stimulus was *close* or *far away* from this L1 vowel category. Subjects completed 28 trials (two repetitions of 14 English /b/-V-/t/ words).

5. Vowel-space mapping

In separate experiments, subjects found best exemplars for /b/-V-/t/ words in both English and their L1. Subjects first heard the speaker read *The North Wind and the Sun* in the respective language to familiarize them with the talker. On each trial, subjects saw a /b/-V-/t/ word on the computer screen (e.g., *bot*), as well as a more common word that had the same vowel (e.g., *hot*), and heard a stimulus (synthesized /b/-V-/t/ embedded in a natural carrier sentence). They rated on a continuous scale how far away the /b/-V-/t/ that they heard was from being a good exemplar of the printed word. Their ratings were given by mouse clicking on a continuous bar presented on a computer screen.

A goodness optimization procedure (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003; Iverson *et al.*, 2006) was used to iteratively change the stimuli that subjects heard on each trial, to search through the multidimensional stimulus space for better exemplars. Estimates of best exemplar locations were able to be found after 35 trials per vowel, despite the large stimulus set (109,375 stimuli). The procedure involved searching along individual vectors (i.e., one-dimensional straight-line paths crossing through the five-dimensional stimulus space), and finding the best exemplar on each vector. There were a total of seven search vectors and five trials per vector for each vowel. The vectors were chosen so that Vector 1 would allow most subjects to find a close approximation of their best exemplar (the search path passed through frequencies measured from natural productions), Vectors 2–6 orthogonally varied the five acoustic dimensions over a wide range, and Vector 7 fine tuned the position of the best exemplar.

Specifically, Vector 1 was a straight-line path that passed through two points: (1) the F1 and F2 frequencies at the beginning and ending of the natural productions of the target vowel, and (2) a neutral stimulus in the middle of the vowel space (F1=500 Hz and F2=1500 Hz, at both the onset and offset); duration was not varied along this vector, so each of the points was thus defined by four frequency values. Vector 2 varied duration, keeping formant frequencies fixed. Vector 3 varied the onset F1 and F2 frequencies (i.e., duration and offset formant frequencies were fixed) along the same basic path as the first vector (i.e., through a straight-line path including a neutral vowel and the onset formant frequencies of the natural production). Vector 4 was orthogonal to Vector 3 in the F1/F2 onset space. Vectors 5 and 6 were analogous to Vectors 3 and 4, except that the offset F1 and F2 frequencies were varied. Vector 7 varied all dimensions, passing through the best value found thus far on each dimension and the neutral vowel.

The endpoints of all vectors were constrained by the boundaries of the vowel space. For example, Vector 1 for *beat* crossed diagonally across the vowel space, starting from the high-front boundary of the space (i.e., low F1 and high F2), passing through the middle of the space, and ending at the low-back boundary of the space (i.e., high F1 and low F2). That is, the search spanned the entire space so that listeners had freedom to choose whatever acoustic values that they thought were the best, rather than being constrained to stimuli near *beat*.

The best exemplars were found for each vector over five trials. On the first two trials, subjects heard the most extreme stimuli that it was possible to synthesize along the vector (e.g., in the case of *beat*, they heard extreme high-front and low-back vowels, with the order of these two trials randomized). The selection of stimuli on the remaining trials was based on the subjects' judgments, using formulas that were designed to find stimuli along the path that would be perceived as better exemplars. On the third trial, subjects heard a stimulus that was selected by a weighted average of the first two stimuli, according to the equation

$$c = a^* \frac{f(b)}{f(a)+f(b)} + b^* \frac{f(a)}{f(a)+f(b)}, \quad (1)$$

where a and b are the positions on the search path for the first two trials; $f(a)$ and $f(b)$ are the goodness ratings for the stimuli on those trials (the goodness responses of *close* to *far away* were scaled from 0 to 1); and c is the new path position selected for the third trial. On the fourth and fifth trials, the stimuli were selected by finding the minimum of a parabola that was defined by the equation

$$\min = \frac{b - 0.5^*\{[b-a]^{2*}[f(b)-f(c)] - [b-c]^{2*}[f(b)-f(a)]\}}{[b-a]^*[f(b)-f(c)] - [b-c]^*[f(b)-f(a)]}, \quad (2)$$

where b is the path position of the best stimulus found thus far; a and c are the most recently tested positions on either

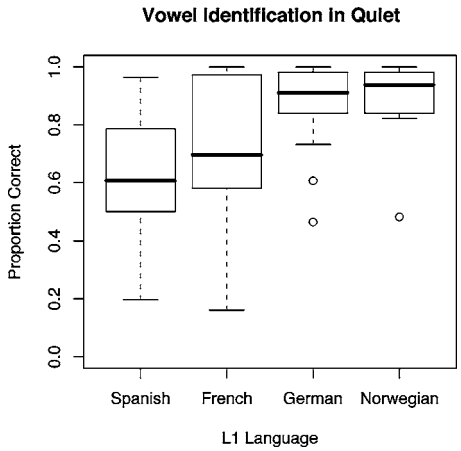


FIG. 1. Accuracy of natural English vowel identification in quiet for the four language groups. The boxplots display the distribution of individual differences (quartile ranges), with outliers indicated by circles (i.e., points that would otherwise make the whiskers longer than 1.5 times the interquartile range).

side of b; and $f(a)$, $f(b)$, and $f(c)$ are the goodness ratings for those stimuli. At the completion of the fifth trial, subjects were allowed to repeat the search if it had produced a poor exemplar (i.e., they made an explicit judgment about whether the stimulus was or was not close to sounding good). If the best exemplar was judged to be close, the parameters of the best stimulus found thus far were passed onto the next stage of the search algorithm (i.e., to search along the next vector).

The searches for the different words were interleaved. Specifically, the search progressed stage by stage, such that listeners completed Vector 1 for all words, then completed Vector 2 for all words, etc., with the word order randomized for each vector. Listeners thus switched vowel categories relatively frequently (they had five trials in a row with the same vowel) rather than repeatedly making judgments on single vowels.

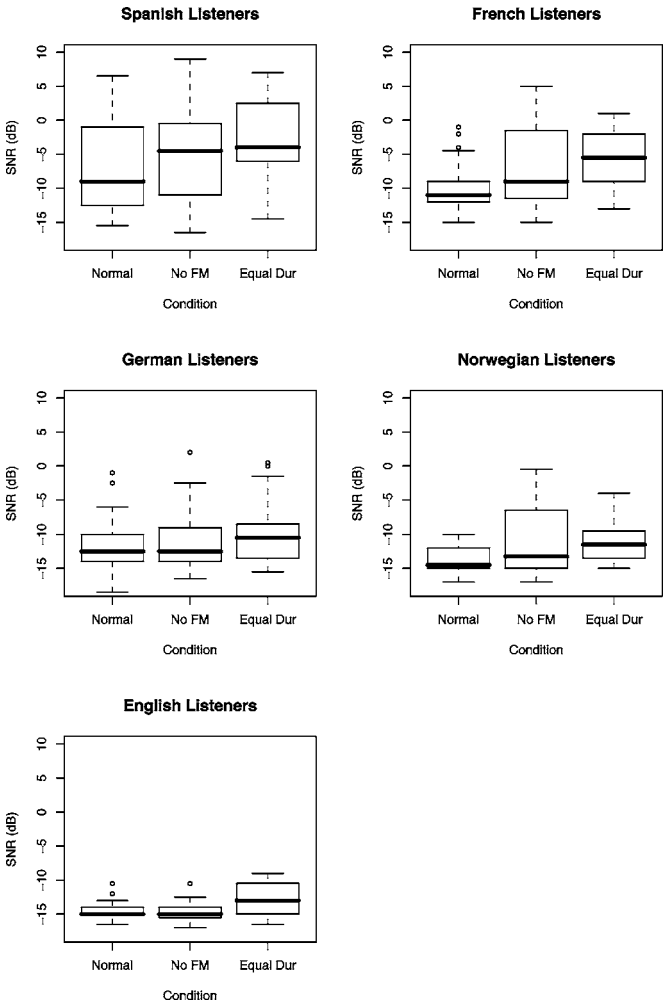


FIG. 2. Boxplots of the SNR thresholds (50% identification accuracy) for the different language groups, for natural vowels (Normal), vowels in which the formant movement had been flattened (No FM), and vowels in which duration had been equated (Equal Dur).

III. RESULTS AND DISCUSSION

A. English vowel identification in quiet

Figure 1 displays the accuracy with which each L1 group identified English vowels in quiet. There were obvious differences between the groups; Germans and Norwegians were more consistently accurate than Spanish and French listeners. A one-way ANOVA on arcsine-transformed scores demonstrated that this effect of L1 was significant, $F(3,79)=10.20$, $p<0.001$.

B. English vowel identification in noise

Figure 2 displays the threshold SNR values for each group of listeners (i.e., the level at which English vowels were identified 50% correct). A repeated-measures ANOVA examined the effect of L1 and stimulus condition (normal, no formant movement, and no duration contrast). There was a significant main effect of L1, $F(4,91)=23.39$, $p<0.001$, with Spanish and French speakers having poorer thresholds than German, Norwegian, and English speakers (i.e., the same pattern as for identification in quiet). There was also a significant main effect of condition, $F(2,184)=12.35$, $p<0.001$; across language groups, SNR thresholds were raised by an average of 1.5 dB when formant movement was flattened and 2.6 dB when duration was equated. There was no significant interaction between L1 and condition, $p>0.05$. Figure 2 makes it appear as if the median change in thresholds may have varied for the different language groups (e.g., Spanish speakers had larger differences in terms of the medians), but the between-language differences were small compared to the individual variability.

The results thus indicate that the L1 groups use formant movement and duration to similar extents when recognizing English vowels. Although the differences may seem small, a 1 dB change in the SNR translates into about a 7 percentage-point change in recognition accuracy (estimated by inspection of the psychometric functions for the present experiment). Thus, flattening formant movement reduced recognition accuracy by about 10 percentage points and equating duration reduced recognition accuracy by about 18 percentage points, which is comparable to the respective 13 and 14 percentage-point reductions that we found previously with L1 English speakers who had cochlear implants (Iverson *et al.*, 2006).

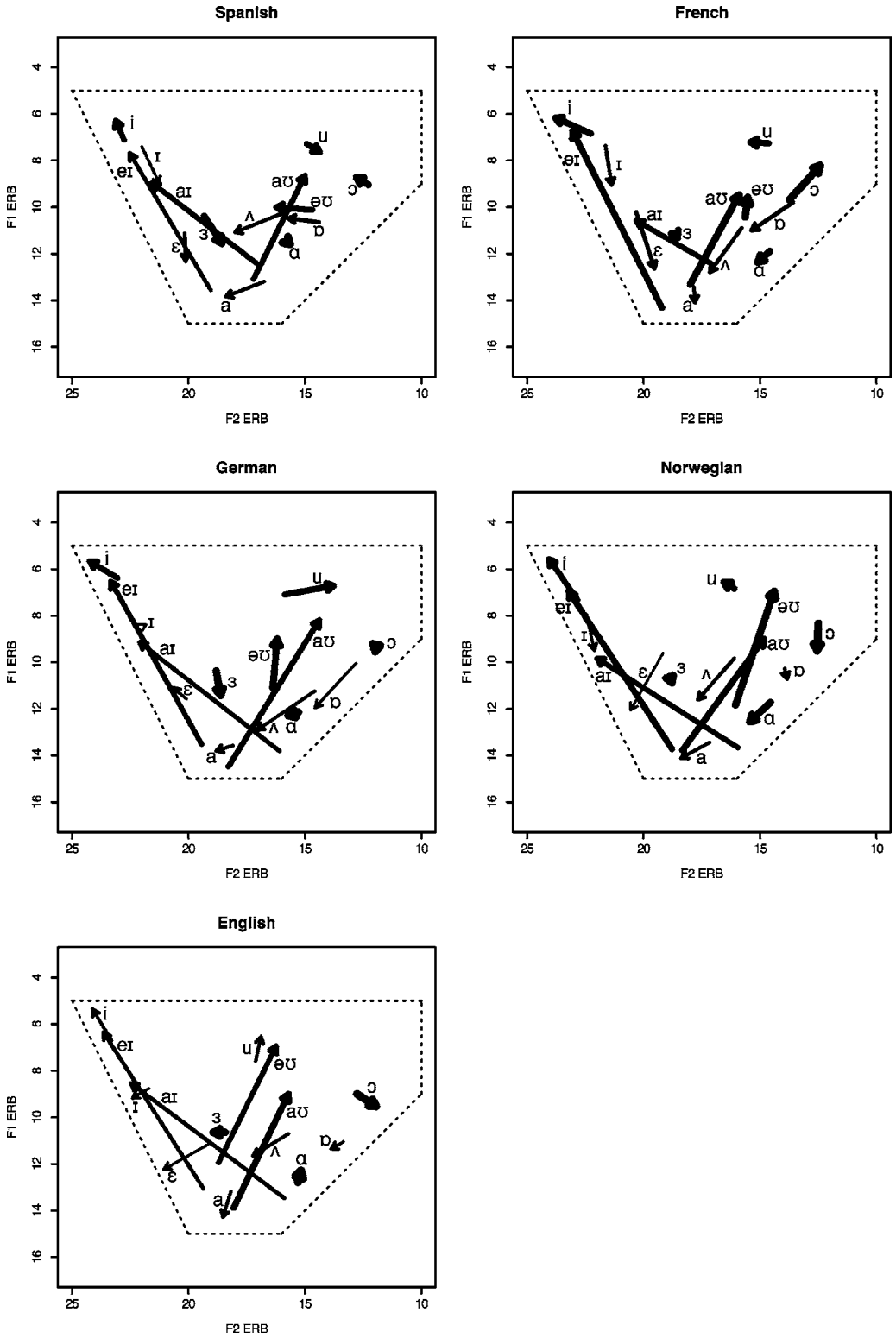


FIG. 3. Average best exemplar locations of English vowels for the different language groups. Each vowel is represented as an arrow from the starting F1 and F2 frequencies to the ending F1 and F2 frequencies (i.e., indicating the degree of formant movement). Duration is indicated by the weight of the line (i.e., thicker lines for longer vowels). Dotted lines indicate the boundaries of the vowel space (i.e., limits of the vowels that had been synthesized).

C. English vowel spaces

Figure 3 displays the mean best exemplars for English vowels by each language group. Broadly speaking, the results from the language groups were similar, although there were some apparent differences for individual vowels (e.g., *boat* was often dissimilar to that of English listeners). That is, all L1 groups contrasted both duration and formant movement, and the vowels were in positions in the space that were near to those chosen by L1 English speakers.

In order to evaluate how accurately the L2 English speakers approximated L1 English vowels, the distance was calculated between each individual's best exemplars and the

average best exemplars for L1 English subjects. These distances were calculated separately for F1/F2 location, formant movement, and duration. The F1/F2 location accuracy was measured by averaging the beginning and ending frequency of each vowel for F1 and F2, giving a two-dimensional F1/F2 coordinate for that vowel with no formant movement. The Euclidean distance (i.e., root mean square) was then calculated between the F1/F2 locations of each individual's English best exemplars and the L1 English averages. Formant movement accuracy was measured by subtracting the F1/F2 location values above, so that each vowel was represented as a vector representing the direction and magnitude of F1/F2 formant movement, with the center of each line passing through zero (i.e., normalizing the vowel's location in the vowel space). As above, Euclidean distances between these formant movement vectors were measured for each individual's vowels and the L1 English averages. Duration accuracy was quantified by calculating the average absolute-value difference between the durations of each individual's best exemplars and those of the L1 English averages.

The accuracy measures are displayed in Fig. 4. Accuracy in best exemplar locations had the same basic pattern as identification accuracy for English vowels in quiet, with Norwegians and Germans being more accurate than French or Spanish speakers. Separate one-way ANOVAs confirmed that the L1 groups were significantly different in terms of F1/F2 location accuracy, $F(4,95)=10.85$, $p<0.001$, formant movement accuracy, $F(4,95)=8.98$, $p<0.001$, and duration accuracy, $F(4,95)=9.12$, $p<0.001$.

Pearson correlations compared the individual differences in vowel-space accuracy measures with arcsine-transformed identification accuracy for English vowels in quiet. F1/F2 location was significantly correlated with identification across language groups, $r=-0.63$, $p<0.001$, and within each language group: Spanish, $r=-0.65$, $p<0.001$; French, $r=-0.49$, $p=0.034$; German, $r=-0.62$, $p=0.002$; and Norwegian, $r=-0.54$, $p=0.022$. Formant movement accuracy was significantly correlated with identification across language groups, $r=-0.54$, $p<0.001$, and within language groups for Germans, $r=-0.65$, $p=0.001$, and Norwegians, $r=-0.53$, $p=0.024$; the correlations were not significant within the Spanish, $r=-0.36$, and French, $r=-0.32$, language groups, $p>0.05$. Duration accuracy was significantly correlated, although weakly, with identification across language groups, $r=-0.27$, $p=0.014$, but was not significant within the Spanish, $r=-0.13$, French, $r=-0.39$, German, $r=-0.37$, and Norwegian, $r=0.01$, language groups, $p>0.05$.

Separate ANOVAs tested whether these relationships between vowel-space accuracy and identification differed between the L1 groups; arcsine-transformed identification accuracy was the dependent measure, L1 was an independent factor, and each vowel-space measure was entered in separate analyses as a covariate. There were no significant interactions of L1 group with F1/F2 location accuracy, formant movement accuracy, or duration accuracy, $p>0.05$. This suggests that the relationships between the accuracy measures and identification were similar for each L1. Confirming our previous analyses, in each of these ANOVAs there were

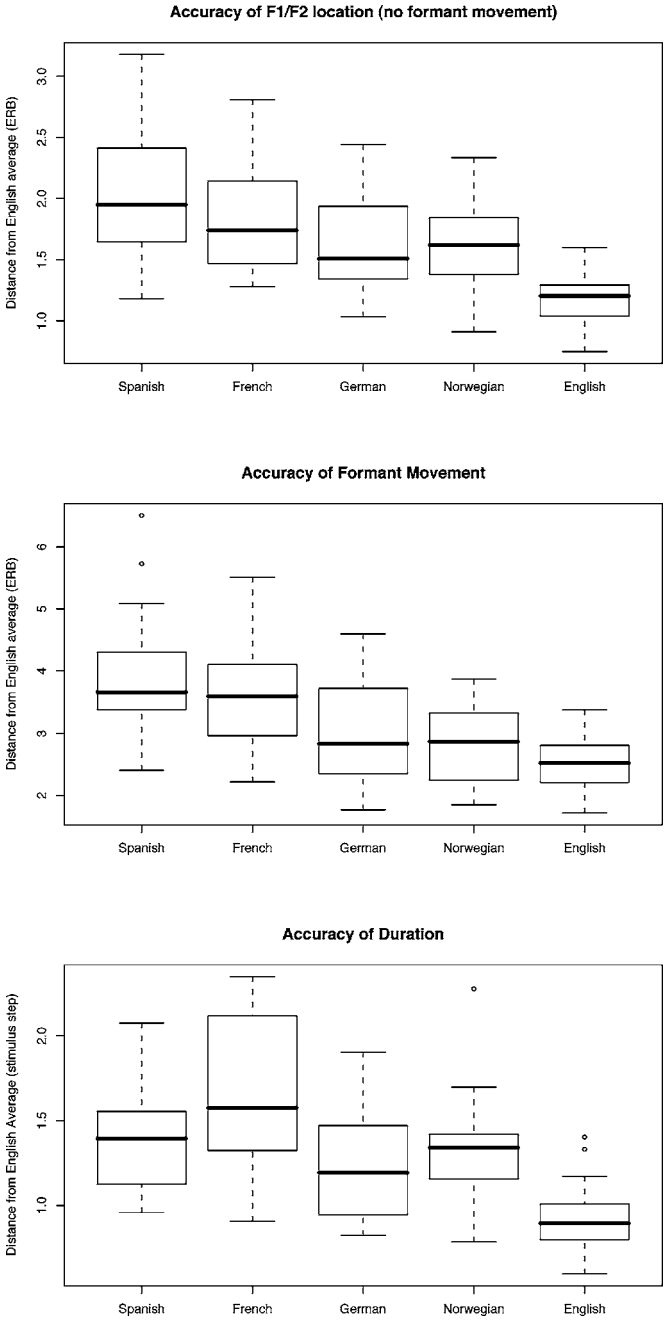


FIG. 4. Boxplots of the accuracy of each language group's English best exemplars (i.e., distance from the averages of native English speakers), in terms of F1/F2 location (i.e., static target frequencies), formant movement (i.e., direction and magnitude of change in F1 and F2 during each vowel), and duration.

significant main effects of L1 [F1/F2 location ANOVA: $F(3,75)=14.25$, $p<0.001$; formant movement accuracy ANOVA: $F(3,75)=11.98$, $p<0.001$; and duration accuracy ANOVA: $F(3,75)=10.72$, $p<0.001$], and each accuracy measure [F1/F2 location: $F(3,75)=34.78$, $p<0.001$; formant movement accuracy: $F(3,75)=16.26$, $p<0.001$; and duration accuracy: $F(3,75)=5.53$, $p=0.021$].

The results from the L2 English vowel spaces are thus in accord with those of English vowel identification in noise. That is, there were large differences in accuracy between L1 groups, but their reliance on F1/F2 location, formant movement, and duration was similar. Despite this commonality, it

should be noted that individual differences in the vowel-space accuracy measures were not always strongly correlated. For example, the correlation between F1/F2 location and duration accuracy across language groups was $r=0.30$, $p=0.004$, and the correlation between F1/F2 location and formant movement accuracy was $r=0.55$, $p<0.001$. The magnitude of these correlations leaves most of the variance unexplained, suggesting that individuals had idiosyncratic patterns of cue use (e.g., being accurate at duration but poor at F1/F2 location, rather than being equally accurate with all cues). That is, individual differences in cue use exist, but they do not appear to be strongly related to L1 background.

D. L1 assimilation and L1 vowel spaces

Table I displays the L1 vowels which were judged, on average, to be the most closely related to each English vowel, as well as listing the average assimilation rating for that vowel (0 for different to 1 for same). The closest vowel was defined by combining identification frequency (i.e., the proportion of trials in which each L1 response category was judged to be the closest) and the average assimilation rating for that vowel (i.e., the two numbers were multiplied and the maximum defined the closest vowel). For Spanish and French listeners, multiple English vowels often assimilated to the same L1 category. For example, Spanish listeners thought that the English vowels /a/, /aɪ/, /aʊ/, and /ɑ/ were all related to the Spanish /a/, with varying degrees of assimilation. However, Germans and Norwegians assimilated most English vowels to a unique L1 vowel. The only exceptions were English /ɔ/ and /əʊ/ which assimilated to German /o/, and English /aʊ/ and /əʊ/ which assimilated to Norwegian /æʉ/. Thus, there should be less pressure for Germans and Norwegians to learn new vowel categories when listening to English vowels, because they would not make many errors, in theory, if they simply used their existing L1 vowel categories to understand English. In contrast, Spanish and French listeners must learn new categories to avoid confusing English vowels.

Figure 5 displays the average L1 vowel spaces for each listener group. At its most basic level, this illustrates the substantial ways in which the vowel spaces of these listeners differed both in terms of the numbers of vowels and the use of formant movement and duration. Given these large differences in L1 vowel systems, it is notable that the English vowels in Fig. 3 were so similar across groups. That is, it is immediately apparent that subjects did not simply use their existing L1 vowel categories when listening to English (e.g., many of the English best exemplars used by Spanish listeners have no obvious counterpart in Spanish).

Item analyses (Table I) were conducted to determine whether the best exemplars of L2 English vowels were closer to L1 English vowels (i.e., average best exemplars for L1 English speakers) or to the closest vowel in each listener's L1. For example, Spanish listeners judged that English /eɪ/ was closest to Spanish /e/. For each Spanish speaker, we calculated how far their own best exemplar for English /eɪ/ was from the average /eɪ/ chosen by L1 English speakers (Fig. 3) and from the average Spanish /e/ (Fig. 5). The dis-

TABLE I. Degrees of assimilation of English vowels into L1 categories, and item analyses of whether L2 English best exemplars were closer to the L1 English vowel or the closest L1 vowel.

Assimilation			Item Analysis (t statistic)		
English vowel	Closest L1 vowel	Average rating	F1/F2 location	Formant movement	Duration
Spanish speakers					
i	i	0.78	0.49	−1.93	−2.40
ɪ	i	0.71	−0.89	−3.63[a]	−1.43
eɪ	e	0.44	0.05	−6.24[a]	1.89
ɛ	e	0.73	−2.19	−3.96[a]	1.89
a	a	0.75	−1.06	2.22	−1.43
aɪ	a	0.39	−8.52[a]	−4.21[a]	−1.56
aʊ	a	0.37	−4.90[a]	−4.76[a]	−1.36
ɑ	a	0.62	−1.80	−2.52	−6.17[a]
ɒ	o	0.72	2.32	0.78	2.40
ɔ	o	0.64	−4.02[a]	0.86	−2.03
əʊ	o	0.43	0.85	1.30	−0.55
ɜ	o	0.57	−18.13[a]	−3.74[a]	−6.00[a]
ʌ	o	0.63	−3.04	−1.20	0.77
u	u	0.71	0.79	−0.19	1.89
French speakers					
i	i	0.86	0.32	0.72	1.47
ɪ	i	0.87	−0.78	−4.29[a]	−0.56
eɪ	ɛ	0.70	−2.57	−43.13[a]	2.23
ɛ	ɛ	0.86	−0.29	0.20	−0.39
a	a	0.91	0.05	−1.59	−2.28
aɪ	a	0.57	−2.15	−0.37	−1.16
aʊ	a	0.47	−2.71	−3.86[a]	−2.19
ɑ	ɑ	0.80	−5.16[a]	−0.47	−1.28
ɒ	ɔ	0.83	−0.30	−4.94[a]	0.69
ɔ	o	0.84	−1.01	0.05	−2.76
əʊ	o	0.74	0.28	1.01	5.46[b]
ɜ	ø	0.80	−4.73[a]	−2.07	−4.47[a]
ʌ	ø	0.85	−3.75[a]	0.82	−0.77
u	u	0.82	−0.19	−1.22	2.28
German speakers					
i	i	0.83	−1.72	−6.46[a]	1.90
ɪ	ɪ	0.87	−0.74	−2.04	2.75
eɪ	e	0.53	−11.14[a]	−44.80[a]	−0.94
ɛ	ɛ	0.85	−2.09	−1.24	−1.10
a	a	0.75	−1.36	−1.95	−1.58
aɪ	aɪ	0.81	−6.30[a]	1.62	−0.59
aʊ	aʊ	0.78	−0.65	4.39[b]	−1.58
ɑ	aː	0.75	−1.61	−1.43	−1.75
ɒ	ɔ	0.84	−2.75	1.56	−0.90
ɔ	o	0.65	−1.35	−2.33	0.21
əʊ	o	0.49	−3.56[a]	−0.82	1.38
ɜ	ø	0.63	−8.44[a]	−3.15	1.74
ʌ	ʊ	0.74	−8.22[a]	0.39	0.04
u	u	0.71	0.25	−6.08[a]	0.68
Norwegian speakers					
i	iː	0.87	−0.94	−2.25	−0.16
ɪ	i	0.85	−5.16[a]	−4.13[a]	0.46
eɪ	ɑi	0.77	−0.46	5.71[b]	0.87
ɛ	ɛ	0.86	0.90	−0.79	−1.46
a	æ	0.85	−2.60	−0.22	0.00
aɪ	ɛi	0.85	−7.55[a]	5.47[b]	−1.13
aʊ	æʉ	0.68	−4.72[a]	0.98	−1.09
ɑ	ɑː	0.84	−1.41	1.80	1.46
ɒ	ɔ	0.85	−0.76	0.54	−0.46
ɔ	oː	0.83	0.03	1.01	0.05

TABLE I. (*Continued.*)

Assimilation			Item Analysis (t statistic)		
English vowel	Closest L1 vowel	Average rating	F1/F2 location	Formant movement	Duration
ɘʊ	æʉ	0.59	4.26[b]	0.76	1.26
ɜ	øː	0.81	−9.28[a]	−7.53[a]	−0.32
ʌ	œ	0.84	−1.05	−1.49	−0.94
u	ʉː	0.83	−1.82	−1.76	1.12

[a] $p<0.003$ closer to L1 English vowel.
[b] $p<0.003$ closer to L2 vowel.

tance calculations used the same metrics as in the accuracy analyses of Fig. 4. Paired *t* tests were used to determine whether the L2 English best exemplars were significantly closer to L1 English vowels (indicating learning), significantly closer to the closest L1 vowel of the listener (indicating L1 assimilation), or were nonsignificant (indicating either that the L2 English best exemplars were inbetween these two vowels, or that the variability was higher than the difference between these vowels). The significance level of $p<0.003$ was chosen to correct for multiple tests.

Spanish speakers clearly learned new aspects of the English vowel system. For diphthongs, they chose formant movement for /eɪ/, /aɪ/, and /aʊ/ that was more English-like than their Spanish patterns of formant movement, and they likewise had more English-like F1/F2 locations for /aɪ/ and /aʊ/. For monophthongs, they chose English-like F1/F2 locations for /ɔ/ and /ɜ/, English-like patterns of formant movement for /ɪ/, /ɛ/, and /ɜ/, and English-like durations (i.e., longer than Spanish) for /ɑ/ and /ɜ/. There was no evidence that these listeners preferred significantly more Spanish-like vowels when listening to English.

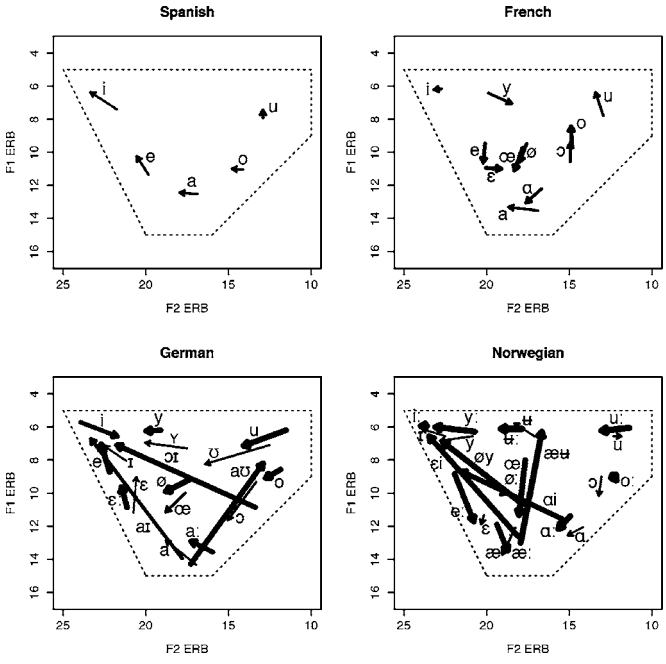


FIG. 5. Average best exemplar locations of L1 vowels for the different language groups. Each vowel is represented as an arrow from the starting F1 and F2 frequencies to the ending F1 and F2 frequencies (i.e., indicating the degree of formant movement). Duration is indicated by the weight of the line (i.e., thicker lines for longer vowels). Dotted lines indicate the boundaries of the vowel space (i.e., limits of the vowels that had been synthesized).

French speakers also appeared to have acquired new English vowels. For diphthongs, they had English-like formant movement for /eɪ/ and /aʊ/. For monophthongs, they had English-like F1/F2 locations for /ɑ/, /ɜ/, and /ʌ/, formant movement for /ɪ/ and /ɒ/, and durations for /ɜ/. There was evidence that they preferred French durations for /əʊ/; this was a small difference and occurred because they chose slightly longer durations for that vowel than did English speakers.

Despite the fact that German listeners would not need to form many new English categories in order to distinguish English vowels, they showed evidence of learning. Germans chose more English-like F1/F2 locations for /eɪ/, /aɪ/, /əʊ/, /ɜ/, and /ʌ/, and more English-like formant movement for /i/, /eɪ/, and /u/. Their formant movement for /aʊ/ was significantly more like the corresponding vowel in German, indicating a degree of L1 assimilation. None of the durations were significantly more like L1 English or L1 German, because the corresponding English and German vowels had very similar durations.

Norwegians had some L1 assimilation for diphthongs; their English vowels were more significantly like Norwegian in terms of F1/F2 location for /əʊ/, and in terms of formant movement for /eɪ/ and /aɪ/. However, these listeners still were more English-like for many vowels; their English vowels were significantly more like L1 English vowels in terms of F1/F2 location for /ɪ/, /aɪ/, /aʊ/, and /ɜ/, and in terms of formant movement for /ɪ/ and /ɜ/. None of the durations were significantly more like L1 English or L1 Norwegian.

One of the claims of Flege's Speech Learning Model (SLM) (Flege 1995, 2003) is that these patterns of learning ought to be predictable from assimilation. That is, vowels that are weakly assimilated into L1 categories should be easier to learn than vowels that are more strongly assimilated. To test this possibility, an ANOVA was conducted with the average assimilation rating for the closest vowel as the dependent measure (i.e., as listed in Table I), and L1 and learning coded as independent categorical variables. For the learning variable, a vowel was coded as "learned" if it was significantly closer to an L1 English vowel on any of the three accuracy measures, and "not learned" if it was not. There was a main effect of L1, $F(3,48)=7.81$, $p<0.001$. This occurred because the assimilation ratings of Norwegians were higher than those of Spanish listeners, which makes sense given the differences in the vowel systems (i.e., Norwegians have a more crowded L1 space, and thus have more vowels that are acoustically close to English vowels). However, there was no significant main effect of learning or interaction, $p>0.05$. Thus there was little evidence that the vowels that were weakly assimilated into the L1 vowel system were easier to learn.

IV. GENERAL DISCUSSION

The results demonstrated that there were large and consistent differences between the language groups in terms of

their identification accuracy in quiet and in noise, as well as the degree to which their English best exemplars matched those of L1 English speakers. Although these group differences seem to be obvious effects of L1 vowel systems, with individuals who have larger L1 vowel systems being more accurate with L2 English vowels, this interpretation is qualified by the fact that our subject populations were not selected to be matched based on English experience, and many relevant factors could not have been matched across L1 groups even if we had tried (e.g., the fact that English television programs are dubbed in Germany but subtitled in Norway). That being said, there remains a clear effect of L1; it is difficult to imagine that other factors could explain, for example, the fact that Norwegians were much more accurate at identifying English vowels than were Spanish speakers.

Beyond these large overall differences, there was surprisingly little evidence that the language groups perceived English vowels in fundamentally different ways, despite the large differences in L1 vowel systems and the heterogeneity of the subject groups. For example, SNR thresholds for vowels increased when formant movement was flattened, with no reliable differences between language groups. This result demonstrates that even the subtle patterns of formant movement among English monophthongs were important to L2 listeners, as has been found previously for native English speakers (e.g., Assmann and Katz, 2005; Hillenbrand and Nearey, 1999; Iverson *et al.*, 2006). Moreover, the accuracy with which listeners represented formant movement in their best exemplars was correlated with identification accuracy, and this relationship did not differ reliably between groups. That is, even though Germans and Norwegians were more accurate with regard to formant movement than were Spanish and French speakers, there is evidence that formant movement was an important part of vowel perception for all groups.

Identification accuracy in noise was likewise reduced when duration was equated, and this reduction was not significantly different between the language groups. The accuracy with which duration was represented in the best exemplars was relatively weakly correlated with identification accuracy in quiet, reaching significance when all subjects were included but not when calculated within each language group. This seems to confirm the status of duration as a more secondary cue in English (e.g., Hillenbrand *et al.*, 2000). That is, the representation of F1/F2 target frequencies is a more significant cause of individual differences in vowel recognition, but the use of duration can have value when the formant information is less clear (e.g., noisy conditions).

The L1-related differences in F1/F2 targets, formant movement, and duration are particularly notable given that we previously found few such differences when testing cochlear implant users (Iverson *et al.*, 2006). That is, postlingually deafened cochlear implant users were nearly as accurate in their best exemplar locations as were normal-hearing individuals, despite the fact that the cochlear implant users averaged only 74% correct when recognizing natural vowels. The vowel-space mapping task thus appears to be relatively unaffected by peripheral distortions like these, probably because the task demands are low (e.g., listeners can repeatedly listen to the stimuli). These low task demands, as well as the fact that the vowel judgments were made with reference to a particular talker's voice, may have contributed to the general similarity of the mean best exemplars among L1 groups in the present study (i.e., Fig. 3). However, the individual differences in these L2 English best exemplars and their correlation with identification accuracy suggest that the task was still sensitive to variation in the underlying representations for these phonemes.

Although the L2 speech perception literature has emphasized L1 assimilation (e.g., Best, 1995; Best *et al.*, 1988; Best *et al.*, 2001; Flege, 1995, 2003), the present results are mixed with regard to the role of assimilation. On one hand, the large overall differences between L1 groups could be viewed as being in accord with L1 assimilation. That is, the assimilation patterns listed in Table I should have offered an advantage to German and Norwegian speakers at the start of their English studies because they had few instances where multiple English vowels mapped onto a single L1 category; the present results suggest that this advantage may persist even after individuals have years of English experience. However, these L1 assimilation patterns also imply that Germans and Norwegians would have been under less pressure to learn English vowel categories; our item analyses suggest instead that they used many new vowels in English, rather than simply transferring their L1 vowels to English. This learning by Germans and Norwegians was particularly surprising given that their L1 vowel inventories were already quite crowded; these individuals were expected to have difficulty learning new vowel categories.

Degree of assimilation was poor at predicting which individual vowels were learned or not learned. For example, Spanish and Norwegian speakers did not demonstrate learning for English /əʊ/ despite having relatively low assimilation ratings, while Germans demonstrated learning for English /aɪ/ despite having relatively high L1 assimilation ratings. Listeners were able to learn many of the biggest and most obvious differences between the L1 and L2 vowels (e.g., Spanish and French speakers learning to add formant movement to English diphthongs), but were also able to learn subtle aspects, such as Spanish, French, and Norwegian listeners learning more English-like formant movement for /ɪ/.

Although this relatively poor correspondence of assimilation and L2 learning is contrary to SLM (Flege, 1995, 2003), it is worth noting that the weight of the evidence for SLM has been from L2 production data (e.g., Flege, 1987, 1995, 2003; Bohn and Flege, 1992). Moreover, the perceptual evidence for SLM has mostly involved having L2 learners identify pairwise contrasts (e.g., /i/ vs /ɪ/; Flege, *et al.*, 1997; 1999); such restricted identification tasks can reflect perceptual sensitivity as much as categorization. The present study is unique in its examination of the perception of entire vowel systems, and in the use of a task that allows for direct comparisons of the underlying L1 and L2 perceptual representations in multiple dimensions. It may simply be the case that assimilation has a stronger role in constraining the learning of new productions than in the learning of perceptual representation for new vowels.

The view that emerges from the present study is that L2 vowel learning is quite pervasive, with individuals learning even when L1 assimilation is sufficient to distinguish L2 vowels, and individuals learning secondary cues for vowels (e.g., formant movement in monophthongs, and duration) rather than simply learning more primary cues (e.g., static F1/F2 targets). This more holistic pattern of learning (i.e., learning primary and secondary cues together) is compatible with the notion that the underlying categories for vowels are phonetically detailed (Goldinger, 1996, 1998; Hawkins and Smith, 2001; Johnson, 1997; Nygaard *et al.*, 1995, Nygaard and Pisoni, 1998; Pisoni, 1997). This kind of exemplar learning implies that listeners would learn the details of a vowel all at once, rather than only learning whichever individual cues seem best for distinguishing categories. That being said, the evidence for this kind of holistic learning in the present study is mixed. In support of this conclusion, the between-group differences in accuracy of F1/F2 location, formant movement, and duration all follow the same basic pattern (i.e., Spanish and French speakers being less accurate), suggesting that language groups who have poor representations of primary cues also have poor representations of the more secondary cues. However, the individual differences in F1/F2 location and duration accuracy were only weakly correlated, demonstrating that individuals who were accurate at one cue were not necessarily accurate at all others. These idiosyncratic patterns of cue weighting (e.g., individuals representing duration more accurately than target formant frequencies) were not apparent in the cross-language comparisons simply because they were not strongly related to L1 background. It is thus possible that L2 learners may engage in cue-based learning, even though learning may appear to be holistic when looking at entire vowel systems and across L1 groups.

Although vowel category learning seems pervasive across L1 groups in the present study, learning L2 vowels is not always easy. For example, even highly experienced bilingual Spanish-Catalan speakers have difficulty with the Catalan /e/-/ɛ/ distinction if their first language was Spanish (Pallier *et al.*, 2001). In the present study, Spanish and French speakers made many errors recognizing English vowels despite having years of experience. Assimilation models may be able to explain some of the learning problems for individual vowels, but it is clear from the present results that assimilation alone does not fully explain the difficulties that individuals have when learning an L2 vowel system.

ACKNOWLEDGMENTS

We are grateful to Anke Sennema for subject recruitment and hosting the data collection at University of Potsdam, Germany, Dawn Behne for hosting the data collection at Norwegian University of Science and Technology, and Eivind Torgersen for advice on the Norwegian stimuli. This research was funded by Grant No. RES-000-23-0838 from the Economic and Social Research Council of the UK.

Assmann, P. F., and Katz, W. F. (**2005**). "Synthesis fidelity and vowel identification," J. Acoust. Soc. Am. **117**, 886–895.

Best, C. T. (**1995**). "A direct-realist view of cross-language perception," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Balitmore, MD), pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**, 775–794.

Boersman, P., and Weenink, D. (**2005**). "Praat: Doing phonetics by computer," [Computer software] (University of Amsterdam, Amsterdam, The Netherlands).

Bohn, O.-S. (**1995**). "Cross-language speech perception in adults: First language transfer doesn't tell it all," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD).

Bohn, O-S., and Flege, J. E. (**1990**). "Interlingual identification and the role of foreign language experience in L2 vowel perception," Appl. Psycholinguist. **11**, 303–328.

Bohn, O.-S., and Flege, J. E. (**1992**). "The production of new and similar vowels by adult German learners of English," Stud. Second Lang. Acquis. **14**, 131–158.

Cebrian, J. (**2006**). "Experience and the use of non-native duration in L2 vowel categorization," J. Phonetics **34**, 372–387.

Delattre, P. (**1965**). *Comparing the Phonetic Features of English, French, German, and Spanish* (Harrap & Co, London).

Escudero, P., and Boersma, P. (**2004**). "Bridging the gap between L2 speech perception research and phonological theory," Stud. Second Lang. Acquis. **26**, 551–585.

Evans, B. G., and Iverson, P. (**2004**). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," J. Acoust. Soc. Am. **115**, 352–361.

Evans, B. G., and Iverson, P. (**2007**). "Plasticity in vowel perception and production: A study of accent in young adults," J. Acoust. Soc. Am. **121**, 3814–3826.

Flege, J. E. (**1987**). "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," J. Phonetics **15**, 47–65.

Flege, J. E. (**1989**). "Differences in inventory size affect the location but not the precision of tongue positioning in vowel production," Lang Speech **32**, 123–147.

Flege, J. E. (**1995**). "Second Language speech learning: Theory, findings, and problems," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 233–277.

Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin, Germany).

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "The effect of experience on nonnative subjects' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J. E., MacKay, I. R. A., and Meador, D. (**1999**). "Native Italian speakers' production and perception of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Flege, J. E., Schirru, C., and MacKay, I. R. A. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. **22**, 1166–1183.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Gottfried, T., and Beddor, P. S. (**1988**). "Perception of spectral and temporal information in French vowels," Lang Speech **31**, 57–75.

Hawkins, S., and Smith, R. (**2001**). "Polysp: a polysystemic, phonetically-rich approach to speech understanding," Italian Journal of Linguistics **13**, 99–188.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized hVd utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

Iverson, P., and Evans, B. G. (**2003**). "A goodness optimization method for investigating phonetic categorization," Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona, Spain.
Iverson, P., Smith, C. A., and Evans, B. G. (**2006**). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement," J. Acoust. Soc. Am. **120**, 3998–4006.
Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar model," *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenix (Academic, San Diego, CA), pp. 145–165.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.
Krisfoffersen, G. (**2000**). *The Phonology of Norwegian* (Oxford University Press, Oxford, UK).
Ladefoged, P., and Broadbent, D. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.
Lee, B., Guion, S. G., and Harada, T. (**2006**). "Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals," Stud. Second Lang. Acquis. **28**, 487–513.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (**2001**). "Category restructuring during second-language (L2) speech acquisition," J. Acoust. Soc. Am. **110**, 516–528.
McAllister, R., Flege, J. E., and Piske, T. (**2002**). "The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian," J. Phonetics **30**, 229–258.
Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Le Besnerais, M. (**2003**). "Production and perception of foreign vowels: does the density of the system play a role?" Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.
Morrison, G. (**2002**). "Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results," *Proceedings of the North West Linguistics Conference 2002*, edited by G. S. Morrison and L. Zsoldes (Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada), pp. 29–48.
Munro, M. J., Flege, J. E., and MacKay, I. R. A. (**1996**). "The effects of age of second language learning on the production of English vowels," Appl. Psycholinguist. **17**, 313–334.
Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech perception," Percept. Psychophys. **60**, 355–376.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1995**). "Effects of stimulus variability on perception and representation of spoken words in memory," Percept. Psychophys. **57**, 989–1001.
Pallier, C., Colome, A., and Sebastian-Galles, N. (**2001**). "The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries," Psychol. Sci. **12**, 445–449.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Pisoni, D. B. (**1997**). "Some thoughts on 'normalization' in speech perception," *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenix (Academic, San Diego, CA), pp. 9–32.
Stockwell, R. P., and Bowen, J. D. (**1965**). *The Sounds of English and Spanish* (University of Chicago Press, Chicago, IL).
Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. (**2005**). "Contextual variation in the acoustic and perceptual variation of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–1762.

Factors affecting vowel formant discrimination by hearing-impaired listeners[a)]

Chang Liu[b)] and Diane Kewley-Port[c)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 24 August 2005; revised 17 July 2007; accepted 7 August 2007)

The goal of this study was to measure the ability of adult hearing-impaired listeners to discriminate formant frequency for vowels in isolation, syllables, and sentences. Vowel formant discrimination for F1 and F2 for the vowels /ɪ ɛ æ ʌ/ was measured. Four experimental factors were manipulated including linguistic context (isolated vowels, syllables, and sentences), signal level (70 and 95 dB SPL), formant frequency, and cognitive load. A complex identification task was added to the formant discrimination task only for sentences to assess effects of cognitive load. Results showed significant elevation in formant thresholds as formant frequency and linguistic context increased. Higher signal level also elevated formant thresholds primarily for F2. However, no effect of the additional identification task on the formant discrimination was observed. In comparable conditions, these hearing-impaired listeners had elevated thresholds for formant discrimination compared to young normal-hearing listeners primarily for F2. Altogether, poorer performance for formant discrimination for these adult hearing-impaired listeners was mainly caused by hearing loss rather than cognitive difficulty for tasks implemented in this study.
 [DOI: 10.1121/1.2781580]

PACS number(s): 43.71.Ky, 43.71.Es, 43.66.Fe [KWG] Pages: 2855–2864

I. INTRODUCTION

It is well recognized that hearing-impaired listeners perform more poorly in speech perception tasks than normal-hearing listeners. Previous studies in vowel perception in our laboratory measured the performance of vowel formant discrimination for hearing-impaired listeners for isolated vowels, showing degraded performance compared to normal-hearing listeners (Coughlin *et al.*, 1998; Richie *et al.* 2003). However, little research has been done to investigate vowel formant discrimination for hearing-impaired listeners in more ordinary listening conditions with high-fidelity speech. The purpose of this study was to examine the performance of hearing-impaired listeners for vowel formant discrimination in a variety of stimulus conditions that included sentences with and without an identification task.

The ability to discriminate formant frequency under optimal and ordinary listening conditions has been systematically investigated for normal-hearing listeners in several studies by Kewley-Port and her colleagues (Kewley-Port and Watson, 1994; Kewley-Port and Zheng, 1999; Kewley-Port, 2001; Liu and Kewley-Port, 2004a, b). In optimal listening conditions, in which highly trained participants listened in quiet under minimal stimulus uncertainty testing, listeners showed the most accurate resolution of formant frequency. In more ordinary listening conditions for vowels in sentences, the formant resolution was degraded by a factor of 2.6 (Kewley-Port and Zheng, 1999). Many factors have been shown to affect formant resolution, including level of stimulus uncertainty (i.e., trial by trial variability in vowel stimuli), level of linguistic context, listener training, and background noise. Increases in stimulus uncertainty or level of linguistic context (from isolated vowels to sentences) degraded vowel formant discrimination performance (Kewley-Port and Zheng, 1999; Liu and Kewley-Port, 2004a). Background noise also significantly degraded formant resolution, particularly when noise level was equal to or higher than speech level (Liu and Kewley-Port, 2004b). However, when a complex sentence identification task was added to formant discrimination for vowels in sentence, the performance was the same for formant discrimination without the additional identification task (Kewley-Port and Zheng, 1999; Liu and Kewley-Port, 2004a), implying that the formant resolution for vowels may be independent of the identification of sentences under some conditions.

The above-specified studies used young normal-hearing listeners. However, to date, formant discrimination for hearing-impaired listeners has focused on only isolated vowels. Coughlin *et al.* (1998) examined formant discrimination ability and vowel identification for three groups of listeners: young normal-hearing (YNH), elderly normal-hearing (ENH), and elderly hearing-impaired listeners (EHI). The stimuli were four formant-synthesized American English vowels presented at 70 and 95 dB SPL. Results suggested that thresholds for vowel formant discrimination for F1 were the same across listener groups and stimulus levels. However, EHI listeners had higher thresholds than YNH for F2. For hearing-impaired listeners, vowel formant discrimination improved significantly from 70 to 95 dB SPL, especially for F2. They argued that improved performance was due to the better audibility at 95 dB SPL for the hearing-impaired lis-

[a)]Portions of the data were presented at the 146th Meeting of the Acoustical Society of America (J. Acoust. Soc. Am. 2003, 114, 2361).
[b)]Present address: Dept. of Communication Sciences and Disorders, 1 University Station A1100, The University of Texas at Austin, Austin, TX 78712. Electronic mail: changliu@mail.utexas.edu
[c)]Electronic mail: kewley@indiana.edu

teners. However, EHI listeners still had higher thresholds than YNH for F2 at 95 dB SPL, indicating that audibility alone was not sufficient for hearing-impaired listeners to achieve normal discrimination performance.

Another recent study of formant discrimination for isolated vowels investigated the formant resolution for listeners with hearing loss at three different intensity levels: soft conversation level at 60 dB SPL and two gain conditions with one having a flat frequency response (at 95 dB SPL) and the other having frequency-specific gain (Richie *et al.*, 2003). Results confirmed the findings of Coughlin *et al.* (1998) that hearing-impaired listeners had significantly better discrimination performance at the two gain conditions compared to the low conversational level, although these thresholds were markedly higher than those for normal-hearing listeners for both gain conditions. Thus, both of these two recent studies suggested that there were some factors other than audibility to account for the degradation in formant discrimination for hearing-impaired listeners.

There are several hypotheses to explain deficits in speech perception for hearing-impaired listeners. One hypothesis is that reduced speech-recognition abilities for hearing-impaired listeners are predominantly caused by hearing loss, including the decreased audibility and secondary distortions accompanying the loss of pure-tone sensitivity. In studies by Humes and colleagues (Humes *et al.*, 1987; Humes *et al.*, 1990), the nonsense syllable test was presented at different levels in quiet and noise conditions for hearing-impaired listeners, as well as normal-hearing listeners with simulated hearing loss. Results suggested that there was no significant difference in speech-recognition performance between noise-masked normal listeners and hearing-impaired listeners. Thus, the authors concluded that the loss of audibility was the major factor contributing to degraded speech recognition. Needleman and Crandell (1997) found that speech reception thresholds for the Speech Perception in Noise test were significantly better for noise-masked, normal-hearing listeners than those for hearing-impaired listeners. They concluded that reduced frequency selectivity might account for the degraded speech recognition for hearing-impaired listeners. In addition, Hedrick and Jesteadt (1997) reported that, to distinguish /p/-/t/ stops, normal-hearing listeners used cues in the formant transitions as well as the relative amplitude, while hearing-impaired listeners primarily used the relative amplitude in the F4 and F5 regions. They argued that low perceptual weighting given to formant transition cues by hearing-impaired listeners may be partially due to a peripheral physiology deficit that resulted in a poor representation of dynamic formant transition. Thus, results from these studies suggest that reduced spectral and temporal resolution, in addition to reduced audibility, contributed to hearing-impaired listeners' difficulty in understanding speech signals, consistent with the two studies of formant discrimination described earlier (Coughlin *et al.*, 1998; Richie *et al.*, 2003). However, these discrimination studies used quite artificial, steady-state vowels. The main purpose of this study was to use more normal and dynamic speech stimuli to assess formant discrimination in syllables and sentences as an extension of Liu and Kewley-Port's study (2004a). Specifically, that study of formant discrimination in high-fidelity syllables and sentences with young normal-hearing listeners was extended here to comparable adult hearing-impaired listeners.

Another hypothesis about speech recognition proposes that poor performance is due to cognitive factors other than hearing loss. Rakerd *et al.* (1996) assessed the cognitive demands of speech listening for young and elderly listeners with normal hearing and hearing loss using a dual-task method that included a digit memorization task and speech understanding task. Results suggested that, when cognitive load was high, hearing-impaired listeners had higher susceptibility, i.e., lower speech understanding performance and more errors in the digit memorization, than normal-hearing listeners had. They concluded that cognitive factors could prominently contribute to the difficulty of speech understanding for hearing-impaired listeners. On the other hand, degraded functions of formant discrimination were nearly equivalent between young (Richie *et al.*, 2003) and elderly hearing-impaired listeners (Coughlin *et al.*, 1998), suggesting that cognitive effects of aging were not important factors and rather that hearing impairment was the primary factor. However, it should be noted that because only isolated vowels were used in both studies, cognitive processing demands were limited. If higher level processing is required for target vowels embedded in sentences, it is unknown whether or not cognitive factors may additionally degrade discrimination performance for the hearing-impaired listeners. In our previous studies using sentences (Kewley-Port and Zheng, 1999; Liu and Kewley-Port, 2004a), a dual task condition (discrimination and identification) was included. Even though performance for formant discrimination for normal-hearing listeners was not degraded, an increase in cognitive load caused by the dual task might have affected the hearing-impaired listeners based on research by McCoy *et al.* (2005) and Wingfield *et al.* (2005). They suggested that consequences of even a mild hearing loss might cause elderly listeners to exert extra effort to successfully perceive speech. Thus a secondary goal of this study was to examine effects of cognitive load on vowel formant discrimination for hearing-impaired listeners using the same dual task of discrimination with identification in sentences as an extension of the Liu and Kewley-Port (2004a) study. In this additional identification task, listeners were required to identify both the target word and the carrier sentence (i.e., listeners had to store and recall additional items from working memory), and then discriminate formant frequency difference. However, given that this additional identification did not affect discrimination performance for normal-hearing listeners and that our hearing-impaired listeners were young to middle-aged, effects of this additional identification were not predicted.

In summary, four factors were manipulated to meet the purposes of the present study, signal level: 70 dB SPL and 95 dB SPL; formant frequency (F1 and F2 of four vowels, i.e., eight vowel formants); level of linguistic context (isolated vowels, syllables, and sentences); and cognitive load (discrimination in sentences with and without identification).

TABLE I. Acoustic parameters of isolated vowels and /bVd/ syllables.

	ɪ	ɛ	æ	ʌ
F1 (Hz)	430	581	678	700
F2 (Hz)	1454	1960	2078	2132
Isolated vowel duration (ms)	107	125	198	123
Syllable duration (ms)	166	187	310	188

II. METHOD

A. Stimuli

Four American English vowels /ɪ ɛ æ ʌ/ were used as stimuli and these vowels covered a wide range of formant frequencies from a low F1 to a high F2 over the English vowel space (see Table I). Thresholds for formant discrimination for F1 and F2 of these four vowels were measured in three types of linguistic context: isolated vowels, syllables, and sentences. In addition, for the sentence context, formant discrimination was examined with and without an additional word/sentence identification task described in the following.

The four original /bVd/ syllables recorded from a female talker were the same ones used in an earlier study (Liu and Kewley-Port, 2004a). Standard vowels without formant shifts and test vowels with formant changes were synthesized using a modified STRAIGHT procedure (Kawahara *et al.*, 1999). F1 and F2 for each standard syllable were shifted systematically by 24 steps using a linear scale. The range of formant frequency shifts was from 0.9% to 17% of the formant frequency for each F1 and F2. The procedure to shift target formant frequencies is described briefly as follows: First, a matrix in MATLAB representing the spectrogram (amplitude × time × frequency) of the standard syllable, was obtained by the analysis in STRAIGHT. Second, to shift a formant peak, the temporal location of the formant across the syllable, including transitions, was visually identified. Third, in each time frame (i.e., one spectrum), the formant shift was manipulated for the portion between the valleys on either side of the formant peak. Amplitude in the low-frequency valley was adjusted to be a constant across the frequency range corresponding to the frequency shift, while the high-frequency valley was collapsed by replacing the original amplitude values with the shifted peak, such that the shift in the selected formant frequency resulted in no change in other formants. Thus, detail in the formant peaks was preserved in this procedure, with the valleys only somewhat changed. Finally, this modified two-dimensional matrix was reloaded into STRAIGHT and used with other unchanged acoustic parameters such as F0 and amplitude contours for resynthesis [for more details, see Liu and Kewley-Port (2004c) and their Fig. 1]. Formant shifts included the steady-state portions as well as the onset and offset formant transitions in the /bVd/ syllables. The isolated vowels were subsequently edited by deleting the formant transition at the beginning and the end of the syllable such that only the steady-state vowel nucleus remained. Durations for the standard isolated vowels ranged from 107 to 198 ms (see Table I). Liu and Kewley-Port (2004a) reported that syllables, phrases, and sentences resynthesized from a female talker using STRAIGHT were essentially indistinguishable from the original in informal listening tasks. They termed the STRAIGHT-synthesized speech "high-fidelity" speech because of its high quality that resembled natural speech. These high-fidelity speech stimuli have been used for several recent studies on formant discrimination (Liu and Kewley-Port, 2004a,b,c).

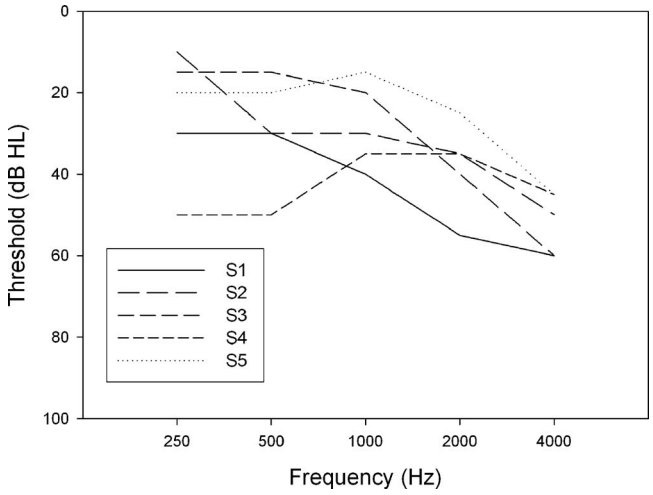


FIG. 1. Audiograms of the five hearing-impaired listeners from 250 to 4000 Hz, showing the configuration of hearing loss over the frequency range presented in this study.

The three nine-word sentence frames in which the word /bVd/ was embedded near the beginning, in the middle, or near the end position are shown in Table II. These three sentence frames recorded from the same female talker, sharing several common words with similar linguistic meanings, were selected in order to increase listeners' difficulty for identifying the sentence in the additional identification task. The standard and test syllables were inserted into the sentences using a digital waveform editor. Essentially, the original syllable was replaced by the resynthesized syllable. All the stimuli were presented at two levels, 70 and 95 dB SPL. Because Richie *et al.* (2003) reported that there was no difference in performance for formant discrimination for hearing-impaired listeners between a shaped and a flat gain of 95 dB SPL, both of which increased the level of the speech signal to well above threshold for octave frequencies below 4000 Hz, only the flat gain was employed in this study.

B. Listeners

Five adult listeners from 21 to 55 years old with hearing loss participated in this study. They were American-English native speakers and were paid for their participation in this study. The age range for hearing-impaired listeners was cho-

TABLE II. Sentences used in the experiment of vowel formant discrimination.

Three sentences used for linguistic context of sentences:
The /b**V**d/ word is first on the page now.
The first word is /b**V**d/ on the page again.
The first word on the book is /b**V**d/ now.

sen as 20–50 years in order to compare results obtained here to those reported by Liu and Kewley-Port (2004a). All listeners had normal middle ear functions. To assure audibility of the vowels at 95 dB SPL, the audiometric criteria for listener selection were thresholds between 25 and 60 dB HL at 2000 Hz and no thresholds greater than 65 dB HL within the range of 250–4000 Hz. Magnitude and configuration of hearing loss are shown in Fig. 1, indicating mild to moderate high-frequency sloping hearing loss of cochlear origin.

C. Procedures

Speech stimuli were presented to the right ears of listeners who were seated in a sound-treated, IAC booth, via TDH-39 earphones. Stimulus presentation was controlled by TDT modules including a 16-bit D/A converter, a programmable filter, and a headphone buffer with a sample rate of 11 025 Hz. A low-pass filter with a cutoff frequency of 5000 Hz and a slope of 80 dB/octave, and an attenuation level set by the calibration procedure, was configured in the programmable filter. The standard vowel /ɛ/ lengthened to a duration of 3 s was used for calibration. The sound-pressure level measured in an NBS-9A 6-c^3 coupler by a Larson-Davis sound-level meter (Model 2800) using the linear setting was adjusted to be either 70 or 95 dB SPL. The amplification of the other standard syllables was adjusted such that the signal level was also either 70 or 95 dB SPL.

Thresholds of vowel formant discrimination were measured for F1 and F2 for the four English vowels, using a three-interval, forced-choice procedure with a two-down, one-up tracking algorithm, estimating 71% correct responses (Levitt, 1971). There were three levels of linguistic context: isolated vowel, syllable, and sentence. For isolated vowels and syllables, the standard isolated vowel or syllable was presented in the first interval, followed by a standard and a test vowel/syllable randomly ordered in the second and third intervals. The listener's task was to indicate which of the two test intervals contained the different vowel/syllable. For sentences, the standard syllable was presented alone in the first interval followed by two test intervals, each containing the same sentence frame selected from the three sentence frames (Table II). The standard and test syllables were imbedded in the sentence, and listener's task was to detect which of the two intervals contained the test syllable. For the sentence discrimination plus identification task, listeners were asked to identify which one of the four target words contained the test vowel that they heard and which one of the three sentences was presented, and then to indicate which test interval contained the vowel with the formant shift. In this complex task, listeners had to remember the sentence, target word, and test interval, then circle their target word (e.g., bid) and sentence (e.g., the first word is /bVd/ on the page again) responses on a paper form, and then complete the discrimination task using the computer keyboard (e.g., the third interval).

Following the terminology in Watson's research (Watson, 1987), level of stimulus uncertainty can be categorized into four groups: low, medium, high, and very high. Results from Kewley-Port (2001) indicated that formant thresholds were not significantly different for medium levels of stimulus uncertainty ranging from 8 to 22 stimuli per block. Thus, in order to simulate ordinary listening conditions, the level of stimulus uncertainty in the present experiment was medium, i.e., all eight formants were randomly presented in each block for the isolated vowel and syllable tasks. For the sentence tasks, the 24 test conditions (two formants for the four vowels in three positions) were split equally into two groups such that trials within one block presented 12 conditions. There were 96 trials in each block for all conditions. All conditions were presented in a sequence randomized daily for each listener. The duration of each block was approximately 7 min for isolated vowels and syllable, and 14 min for sentences. Each daily session lasted 1.5–2 h. After extensive training, a threshold ΔF in hertz for each listener was averaged from the mean reversals over the last four blocks in which performance was judged as stable by visual inspection. Because the three sentences had quite similar meaning and structure, an average threshold was calculated over the three positions for the sentence discrimination conditions, with and without identification.

III. RESULTS

A. Formant frequency, linguistic context, and signal level

Average thresholds across listeners for the three linguistic contexts and two signal levels are shown in Fig. 2. Overall, thresholds increased as the formant frequency increased in the F2 region. A three-factor (formant frequency × linguistic context × signal level) repeated-measures analysis of variance (ANOVA) with ΔF as the dependent variable showed that thresholds for formant discrimination were affected significantly by formant frequency $[F(7,28)=15.404, p<0.001]$ and linguistic context $[F(2,8)=4.908, p<0.05]$. The effect of signal level was not significant $[F(1,4)=4.657, p=0.097]$. None of the two-way and three-way interaction effects was significant ($p>0.05$), suggesting that threshold-frequency functions showed similar patterns across linguistic context and signal level.

In order to measure the simple main effect of linguistic context under each signal level, two-way (formant frequency × linguistic context) repeated-measures ANOVAs were completed for 70 and 95 dB SPL. Formant frequency had significant effects on ΔF for both 70 and 95 dB SPL ($p<0.05$). For the speech level at 70 dB SPL, there was no significant effect of linguistic context $[F(2,8)=2.160, p>0.05]$, whereas for the speech level at 95 dB SPL, linguistic context significantly affected formant discrimination $[F(2,8)=7.152, p<0.05]$. A post hoc Tukey test at the 95 dB SPL suggested that there was a significant difference in ΔF between the syllable and sentence contexts ($p<0.05$), while no significant difference was obtained between the isolated vowel and either syllable or sentence contexts. Although significantly better performance for isolated vowels than for syllables was reported for normal-hearing listeners in our previous studies (Liu and Kewley-Port, 2004a, b), there was no significant difference in thresholds between isolated vowels and syllables for the hearing-impaired listeners as illus-

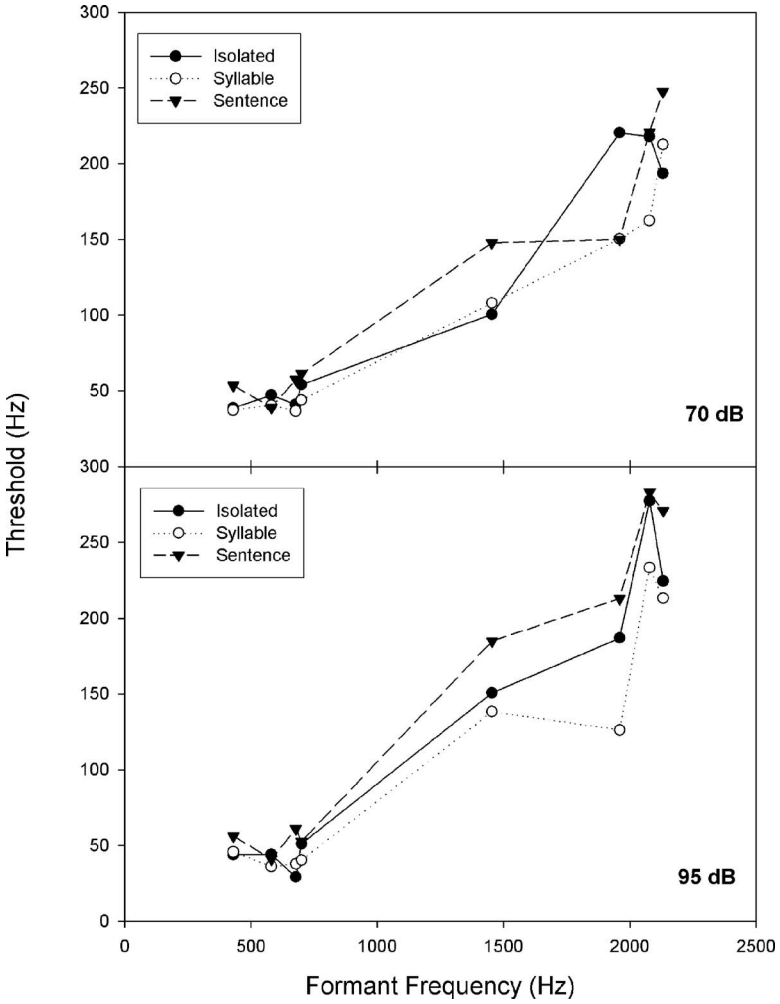


FIG. 2. Thresholds for discrimination of F1 and F2 in hertz as a function of formant frequency for three linguistic contexts: isolated vowel (solid line), syllable (dotted line), and sentence (dashed line) at two levels, 70 (top panel) and 95 (bottom panel) dB SPL.

trated in Fig. 3, even though there were slightly higher thresholds for isolated vowels than for syllables at both 70 and 95 dB SPL. Overall results suggest that linguistic context has a greater effect on formant discrimination at the high signal level than at the low signal level.

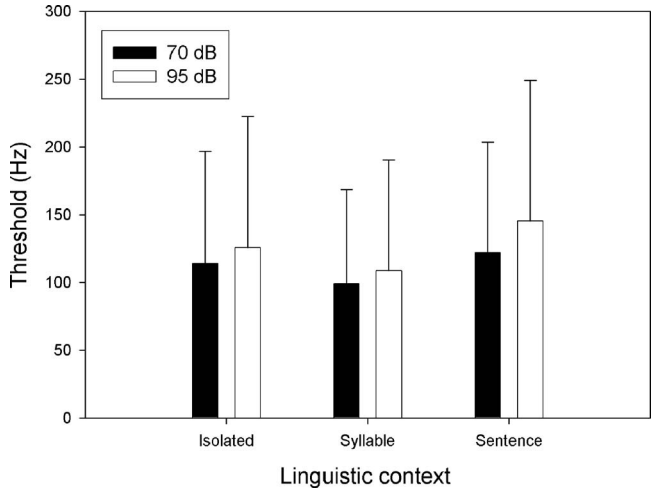


FIG. 3. Thresholds for discrimination of F1 and F2 in hertz as a function of the linguistic context: isolated vowel, syllable, and sentence, averaged across formants at two speech levels, 70 (black) and 95 (white) dB SPL. Error bars display standard deviations.

The above-mentioned three-way ANOVA results suggested a nonsignificant effect of signal level (p=0.097). Overall, in Fig. 3, given the same linguistic context, average thresholds for 70 dB were unexpectedly lower than thresholds for 95 dB SPL, based on better performance at higher levels for the two previous studies (Coughlin *et al.*, 1998; Richie *et al.*, 2003). In particular, formant thresholds on average increased by 19% from 70 to 95 dB SPL for sentence context, while they increased by only 10% and 11% for isolated vowel and syllable context, respectively. This indicates that signal level may have greater effects on performance of formant discrimination for the longer linguistic context, than for the shorter linguistic contexts.

B. Discrimination with identification task

The discrimination with identification task required listeners to listen to the three stimuli, indicate on paper the target word and sentence frame heard, and then to indicate which stimulus had a different formant using a keyboard response. The purpose of intentionally making this a complex auditory task was to increase cognitive load, which refers to "the total amount of mental activity imposed on working memory at an instance at time" (Cooper, 1998). Compared to formant discrimination in sentence only, an additional identification task required a greater amount of men-

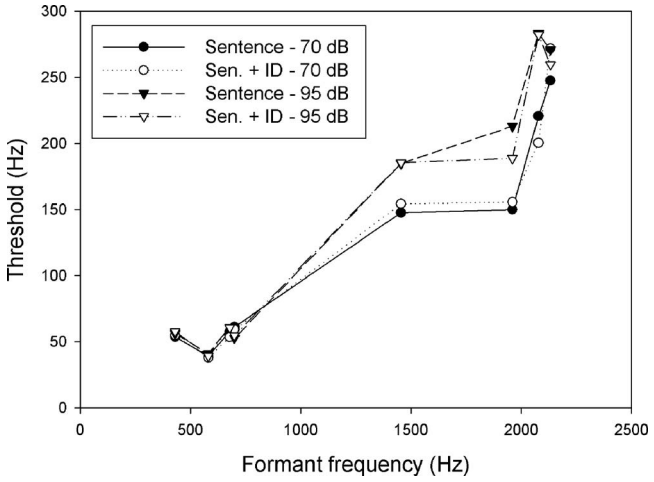


FIG. 4. Thresholds for discrimination of F1 and F2 in hertz as a function of formant frequency for sentences with and without the additional identification task at two speech levels, 70 and 95 dB SPL.

tal activity in working memory, i.e., memorization of the target word and carrier sentence. The resulting formant discrimination functions were, however, essentially similar for the sentence task with and without the additional identification task for both 70 and 95 dB SPL, as shown in Fig. 4. A three-way (formant frequency × with and without ID × signal level) repeated-measure ANOVA showed that formant frequency [$F(7,28)=21.565$, $p<0.001$] had a significant effect on formant thresholds, whereas there were no significant effects of signal level [$F(1,4)=5.531$, $p=0.078$] and the added identification task [$F(1,4)=0.813$, $p>0.05$]. In addition, no significant effects of any two-way and three-way interactions were found ($p>0.05$), suggesting similar patterns for the discrimination functions across the identification task and signal level (see Fig. 4). For the sentence task with and without the identification task, formant discrimination was better for 70 dB SPL than for 95 dB SPL, mainly in the F2 region (24% lower thresholds at 70 dB SPL).

IV. DISCUSSION

A. Factors affecting formant discrimination for hearing-impaired listeners

1. Formant frequency

As shown in Fig. 2, thresholds increased with the increasing formant frequency, mostly in the F2 region for hearing-impaired listeners. Figure 2 shows that thresholds were relatively constant in the F1 region, while they increased linearly at the F2 region. This is consistent with Kewley-Port and Watson's (1994) finding that a piecewise-linear model fit well to formant thresholds, ΔF, for isolated vowels for normal-hearing listeners. This same regression model with the break point at 800 Hz was applied to the present data. For the speech level at 70 dB SPL, because there was no significant effect of linguistic context, average thresholds over the three linguistic contexts were used in the regression analyses. There was no significant linear relationship between ΔF and formant frequency in the F1 region ($p>0.05$), where thresholds were nearly constant at 50 Hz. However, thresholds increased linearly with formant frequency ($p<0.05$) at a rate of 109 Hz/1000 Hz at F2 region above 800 Hz. For the speech level at 95 dB SPL, because thresholds for sentences were significantly higher than thresholds for syllables, regression analyses were conducted for sentence (long linguistic context), thresholds and average thresholds of isolated vowels and syllables (short linguistic context) respectively. In the F1 region, results at 95 dB SPL showed no significant linear relationship between ΔF and formant frequency with an average ΔF at 41.0 and 52.7 Hz ($p>0.05$) for short and long linguistic contexts, respectively. On the other hand, thresholds increased linearly with formant frequencies at F2 region above 800 Hz at a rate of 123 Hz/1000 Hz ($p<0.05$) for the short linguistic context and at a rate of 149 Hz/1000 Hz ($p<0.05$) for the long linguistic context. Overall, the threshold-frequency functions for hearing-impaired listeners were fit well to the piecewise-linear model for different linguistic contexts and signal levels.

2. Linguistic context

As shown in Fig. 3, thresholds of formant discrimination were poorest for the sentence context, the best for the syllable context, with the isolated vowel in between. This pattern of the context effects for hearing-impaired listeners is different from normal-hearing listeners in Liu and Kewley-Port (2004a,b), who used the same stimuli as the present study. Comparing across their two studies, the best performance for formant discrimination was in isolated vowels (Liu and Kewley-Port (2004b), and the poorest performance was in sentences, while syllable performance was in between (Liu and Kewley-Port 2004a). As shown in Table I, the syllable duration for each vowel was almost 50% longer than duration for the isolated vowel. The reversal effect of syllable context for hearing-impaired compared to normal-hearing listeners might be accounted for by two possibilities. One explanation is that, for the formant discrimination task, hearing-impaired listeners may rely on the integration of static and dynamic cues while normal-hearing listeners primarily rely on the static cues for the formant discrimination task. In other words, perceptual weighting was mainly on the steady state of vowels for normal hearing listeners, while both the steady and dynamic state cues were used for hearing-impaired listeners. Another possibility is that hearing-impaired listeners with high-frequency hearing loss may benefit from the low frequency of F2 during the onset formant transition (more than 20 ms for all the four vowels) of the /bVd/ syllable used in this study. As the linguistic context was increased to sentences, thresholds for formant discrimination were generally elevated for hearing-impaired listeners, similar to normal hearing listeners (Kewley-Port and Zheng, 1999; Liu and Kewley-Port, 2004a), although significantly elevated only at the 95 dB SPL level.

3. Signal level

Figure 3 shows that, given the same linguistic context, thresholds for formant discrimination were higher for 95 dB SPL than for 70 dB SPL. An additional analysis separating F1 and F2 suggested that formant thresholds in the F1 region

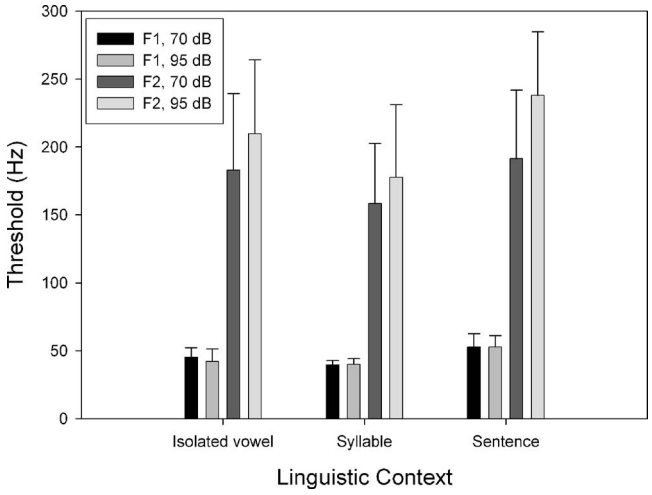


FIG. 5. Comparison of average ΔF for F1 and F2 between 70 and 95 dB SPL in the three linguistic contexts: isolated vowel, syllable, and sentence. Error bars display standard deviations.

were nearly the same, i.e., 1.6% change, between 70 and 95 dB SPL, while thresholds in the F2 region for 95 dB SPL were 18% higher than for 70 dB SPL over the three linguistic contexts (see Fig. 5). This reverse level effect is different from the two previous studies of formant discrimination and several other studies of speech perception for hearing-impaired listeners. Several studies (Ching *et al.*, 1998; Studebaker *et al.*, 1999; Turner and Cummings, 1999) reported that speech recognition was improved with the intensity levels up to almost 100 dB SPL for listeners with mild/moderate hearing loss (the same with our listeners), although for listeners with severe hearing loss there was a rollover level effect at high intensity levels. The difference in the level effects between the present study and other studies could be due to the difference in auditory tasks (discrimination versus identification) and audiogram configuration. Several factors that may account for the reverse level effect in the present study are examined in the following, including audibility, frequency selectivity, upward spread of masking, and forward and backward masking of vowel formants in sentence context.

In order to examine audibility of the vowels for the hearing-impaired listeners, excitation patterns appropriate for hearing-impaired listeners (Moore and Glasberg, 2004) were calculated for the four vowels at 70 and 95 dB SPL for each listener. The F1 and F2 peaks in the excitation patterns of the four vowels at 70 and 95 dB SPL were well above the audiogram for all four vowels for most hearing-impaired listeners, indicating that most of the formants tested were audible (e.g., see Fig. 6 that displays excitation patterns for the /I/ vowel and audiogram for listener S2 in the present study in the upper panel). F2 for /I/ is clearly audible at 95 dB SPL and therefore poor audibility is not the source of the reverse level effect. Consider that Coughlin *et al.* (1998) and Richie *et al.* (2003) reported that hearing-impaired listeners performed better in formant discrimination at the high intensity level (95 dB SPL) than at the low intensity level (70 dB SPL for the study of Coughlin *et al.* and 60 dB SPL for the study

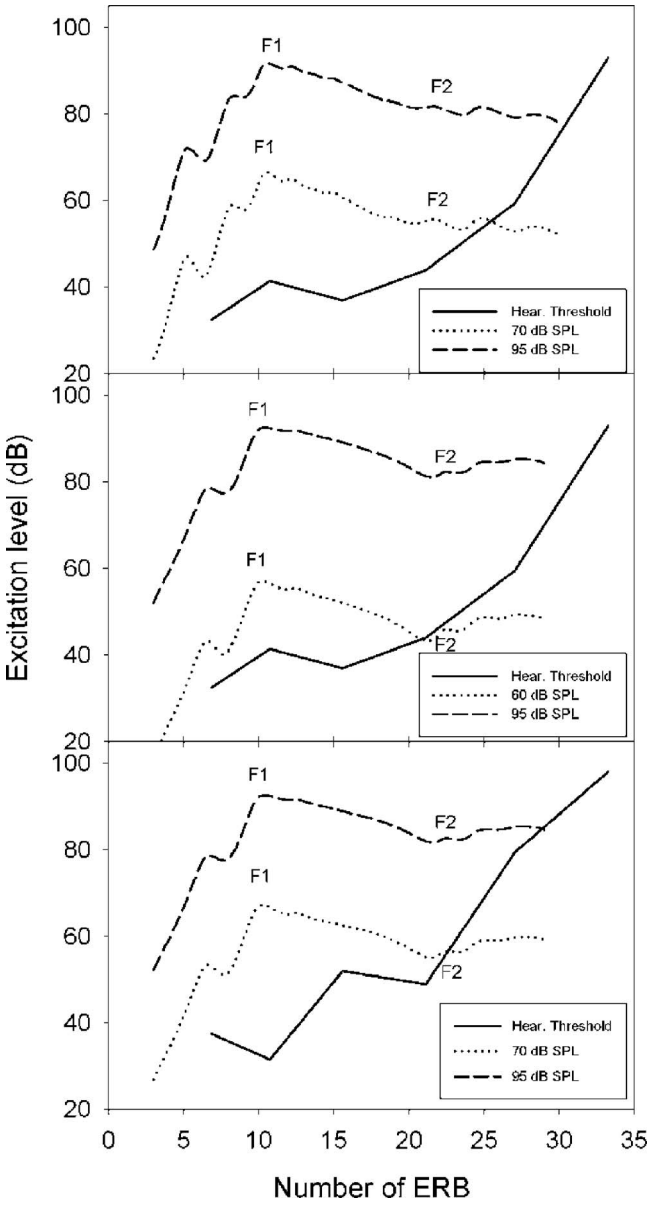


FIG. 6. Excitation patterns of the /I/ vowel at the low (60/70 dB SPL, dotted line) and high (95 dB SPL, dashed line) levels in the present study (top), the studies of Richie *et al.* (2003, middle) and Coughlin *et al.* (1998; bottom) for one listener and the audiogram of this listener of each study (solid line). The listener S2 in this study (top) was the same listener in the study of Richie *et al.* (middle), while the listener in the study of Coughlin *et al.* showed a similar audiogram with the listener of the other two studies. The excitation pattern can be considered as the distribution of excitation level on the ordinate as a function of characteristic frequency on the abscissa (Moore and Glosberg, 1983). Note our low-pass filter cutoff of 5000 Hz is equal to 33.6 ERB.

of Richie *et al.*) as expected from audibility. In fact, excitation patterns, calculated for each vowel for each listener used in their studies, suggested that F2 was inaudible or partially audible at low levels, while fully audible at 95 dB SPL. For example, observe evidence for poor audibility of F2 at low levels in the middle panel of Fig. 6 for listener S2 who participated in the study of Richie *et al.* (2003) and in the lower panel of Fig. 6 for a listener from Coughlin *et al.* (1998) with a similar audiogram to S2. Thus, although in the two previous studies, audibility appeared to be a major factor degrad-

ing discrimination performance at low signal levels, this explanation fails to account for the observed reverse level effect here.

A possible mechanism that may explain the reverse level effect for hearing-impaired listeners is reduced frequency selectivity at 95 dB SPL compared to 70 dB SPL. Auditory filters broadened somewhat with an increase of signal level (Moore and Glasberg, 1997; Baker and Rosen, 2002). As shown in Fig. 6, peak-to-valley contrasts of F1 and F2 in the excitation patterns for the four vowels in the present study were slightly reduced (less than 1 dB on average) from 70 to 95 dB SPL for hearing-impaired listeners. Apparently a reduction in peak-to-valley contrasts of F1 and F2 as signal level increased may play a limited role in the increased formant thresholds at 95 dB SPL.

Another possibility accounting for the higher formant thresholds at 95 dB SPL could be the upward spread of masking produced by F1. As noted earlier, the level effect was found primarily for F2, but not for F1. As signal level is increased, more upward spread of masking on F2 was observed for hearing-impaired listeners (Leek and Summers, 1993; Zwicker and Jaroszowski, 1982). Summers and Leek (1997) found that CV identification increased with F1 attenuation up to 18 dB for both normal-hearing and hearing-impaired listeners. Thus, upward spread of masking caused by F1 might increase the difficulty of discriminating vowel formant frequency in F2 at the high presentation level in the present study.

As noted in Sec. III, for the sentence context, the increase in ΔF with the level increase was markedly higher than for isolated vowels and syllables. This suggests that, when target vowels are presented in sentences, some other mechanism may also contribute to poor formant discrimination at high signal levels. One possibility may be that formant discrimination in sentences is affected by forward and backward masking resulting from the speech preceding and following the target vowel. Studies of intensity discrimination for pure tones under forward masking suggested a nonmonotonic function from 20 to 100 dB SPL for normal-hearing listeners (Zeng *et al.*, 1991; Zeng and Turner, 1992) and hearing-impaired listeners (Zeng and Shannon, 1995). In addition, given that vowel recognition was significantly degraded by backward masking according to Dorman *et al.* (1977), backward masking may also affect formant discrimination in sentences in the present study.

Summarizing, elevation of formant discrimination thresholds at high stimulus levels in the present study may be accounted for by two mechanisms, depending on linguistic context. When the linguistic context is short, reduced formant discrimination at higher levels is likely due to the poorer representation of high-frequency formants and greater upward spread of masking on F2. When the target vowel is embedded into a long linguistic context, the forward and backward masking resulting from the preceding and following speech may additionally degrade formant discrimination.

4. Discrimination with identification task

No significant effects of the additional identification task were found on the performance of vowel formant discrimination at either the 70 or 95 dB SPL level, consistent with previous studies for normal-hearing listeners (Kewley-Port and Zheng, 1999; Liu and Kewley-Port, 2004a). Although cognitive load in working memory was increased to process the responses for both tasks, cognitive processing for sentence identification did not affect the processing of the spectrotemporal information necessary for the discrimination task. Our results are in some agreement with a related study by Gordon-Salant and Fitzgibbons (1997). They reported the results of two recall tasks when the availability of context cues and different speech rates were manipulated for young and elderly listeners with either normal or impaired hearing. Results showed that recall performance was better for normal-hearing than for hearing-impaired listeners for both age groups. The authors suggested that speech recognition by elderly listeners with hearing loss was dependent on both auditory and cognitive factors. However, similar to our results with younger listeners, effects of hearing impairment were observed without additional effects of cognitive load in their speech recognition tasks. Taken together, the results support a conclusion that cognitive processing for sentence identification is essentially independent from processing necessary for formant discrimination for both younger hearing-impaired listeners and normal-hearing listeners.

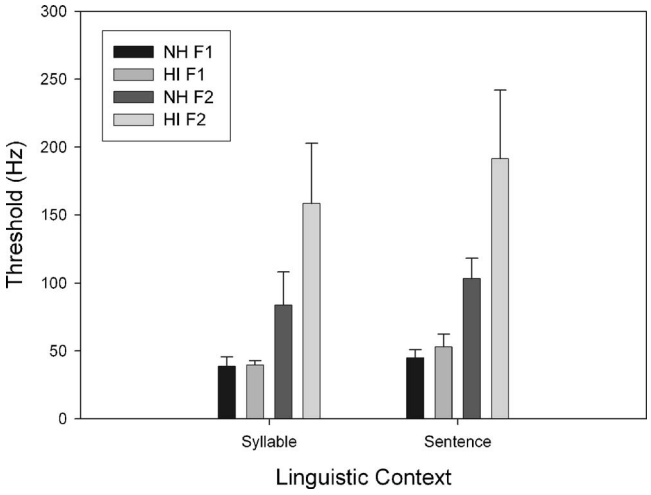


FIG. 7. Comparison of average ΔF for F1 and F2 for hearing-impaired and normal-hearing listeners at 70 dB SPL in the two linguistic contexts: syllable, and sentence. The data for normal-hearing listeners are from Liu and Kewley-Port's study (2004a). Error bars display standard deviations.

B. Comparison between normal-hearing and hearing-impaired listeners

The syllable and sentence tasks in the present study were identical to those presented to young normal-hearing listeners in Liu and Kewley-Port (2004a) at the 70 dB SPL signal level. The formant thresholds for those normal-hearing and our hearing-impaired listeners are shown in Fig. 7. A three-way ANOVA (formant frequency × linguistic context × hearing group) showed significantly higher thresholds for the hearing-impaired group than for the normal-hearing group $[F(1,112)=20.463, p<0.05]$. Simple main effects of hearing status suggested that, under each linguistic context

(syllable and sentence), hearing-impaired listeners had significant higher thresholds than normal-hearing listeners ($p < 0.05$).

Because threshold differences appeared to occur primarily for F2 (Fig. 7), separate analysis of the F1 and F2 data were conducted. Three-way ANOVAs on F1 and F2 (formant frequency × linguistic context × hearing group) suggested that no significant effect of hearing status was found for the F1 region [$F(1,56)=0.648, p>0.05$], while formant thresholds were significantly higher for hearing-impaired listeners for the F2 region [$F(1,56)=20.094, p<0.05$]. This is consistent with the study of Coughlin *et al.* (1998) that reported similar thresholds for F1 but different thresholds for F2 between young normal-hearing and elderly hearing-impaired listeners for isolated synthetic vowels. The present study extends their results to vowels in syllables and sentences to demonstrate that mild-to-moderate sloping hearing impairment had no effect on the auditory processing of F1 at a conversational level of 70 dB SPL that was used in both studies. These results are commensurate with the physiological findings that the internal representation of vowels as a result of sloping sensorineural hearing loss is relatively unchanged for F1, but degraded for F2 (Javal, 1997; Miller *et al.*, 1997).

C. General discussion and conclusions

The aim of the present study was to investigate how two factors, hearing loss and cognitive load, influenced vowel formant discrimination for adult listeners. Thresholds of formant discrimination for syllables and sentences were significantly elevated for hearing-impaired listeners compared to thresholds for young normal-hearing listeners (Fig. 7). However, formant discrimination was elevated in the F2 region by almost 100%, where the greater hearing loss occurred, rather than in the F1 region. When cognitive load was increased, i.e., when a complex sentence identification task was added to the formant discrimination task, results were similar to previous results from normal-hearing listeners (Liu and Kewley-Port, 2004a), suggesting that the additional sentence identification task had little effect on formant discrimination. Thus, higher thresholds for formant discrimination by hearing-impaired listeners for sentences were not due to the difficulty of cognitive task, but rather due to the hearing loss.

The present results showed that high signal levels for speech signals degraded thresholds for formant discrimination for hearing-impaired listeners rather than improved performance when audibility was assured. Several factors were considered to account for the level effect on formant discrimination, including audibility, frequency selectivity, and upward spread of masking on F2. All these factors may interact with each other to affect formant discrimination. In the present study, decreased frequency selectivity and greater upward spread of masking on F2 at the high signal level may have contributed to the reverse level effect of formant discrimination. Furthermore, when the target vowel was embedded in a sentence context, forward and backward masking caused by the preceding and following speech may have also degraded formant discrimination at the high speech level.

In conclusion, adult listeners with hearing loss had higher thresholds for the same linguistic context and the same speech level, primarily for the F2 region, compared to normal-hearing listeners. This study also suggests that degraded vowel formant discrimination in sentences by hearing-impaired listeners was not due to an increased cognitive load, but was caused primarily by auditory factors associated with hearing loss.

ACKNOWLEDGMENTS

This research was supported by the National Institute of Health Grant No. DC-02229 to Indiana University. Zach Burkle is gratefully acknowledged for help in running experiments in the study. The authors are also thankful to two anonymous reviewers and Kenneth Grant for their constructive comments on earlier versions of this manuscript.

Baker, R. J., and Rosen, S. (**2002**). "Auditory filter nonlinearity in mild/moderate hearing impairment," J. Acoust. Soc. Am. **111**, 1330–1339.

Ching, T. Y. C., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Cooper, G. (**1998**). "Research into cognitive load theory and instructional design at UNSW," http://education.arts.unsw.edu.au/CLT_NET_Aug_97.HTML, last viewed on 16 April, 2007.

Coughlin, M., Kewley-Port, D., and Humes, L. E. (**1998**). "The relation between identification and discrimination of vowels in young and elderly listeners," J. Acoust. Soc. Am. **104**, 3597–3607.

Dorman, M. F., Kewley-Port, D., Brady, S., and Turvey, M. T. (**1977**). "Vowel recognition: Inferences from studies of forward and backward masking," Q. J. Exp. Psychol. **29**, 483–497.

Gordon-Salant, S., and Fitzgibbons, P. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Hedrick, M., and Jesteadt, W. (**1997**). "Influence of relative amplitude and presentation level on perception of the /p/-/t/ consonant contrast by normal and impaired listeners," *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Lawrence Erlbaum, Mahwah, NJ), pp. 475–486.

Humes, L. E., Dirks, D. D., Bell, T. S., and Kincaid, G. E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765–773.

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.

Javal, E. (**1997**). "Cochlear excitation patterns in sensorineural hearing loss," *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Lawrence Erlbaum, Mahwah, NJ), pp. 9–34.

Kawahara, H., Masuda-Kastuse, I., and Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.

Kewley-Port, D. (**2001**). "Vowel formant discrimination: Effects of stimulus uncertainty, consonantal context and training," J. Acoust. Soc. Am. **110**, 2141–2155.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1999**). "Vowel formant discrimination: Towards more ordinary listening conditions," J. Acoust. Soc. Am. **106**, 2945–2958.

Leek, M. R., and Summers, V. (**1993**). "Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise," J. Acoust. Soc. Am. **97**, 3127–3137.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Liu, C., and Kewley-Port, D. (**2004a**). "Vowel formant discrimination for high-fidelity speech," J. Acoust. Soc. Am. **116**, 1224–1233.

Liu, C., and Kewley-Port, D. (**2004b**). "Formant discrimination in noise for isolated vowels," J. Acoust. Soc. Am. **116**, 3119–3129.

Liu, C., and Kewley-Port, D. (**2004c**). "STRAIGHT: A new speech synthe-

sizer for vowel formant discrimination," ARLO **5**, 31–36.
McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (**2005**). "Hearing loss and perceptual effort: Down-stream effects on older adults' memory for speech," Q. J. Exp. Psychol. **584**, 22–33.
Miller, R. L., Schilling, J. R., Young, E. R., and Franck, K. R. (**1997**). "Representation of the vowel /ϵ/ in the auditory nerve of cats with a noise-induced hearing loss," *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Lawrence Erlbaum, Mahwah, NJ), pp. 35–48.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C. J., and Glasberg, B. R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Aud. Neurosci. **3**, 289–311.
Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.
Needleman, A. R., and Crandell, C. C. (**1997**). "Speech perception in noise by listeners with hearing impairment and simulated sensorineural hearing loss," *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Lawrence Erlbaum, Mahwah, NJ), pp. 461–473.
Rakerd, B., Seitz, P. F., and Whearty, M. (**1996**). "Assessing the cognitive demands of speech listening for people with hearing loss," Ear Hear. **17**, 97–106.
Richie, C., Kewley-Port, D., and Coughlin, M. (**2003**). "Discrimination and identification of vowels for hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **114**, 2923–2933.
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.
Summers, V., and Leek, M. (**1997**). "Intraspeech spread of masking in normal-hearing and hearing-impaired listener," J. Acoust. Soc. Am. **101**, 2866–2876.
Turner, C. W., and Cummings, K. J. (**1999**). "Speech audibility for listeners with high-frequency hearing loss," American J. of Audiology **8**, 47–56.
Watson, C. S. (**1987**). "Uncertainty, informational masking and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.
Wingfield, A., Tun, P. A., and McCoy, S. L. (**2005**). "Hearing loss in older adulthood: What it is and how it interacts with cognitive performance," Curr. Dir. Psychol. Sci. **14**, 144–148.
Zeng, F.-G., and Shannon, R. (**1995**). "Possible origins of the non-monotonic intensity discrimination function in forward masking," Hear. Res. **82**, 216–224.
Zeng, F.-G., and Turner, C. W. (**1992**). "Intensity discrimination in forward masking," J. Acoust. Soc. Am. **92**, 782–787.
Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). "Recovery from prior stimulation. II. Contribution to intensity discrimination," Hear. Res. **55**, 223–230.
Zwicker, E., and Jaroszowski, A. (**1982**). "Inverse frequency dependence of simultaneous tone-to-tone masking patterns at low levels," J. Acoust. Soc. Am. **71**, 1508–1512.

A detailed study on the effects of noise on speech intelligibility[a)]

Finn Dubbelboer[b)] and Tammo Houtgast
Department of Otolaryngology, VU Medical Center, P.O. Box 7057, 1007 MB, Amsterdam, The Netherlands

(Received 16 May 2007; revised 13 August 2007; accepted 18 August 2007)

A wavelet representation of speech was used to display the instantaneous amplitude and phase within 1/4 octave frequency bands, representing the envelope and the carrier within each band. Adding stationary noise alters the wavelet pattern, which can be understood as a combination of three simultaneously occurring subeffects: two effects on the wavelet levels (one systematic and one stochastic) and one effect on the wavelet phases. Specific types of signal processing were applied to speech, which allowed each effect to be either included or excluded. The impact of each effect (and of combinations) on speech intelligibility was measured with CVC's. It appeared that the systematic level effect (i.e., the increase of each speech wavelet intensity with the mean noise intensity) has the most degrading effect on speech intelligibility, which is in accordance with measures such as the modulation transfer function and the speech transmission index. However, also the introduction of stochastic level fluctuations and disturbance of the carrier phase seriously contribute to reduced intelligibility in noise. It is argued that these stochastic effects are responsible for the limited success of spectral subtraction as a means to improve speech intelligibility. Results can provide clues for effective noise suppression with respect to intelligibility. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2783131]

PACS number(s): 43.72.Dv, 43.71.Es, 43.66.Ts [DOS] Pages: 2865–2871

I. INTRODUCTION

When noise is added to speech, the speech signal is altered by the stochastic processes involved in the interaction. This paper describes the nature and consequences of these interactions in detail. Sometimes speech processing is used to counteract these alterations, for instance in hearing aids and (mobile) communication devices. When noisy speech is recorded by a single microphone, the noise spectrum can be estimated and subtracted from the speech-plus-noise input, an operation known as spectral subtraction (Lim, 1978; Boll, 1979). Spectral subtraction is one of the first "easy-to-implement" reduction schemes among single-microphone noise-reduction techniques and currently often used, for instance in hearing aids and mobile phones. Various alternative techniques have been investigated since, aiming at estimating essential parameters for restoring the speech envelope. Three fundamental differences among these techniques can be distinguished. First, the type of parameter that is estimated [spectral magnitude, log spectral magnitude, complex valued spectral coefficient—Lee (1960); Ephraim and Malah (1984)]. Second, the way this parameter is estimated (expected value, maximum *a posteriori* criterion) and third, the assumptions that are made concerning the amplitude distributions of speech and noise [Gaussian, Laplacian, gamma, and super-Gaussian—Martin, 2002; Breithaupt and Martin, 2003]. Currently, psychoacoustics plays an increasingly important role during the design process, which leads to "perceptually optimized" algorithms. Although positive results have been reported in terms of listening comfort and fatigue, the overall success of signal restoration is somewhat disappointing. The improved quality of the output signal seldom leads to improved intelligibility (Lim and Oppenheim, 1979; Levitt, 1986; WGCA, 1991), suggesting that the exact nature of the speech-noise interactions and their consequences for speech intelligibility are not fully understood.

In order to improve speech intelligibility in noise one should know (1) how speech is physically changed, (2) which of these changes are most detrimental, and (3) how the most detrimental changes can be counteracted, without introducing new distortions. The idea of speech being affected in several ways by adding noise was recognized earlier in work by Drullman and Noordhoek (see the following). One way of analyzing the properties of speech is to consider the speech signal as a sum of amplitude-modulated carriers in adjacent frequency bands; it is known that these modulations are essential for speech intelligibility. The presence of noise (or reverberation) reduces these modulations and therefore reduces intelligibility. This is the basis of the concept of the modulation transfer function (MTF) and the speech transmission index (STI) (Houtgast and Steeneken, 1985). The STI is a widely used measure (IEC, 2003), for estimating intelligibility in auditoria, working places, public areas, etc. Drullman (1995) found that equal MTFs do not necessarily lead to equal intelligibility. Noordhoek and Drullman (1997) compared the effect of two types of modulation reduction on speech perception. In the first set of stimuli a multichannel compression scheme was applied on the temporal speech envelope (deterministic modulation reduction). In the second set, modulations were reduced by adding noise, which was referred to as stochastic modulation reduction. They found that "pure" modulation reduction—the one effect considered

[a)]Presented at the 149th ASA Meeting, April 2005.
[b)]Electronic mail: finndubbelboer@planet.nl

in the STI—could not fully explain the detrimental effect of added noise. Two possible additional noise effects were suggested. First, nonrelevant modulations arising from the stochastic nature of the noise-speech interaction can be responsible for reducing the perceptual distance between speech and noise. Second, the fine structure is damaged, which may affect possible cues that rely on this fine structure. Noordhoek showed that, in case of adding noise, these additional effects grow proportionally with the effect of modulation reduction. Under normal circumstances, their impact remains relatively small and is implicitly included in the experimentally determined relation between the modulation-reduction based STI and speech intelligibility. However, in case of specific types of noise suppression, for instance spectral subtraction, this relation is disturbed and the STI is no longer a reliable predictor (Ludvigsen, 1993). Apparently, noise reduction algorithms cannot compensate for all three noise effects. In fact, it seems that extra effects are introduced, of which the consequences for speech perception are not clearly understood.

The first part of this paper describes a type of signal analysis that enables us to identify different speech alterations involved with additive noise. Thinking of the speech signal in terms of a sum of amplitude-modulated carriers in adjacent frequency bands, three effects can be distinguished: (1) a systematic lift of the envelope equal to the mean noise intensity, (2) stochastic envelope fluctuations, and (3) the corruption of the fine structure. Subsequently, it will be shown how this type of analysis can be applied to isolate each of these effects and how the perceptual consequences of these three effects were measured in a series of listening experiments. Finally, the results will be discussed in relation to the limited effects of noise suppression on speech intelligibility.

II. SIGNAL PROCESSING

For defining the different effects of noise on the speech signal, and for preparing the stimuli for the listening experiment, we used a type of wavelet transformation (WT) (Strang, 1994; Rioul, 1991). By choosing an appropriate mother wavelet, WT can provide a spectrotemporal representation that roughly corresponds to auditory frequency-time analysis. The quality of the "match" depends on how well the spectrotemporal segmentation or tiling (determined by the shape of the wavelet) is in agreement with auditory frequency-time resolutions. In a number of experiments, van Schijndel (1999) determined parameter settings for an optimal "auditory" mother wavelet. This involved the shape of the temporal envelope and the number of cycles, which determine the effective duration and the effective spectral bandwidth of a wavelet, and together the spectrotemporal resolution. This resolution approaches a theoretical limit—dictated by the uncertainty principle (Landau and Polak, 1961)—for Gaussian-shaped envelopes. An appealing property of a Gaussian-shaped wavelet is its symmetry in frequency and time, which is an advantage from the signal processing point of view. Although it does not strictly correspond to auditory filtering, it can be considered a first-order approximation of the auditory filter, which is assumed to be Gammatone shaped (Patterson *et al.*, 1992). A less appealing property of a Gaussian shape is the fact that, when applying wavelet analysis and resynthesis, the reconstructed signal is not identical to the original signal: A Gaussian envelope causes imperfections during inverse transformation. This effect can be counteracted by increasing the amount of overlap between subsequent wavelets, both in time and frequency. This also improves the robustness of the analysis-resynthesis scheme when modifications are involved, as described in Sec. II B. The Gaussian mother wavelet is described by

$$s(t) = \sqrt{\alpha f_0}\exp(i2\pi f_0 t)\exp(\pi(\alpha f_0 t)^2), \quad (1)$$

in which f_0 is the carrier frequency, α is the shape factor, and $\sqrt{\alpha f_0}$ normalizes the energy of the analysis function. The wavelet has an effective bandwidth of $\Delta_f = \alpha f_0$ and an effective duration of $\Delta_t = 1/\alpha f_0$ (van Schijndel, 1999). The effective bandwidth of the analysis function was set to 1/4 octave, roughly corresponding to the critical bandwidth of the auditory system (Florentine *et al.*, 1988). This corresponds to a shape factor of $\alpha = 0.1735$. As a result, the effective duration of the frequency-time window is 5.76 ms at 1 kHz (1.44 ms at 4 kHz). The effective number of periods contained within the Gaussian envelope equals $5.8 (= 1/\alpha)$. The overlap between wavelets in time was set to one wavelet every three periods of the carrier frequency and eight wavelets per octave along the frequency axis. This implies 33 spectral output channels with f_o varying from 250 to 4000 Hz and a total of approximately 16×10^3 wavelet coefficients per second. An overlap-add procedure was used for synthesis back into time domain (Sec. II B): Each wavelet was multiplied by a wavelet coefficient, corresponding to the proper amplitude and phase. The quality of the output of the described analysis-synthesis scheme has previously been evaluated in a listening experiment in which pre- and postprocessed speech were compared (van Schijndel, 1999). Results indicated that processing-related artifacts in the output signal were imperceptible.

A. Signal analysis: The three noise effects

Since the aim of this study is to investigate the perceptual consequences of alterations on the speech signal brought about by adding noise, we need to determine: (1) the physical nature of the alterations and (2) the effect of each of these alterations on speech intelligibility. Wavelet transformation and its inverse were used for both purposes.

After applying wavelet transformation, an input signal is represented by a number of wavelet coefficients, i.e., signal energy within a spectral band integrated over a few milliseconds, which will be referred to as "pixels" throughout the paper. First, speech was subjected to WT yielding a number of bandpass filtered envelopes, each consisting of a row of pixels (Fig. 1). This was done for clean speech and for an identical version of speech corrupted with stationary speech-shaped noise. The first set of pixels can be considered as the input of a noise corruptive system, the second set as the output of such system. By plotting the input pixels versus the

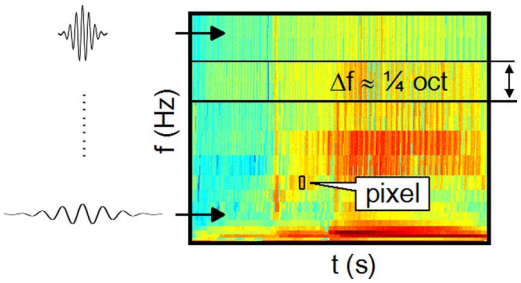


FIG. 1. (Color online) A frequency-time representation of a speech token after wavelet transform. The output can be considered as a number of $\frac{1}{4}$-octave bandpass-filtered envelopes, each represented by a row of pixels. By compressing or expanding the mother wavelet, the entire spectrum can be analyzed. Note that toward higher frequencies both the bandwidth (in hertz) and the number of pixels per time unit increase.

output pixels the effect of noise on speech is captured in detail from pixel to pixel. Figure 2 shows an input/output diagram of pixel levels within one frequency band, 1/4-octave around 1000 Hz. For the other bands, the input/output statistics are essentially similar, dictated by the signal-to-noise (S/N) ratio. For clarity, we choose a +10 dB speech-to-noise ratio in the picture to illustrate the method. Pixel levels are depicted relative to the rms of the clean-speech pixels within the given 1/4-octave band. Figure 2 shows that high-level speech pixels are little or not affected. However, the influence of noise increases toward lower speech levels until the output is fully determined by noise. This effect is illustrated by the black curve, the effect of a systematic lift of speech-pixel levels, equal to the addition of the mean noise intensity $\overline{I_N}$. This is identified as the first noise effect. The pixel cloud around this curve represents the second noise effect: random intensity fluctuations I_{Nr}, reflecting the stochastic nature of noise and of the speech-noise interaction. The position of each output pixel is a result of a combination of speech- and noise-pixel-intensities and underlying phase interactions. These interactions are dominated either by

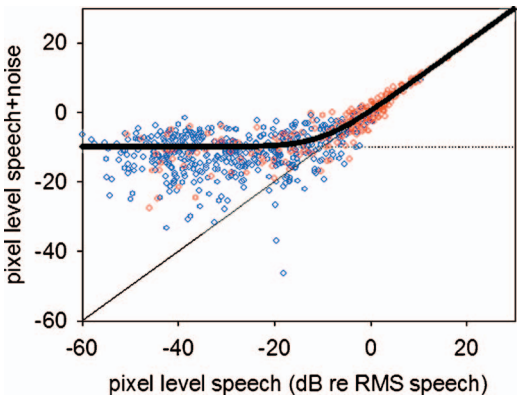


FIG. 2. The input/output relation given (i.e., pixel level speech vs pixel level speech+noise) illustrates how levels and phases of speech pixels within a $\frac{1}{4}$-octave band change after noise is added. Three separate noise effects can be identified: (1) a systematic level effect, i.e., an average increase of each speech pixel with the mean noise intensity, (2) a stochastic level effect caused by random level fluctuations, and (3) a fine-structure effect, represented by changes of the pixel phases. In the picture, the latter effect is illustrated by using color coding: Pixels of which the phase was changed less than ±15° are indicated by the red circles, the remaining pixels are represented by the blue circles.

TABLE I. Combinations of $\overline{I_N}$, I_{Nr}, and φ_{SN} lead to eight different conditions ($\overline{I_N}$: a systematic lift of the speech-envelope equal to the mean noise intensity; I_{Nr}: stochastic envelope fluctuations; φ_{SN}: corruption of the fine-structure). The asterisks indicate the type of modification that was applied to the speech. The condition numbers correspond to the number in the panels of Fig. 3.

Condition	1	2	3	4	5	6	7	8
$\overline{I_N}$		*				*	*	*
I_{Nr}			*		*		*	*
φ_{SN}				*	*	*		*

speech or by noise, dependent on their relative strength. The color coding is used to illustrate the effect on the pixel phase: the third noise effect. The phases of speech-dominated pixels will only be slightly affected. In the picture, pixels of which the phase shift is less than ±15° are indicated by red circles. For the remaining pixels the phases are affected more strongly and are marked by a blue circle. Statistically, a small fraction (about 8%) of the noise-dominated pixels is found to be within the 30° interval around the speech-pixel phase. The ±15° classification threshold is arbitrary and is introduced for illustrative purposes only.

B. Signal analysis and resynthesis for the listening experiments

The second problem addresses the question of how speech intelligibility is affected by each of these effects separately. In the lab, signals can be mixed freely in various S/N ratios while storing the exact copies of the underlying uncorrupted speech and noise files. When speech, noise, and a speech-plus-noise mix are subjected to the same wavelet transformation, pixels from these signals are mutually linked: of each speech-plus-noise pixel in the f-t plane, there exists a corresponding "underlying" speech pixel and noise pixel. Not only pixel levels, also the underlying phases are present. Pixel levels and pixel phases can freely be exchanged among the three files, which leads to new signals after inverse transform. For example, combining the speech-plus-noise phases with the clean-speech pixel levels yields an intact temporal speech envelope with an underlying noise-corrupted fine structure. In doing so systematically, eight different ways of corrupting speech can be realized by combining $\overline{I_N}$, I_{Nr}, and φ_{SN}, the three basic noise effects, as shown in Table I. This includes the uncorrupted speech (condition 1) and the full noise effect (condition 8).

An overview of the physical consequences of the operation is shown in Fig. 3, in a number of input/output diagrams, and will be discussed in the following. At this stage, only signal physics was discussed, no listener was yet involved. The above-described type of signal processing was used as a protocol to compute a large set of specifically corrupted speech. The speech in this set was used as stimuli in a listening experiment with normal hearing subjects.

III. MEASUREMENTS

Four speakers and four normal-hearing listeners participated in a CVC word-score listening experiment. All stimuli

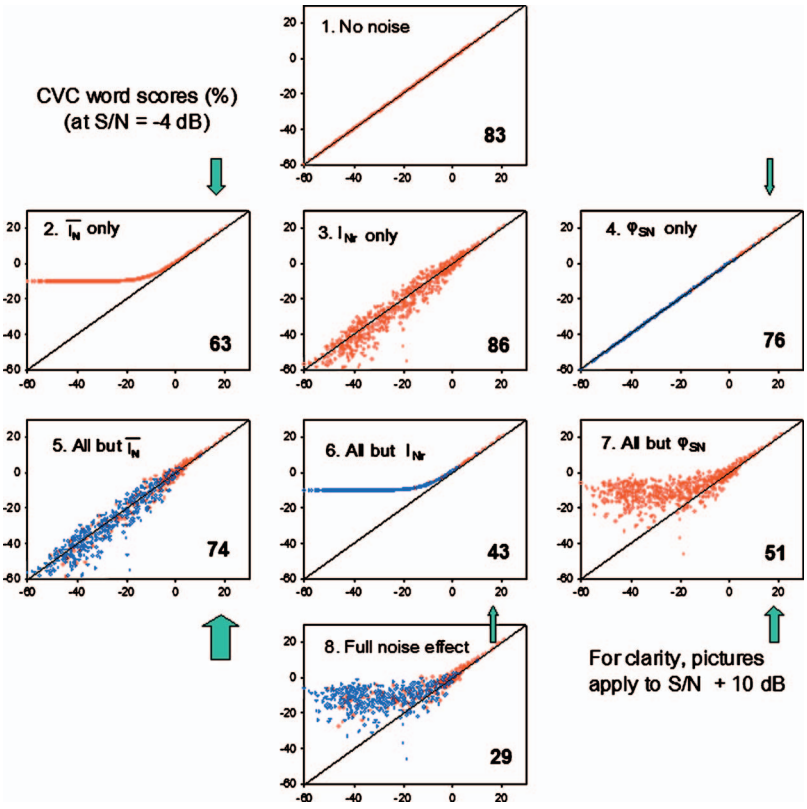


FIG. 3. A visual representation of the eight conditions, according to Table I, displayed in a number of input/output diagrams. The type of noise effect is shown in the upper-left corners, the bottom-right corners show the results of the CVC-test in % word-score. The S/N ratio of the stimuli was set to −4 dB. For illustrative purposes the S/N ratio for the pictures is +10 dB.

were computed in advance and consisted of CVC (consonant-vowel-consonant) words, sampled at 44.1 kHz with a 16-bit resolution. Each condition was measured with one list per speaker, each list containing 50 words. The overall S/N ratio was set to −4 dB, roughly corresponding to the critical S/N ratio for understanding speech in noise for normal hearing listeners (Plomp and Mimpen, 1979; Versfeld *et al.*, 2000). Signals were bandpass filtered, and contained the frequency range between 250 and 4000 Hz.

CVC scores in percent are shown in the right bottom corner of the input/output diagrams in Fig. 3. Numbers in the left upper corner correspond to the combinations of $\overline{I_N}$, I_{Nr} and φ_{SN} given in Table I. The panels are arranged by increasing noise effect, starting from "no noise" (= clean speech) at the top to "the full noise effect" at the bottom. The first row shows the result of imposing one single noise effect. Results from the no noise condition (83%) are considered as reference. Common speech-plus-noise corruption (full noise effect) causes an intelligibility drop to 29%. If we concentrate on the first row, the most detrimental effect is the systematic level increase of the speech envelope, the $\overline{I_N}$ condition, causing a 20% drop to 63%. Second is the corruption of the fine structure, indicated by φ_{SN}, reducing the score to 76%, both results are highly significant ($p<0.01$). The size of the green arrows corresponds to the relative impact of each effect. The effect of random fluctuations I_{Nr} is somewhat puzzling. In fact, the results suggest that adding some randomization to the speech envelope slightly increases intelligibility (from 83% to 86%). This effect is significant ($p<0.05$) and not yet understood. When adding the random fluctuations to any or both of the other noise effects, the effect is always a decrease in score.

In general, adding a second effect to the first (second row) degrades intelligibility, and again $\overline{I_N}$ contributes most as illustrated in panels 6 and 7 compared to 5. Finally, adding a third effect (second row to the bottom panel) leads to the full noise effect and is again dominated by $\overline{I_N}$ (conditions 5–8).

Figure 3 shows that an intensity lift of the speech envelope (equal to the mean noise intensity) is the most detrimental effect: Going from condition 1 to condition 2 causes the largest (single noise-effect) drop in word-score, from 83% to 63%. An alternative way to weigh the relative contribution of each effect is by going from the bottom panel, the full-noise effect, up and compare the result of removing one of these noise effects. The size of the green arrows indicates the effect of removing the corresponding noise effect on intelligibility. Also from this viewpoint, the first noise effect appears to be the most important one. It is interesting to note that this effect is the only effect considered in measures such as the MTF and the STI, in which signal physics is related to speech perception.

IV. DISCUSSION

When considering the effect of additive stationary noise on the speech signal, the second noise effect is often not fully recognized. The general view is that the instantaneous speech intensity is increased by the mean noise intensity, and

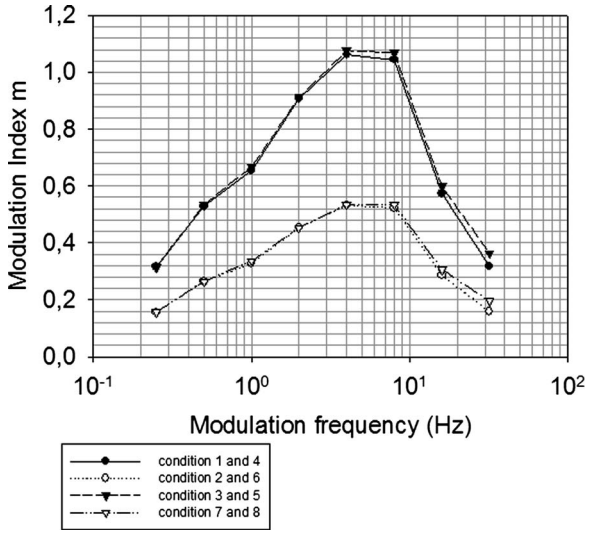


FIG. 4. Two clusters of modulation spectra derived from eight modified envelopes corresponding to the eight conditions illustrated in Fig. 3. The eight conditions fall apart in two groups: a first group of four conditions, including the clean speech, all showing the original speech envelope spectrum, and a second group, including the full noise effect, all showing the same reduced envelope spectra. In or excluding the phase effect results in identical modulation spectra (1–4, 3–5, 2–6, and 7–8).

that the fine structure (the carrier) is corrupted to some extent. The consequences of the introduction of the stochastic level fluctuations, the second noise effect, are not always fully acknowledged. We will discuss this issue with respect to two topics, the MTF-STI concept for predicting speech intelligibility, and the spectral subtraction approach for noise reduction.

A. The MTF and the STI model

The success of models like MTF and STI in predicting intelligibility of noise-corrupted speech is generally recognized. However, these models only take account of the first noise effect, and therefore are based on a simplified image of the actual speech/noise interaction. According to the MTF-STI model, intelligibility reduction is a direct consequence of the extent to which modulations within speech envelopes are reduced. It can easily be shown that these modulation reductions are caused by the first noise effect only (i.e., by the addition of the mean noise intensity), and are not affected by the second and third noise effects. To illustrate this, speech was subjected to exactly the same wavelet analysis and resynthesis as described before. Resulting pixel levels and pixel phases were modified in correspondence to the eight conditions in Table I and Fig. 3. The resulting modified speech envelopes, i.e., arrays of pixel levels, were subjected to a fast Fourier transformation. Normalizing for the mean intensity and integration within 1-octave modulation bands, yielded the eight modulation spectra in the frequency range from 0.25 to 32 Hz. The results are depicted in Fig. 4. The eight conditions fall apart in two groups: A first group of four conditions, including the clean speech, all showing the original speech envelope spectrum, and a second group, including the full noise effect, all showing the same reduced envelope spectra. Note that the two groups only differ by the absence (conditions 1, 3, 4, and 5) or the presence (conditions 2, 6, 7, and 8) of the first noise effect. Within groups, the four conditions with equal envelope spectra would yield equal STIs, and thus equal predicted intelligibility. Figure 3 shows that this is clearly not the case.

Hence, when speech is corrupted by noise (the full noise effect), the intelligibility is reduced as the result of three noise effects, while the modulation-reduction based STI model only accounts for one of these three effects, i.e., the lift of the speech envelope by the mean noise intensity. In normal circumstances, this does not pose a real problem since all three noise effects will depend on the S/N ratio, and are thus highly related. This means that the STI is still uniquely related to intelligibility, as long as the mutual relation among the three noise effects is maintained. However, when the three effects are manipulated individually, disturbing their normal relation given by the full noise effect, the STI predictions will fail. The data in Fig. 3 do illustrate this, and are just an example of specific types of signal processing for which the modulation-reduction based STI approach may fail.

B. Spectral subtraction and the second noise effect

In noise reduction research, the motivation for the concept of spectral subtraction, as a means to restore the original speech envelope, is most convincing when only the first noise effect is considered. However, it will be shown that the second noise effect plays a crucial role in diminishing the expected benefits.

In Fig. 5, the two panels in the upper row refer to the simplified image, only considering the first noise effect. The thin line represents a row of wavelet pixels defining the envelope of a small fragment of speech filtered $\frac{1}{4}$ octave around 1 kHz. These pixels can be considered as the input pixels in the panels of Fig. 3. In condition 1 the speech pixels remain unchanged, in condition 2 the speech pixels are lifted by the mean noise intensity, indicated by the thick line in Fig. 5. Since perception of modulations involves top-valley ratios rather then top-valley differences, the effect of the envelope increment on perceived modulations can be illustrated by equalizing both mean intensities [Fig. 5(b)]. The new envelope is compressed, and therefore the strength of modulations is reduced. In this simplified image, reducing the effect of noise is an extremely simple operation: Estimate the mean noise intensity and subtract this from the speech+noise envelope. Dependent on the accuracy of the noise mean estimation, this operation can fully restore the original speech envelope. However, speech/noise interaction in real-life also includes the second noise effect, and this leads to Figs. 5(c) and 5(d). In this picture, the S/N ratio is 0 dB just as in the upper row. However, now also the second noise effect is present, causing stochastic intensity fluctuation in the speech envelope beside an intensity lift (the relation between speech and speech+noise now corresponds to condition 8 of Fig. 3). In the concept of spectral subtraction, the main target is to neutralize the most damaging noise effect by subtracting the mean noise intensity. However, contrary to the simplified situation depicted in Figs. 5(a) and 5(b), the success of re-

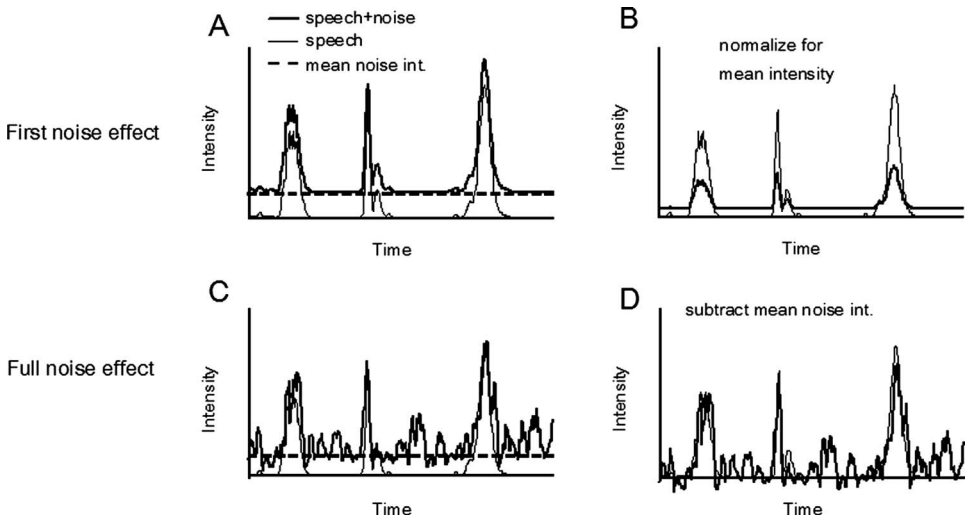


FIG. 5. The intensity envelope of 1 s of clean speech (thin line) is affected by two types of noise: noise consisting of only the first noise effect (upper panels) and the full noise effect (lower panels). The first type causes an intensity lift equal to the mean noise intensity (A). After normalization, the modulations are strongly reduced (B) and therefore intelligibility is reduced. However, this "first noise effect" type of speech corruption is easily counteracted by simply subtracting the noise mean, the essence of spectral subtraction. A more realistic situation is represented by the second noise type: Besides modulation reduction the noise causes additional stochastic level fluctuations in the speech envelope, indicated by (C). The speech envelope is not restored by subtraction: A number of levels have become negative valued and the fluctuations among the remaining noise pixels remain unchanged, as shown in (D).

storing the initial speech envelope is strongly limited. Due to the stochastic envelope fluctuations of noise and the coincidental phase interactions between speech and noise, a result of subtracting the mean noise intensity is that parts of the new envelope become negative-valued. To compensate for this, negative pixel values are usually made equal to zero, which causes so-called "musical noise" after inversed transformation. Moreover, after processing, the fluctuations among the positive-valued pixels remain essentially the same. Figure 6 shows an illustration of these effects. The picture shows three input/output diagrams: Two diagrams were copied from Fig. 3 [Figs. 3(a) and 3(b)], and a new diagram shows the input/output relation after spectral subtraction, i.e., decreasing each output pixel with the mean noise intensity [Fig. 3(c)]. Going from Fig. 3(a) to Fig. 3(c) shows the success of neutralizing the first noise effect (in the laboratory), causing an increase of word scores from 29% to 74%. Although there remains a negative influence of the second (and third) noise effect, the benefits in terms of intelligibility are substantial. Unfortunately, such operation is impossible in practice, since it requires for each pixel not only the speech+noise level, but also the clean speech level. Although spectral subtraction aims at neutralizing the first noise effect too, the result of the operation is quite different. In terms of word scores there appears to be no improvement: Scores drop from 29% to 24%. Instead of shifting pixels toward the $y=x$ line, roughly half of the noise-dominated pixels have become negative-valued (in the picture, these pixels are displayed at the −50 dB line). The other half of the pixels shows an interesting phenomenon: After subtraction, the distribution of the remaining noise-dominated pixels is very similar to the distribution before subtraction. This is illustrated in Fig. 7, showing two histograms of noise-dominated pixels, i.e., pixels of which the phase deviates more than ±15° from the underlying speech phases, before and after applying spectral subtraction. For each distribution, the levels were rms normalized. Hence, the result of spectral subtraction is very different from neutralizing the first noise effect as shown in Fig. 6(b). Figure 6(c) shows that the remaining group of pixels reorganizes into a new cluster with essentially the same distribution as before the operation. Although it can be shown that modulations may increase after spectral subtraction, it may well be that as a result of the unchanged noise statistics it remains equally difficult to extract speech cues from the corrupted signal, explaining why intelligibility remains equally poor.

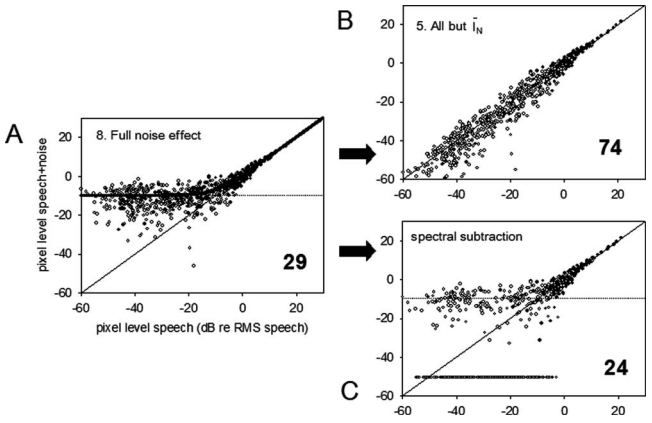


FIG. 6. The input/output diagram from Fig. 2 representing the full noise effect is redrawn in (A), except for the color-coding (now, speech-dominated pixels are represented by closed circles while the open circles depict noise-dominated pixels). (C) An input/output diagram of pixels obtained from speech after being subjected to a basic form of spectral subtraction. Although the number of pixels has been decreased, the distribution of the remaining pixels has not essentially changed compared to (A), and neither have the word scores.

V. CONCLUSIONS

The effect of noise on speech was divided into three subeffects: (1) a systematic lift of the envelope equal to the mean noise intensity, (2) the introduction of stochastic enve-

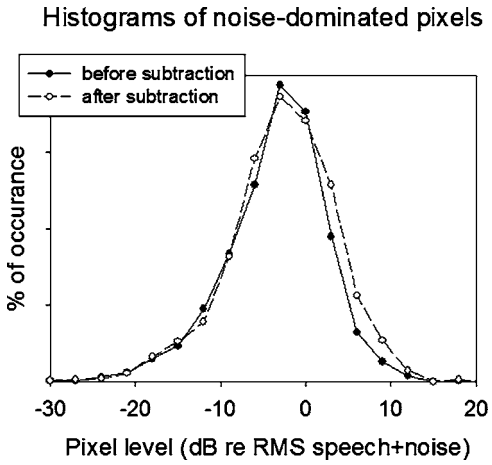


FIG. 7. Two histograms of wavelet pixels derived from two noise-corrupted speech signals. The first type is speech+noise; the second type is speech +noise after spectral subtraction. For each distribution, the levels were rms normalized. The picture shows that the pixel-level distribution is not essentially changed after spectral subtraction.

lope fluctuations and (3) the corruption of the fine structure. Wavelet transformation provides a suitable analysis tool for isolating and identifying these effects, and a strong processing tool for modifying speech by each of these effects separately or in any combination. CVC listening experiments were performed for the various noise effects. It was found that the most detrimental effect of the three subeffects is the systematic envelope lift, as essential speech modulations are reduced as a result of this effect. However, the remaining two effects are not negligible, and appear to be especially detrimental in case of noise suppression. It is argued that especially the introduction of the stochastic level fluctuations prevents spectral subtraction to be successful in terms of improving speech perception in noise.

ACKNOWLEDGMENTS

The preparation of this manuscript was supported by the Mgr J. C. van Overbeek foundation.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Breithaupt, C., and Martin, R. (**2003**). "MMSE estimation of magnitude-squared dft coefficients with superGaussian priors," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 848–851.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

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. **32**, 1109–1121.

Florentine, M., Fastl, H., and Buus, S. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**, 195–203.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

IEC. (**2003**). "Objective rating of speech intelligibility by speech transmission index," IEC 60268-16-2003, (3rd ed., International Electrotechnical Commission, Geneva).

Landau, H. J., and Pollak, H. O. (**1961**). "Prolate spheroidal wave functions, Fourier analysis and uncertainty. II," Bell Syst. Tech. J. **40**, 65–84.

Lee, Y. W. (**1960**). *Statistical Theory of Communication* (Wiley, New York).

Levitt, H. (**1986**). "Hearing impairment and sensory aids: A tutorial review," J. Rehabil. Res. Dev. **23**, xiii–xviii.

Lim, J. S. (**1978**). "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust., Speech, Signal Process. **26**, 471–472.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and bandwidth compression of noisy speech," Proc. IEEE **67**, 1586–1604.

Ludvigsen, C. (**1993**). "Evaluation of a noise reduction method—Comparison between observed scores and scores predicted from STI," Scand. Audiol. Suppl. **38**, 50–55.

Martin, R. (**2002**). "Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. **I**, pp. 253–256.

Noordhoek, I. M., and Drullman, R. (**1997**). "Effect of reducing temporal intensity modulations on sentence intelligibility," J. Acoust. Soc. Am. **101**, 498–502.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," *Auditory Physiology and Perception* (Pergamon, Oxford), pp. 429–446.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Rioul, O., and Vetterli, M. (**1991**). "Wavelets and signal processing," IEEE Signal Process. Mag., October, 14–38.

Strang, G. (**1994**). "Wavelets," Am. Sci. **82**, 250–255.

Van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window," J. Acoust. Soc. Am. **105**, 3425–3435.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Working Group on Communication Aids for the Hearing-Impaired (WGCA). (**1991**). "Speech perception aids for hearing-impaired people: current status and needed research," J. Acoust. Soc. Am. **90**, 637–685.

Perception of music performance on historical and modern commercial recordings

Renee Timmers[a)]
Centre for the History and Analysis of Recorded Music, Department of Music, King's College London, Strand, WC2R 2LS, London, United Kingdom

(Received 15 July 2006; revised 10 July 2007; accepted 22 August 2007)

Performing styles as well as recording styles have changed considerably within the 20th century. To what extent do the age of a recording, the unfamiliarity with performing style, and the quality of a reproduction of a recording systematically influence how we perceive performances on record? Four exploratory experiments were run to formulate an answer to this question. Each experiment examined a different aspect of the perception of performance, including judgments of quality, perceived emotion, and dynamics. Fragments from *Die junge Nonne* sung by famous singers from the start, middle, and second half of the 20th century were presented in a noisy and clean version to musically trained participants. The results show independence of perception of emotional activity from recording date, strong dependence of perceived quality and emotional impact on recording date, and only limited effects of reproduction quality. Standards have clearly changed, which influence judgments of quality and age. Additionally, changes restrict the communication between early recorded performers and modern listeners to some extent as shown by systematically smaller variations in communicated dynamics and emotional valence for older recordings.
 [DOI: 10.1121/1.2783987]

PACS number(s): 43.75.St, 43.75.Cd, 43.38.Md [DD] Pages: 2872–2880

I. INTRODUCTION

Performing style has changed considerably within the 20th century as comparisons between performances on modern and historical recordings suggest. For example, the use of extreme tempo fluctuations as observed in early 20th century recorded performances is now seen as highly inappropriate. Additionally, the use of pitch glides has become much less common among singers and violinists, while, on the other hand, vibrato has become more prominent as an expressive device (e.g., Day, 2000; Philip, 1992).

Similarly, the conditions of recordings and the quality of recording and reproduction of sounds have changed dramatically. Registration material changed for example from tin foil, to wax, to magnetic tape. Recording horns were used in different sizes and shapes. These were later replaced by microphones and electrical amplification before the introduction of the stereo microphone. Reproduction material and equipment also changed dramatically from the use of cylinders, vinyl, and shellac discs to microgroove discs, tape, and CDs, as did the equipment used to replay them. Moreover, technical improvements influenced the recording and reproduction of sound at every stage of these developments. This resulted in considerable changes in, e.g., the recorded frequency range, the noisiness of recordings and reproductions, the recorded acoustics, as well as the balance between different voices (see, e.g., Gelatt, 1956; Copeland, 1991; Day, 2000).

What does this imply for our (current-day) perception and evaluation of performances on record? To what extent do the age of a recording, unfamiliarity with performing styles, and the quality of a reproduction of a recording systematically influence how we perceive performances on record?

On the one hand, historical recordings are an amazingly rich and seemingly objective source of evidence about how music sounded in the past. Although listeners will readily recognize limitations of acoustic recordings from the early 20th century, these limitations decreased with improvements of recording techniques, and, even within these limitations, considerable information about the recorded music is preserved. This concerns, for example, relative variations in tempo and vibrato.

However, on the other hand, evaluations of historical recordings may be rather subjective. Most contemporary listeners are not familiar with the performing styles of the early 20th century, or are most of them familiar with the conditions of early recordings and the quality of the reproduction of early records. Moreover, the quality of historical recordings and the reproduction of these recordings may influence the perception of the recorded performances. For example, the limited frequency range of the recordings may limit the perception of consonants and timbre differences. Similarly the noisiness of reproductions may influence perception of volume or quality.

Whether listeners are able to listen through the differences in recording and reproduction quality and whether listeners are able to understand the intentions of performers even if the performing style is unfamiliar is unclear. So is the effect of this unfamiliarity on the perception of recorded performances. There is some evidence that the understanding of expressive intentions in performed music can be cross-

[a)]Current affiliation: "Music, Mind, Machine" group, Nijmegen Institute for Cognition and Information, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands.
Electronic mail: r.timmers@nici.ru.nl or reneetimmers@solcon.nl

cultural (Balkwill and Thompson, 1999), which suggests that understanding is independent of familiarity with a performing style. On the other hand, other studies have shown an effect of musical training on perception and interpretation of performance (Repp, 1995; Honing, 2007; Timmers *et al.*, 2006), which instead suggests a dependence on familiarity with performing styles.

The reported study set out to examine the influence of the age of a recording and the quality of reproduction on the perception of recorded performances and to compare this to the influence of performance characteristics. The aim was to investigate this in the context of existing recorded material.

Four exploratory experiments were run that each considered a different aspect of perception of performance. The first experiment concerned the perception of the age of a recording, the second concerned the evaluation of the quality and emotionality of a performance, the third concerned the perception of emotional activity and valence, and the fourth concerned the perception of dynamics (see the following for further explanation and see the Appendix for the instruction of each experiment). Perception was assessed through subjective judgments on a rating scale.

All experiments used the same material: Four fragments from *Die junge Nonne*, a late song by Franz Schubert, sung by six famous sopranos reproduced in a "clean" and "noisy" version. The first recording is from 1907 and the latest from 1977. The four fragments consist of musical passages of the song with distinct emotional characteristics: The first and second fragments are negative in emotion in comparing earthly life with a roaring storm and the darkness of one's heart with the grave, while the third and fourth fragments are more positive in character: The nun finds peace in joining the convent. Additionally, the first and third fragments have high emotional activity, while the second and fourth fragments have relatively low emotional activity; the mood turns from distress (F1), depression (F2), and excitation (F3) to resignation (F4). This is the surface meaning of the text. Alternative interpretations include, for example, that finding peace through an "eternal marriage with God" is actually a metaphor for an escape from the torments of earthly life through death.

The two versions concerned clean and noisy reproductions of a recording. The original recording is the same, but the reproduction differs in noisiness: 78 recordings were transferred either in a "flat" way, i.e., without any processing, or they were cleaned using noise-reduction and anticlick software. Tape or digital recordings issued on CD are, on the other hand, already perfectly clean. To get two versions of these recordings, noise was added and the signal was low-pass filtered to some extent (details are explained in Sec. II).

Three analyses of the collected data were run addressing three specific sub-questions. The first analysis tested the effect of recording date and reproduction quality on the perception of performance. The aim of this analysis was to see to what extent subjective judgments of performances depend on recording date and noisiness of the reproduction. It tested whether our perception of performances is essentially influenced by conditions regarding the recording and the age of a performance or whether it is essentially independent.

The second analysis tested the effects of singer and fragment, and, most important, the interaction between these effects on perception of performance. The aim of this analysis was to see to what extent subjective judgments depend on a performer's interpretation of the music. Fragment alone may influence judgments, the overall style of a performer may influence judgments, and the specific interpretation of the music by a performer may influence judgments, resulting in an interaction between the effects of fragment and singer.

Finally, the third analysis tested the relationship between perception and measured aspects of the performances. Strong correlations between judgments and aspects of the performances provide suggestive evidence for the relevance of performance. The three analyses together should highlight the biasing effects of recording date and reproduction quality as well as the impact of characteristics of performances on perception judgments irrespective of recording conditions.

The rationales for the different experiments were the following. The judgments of the age of the recording were included to function as baseline for the ratings of the other experiments. It is a measure of how distant in time participants perceive the different recordings to be. It was the only measure that asked listeners to judge the recording, although they were advised to pay attention to both the recording and the performing style, since historical recordings can be cleaned, LP's can be noisy, and noise can be added to CDs. In all other experiments, listeners were explicitly asked to pay most attention to the performance.

The judgments of quality and affect (Exp2) are of interest, because they may highly depend on familiarity with performing style, as well as on recording/reproduction quality. Nevertheless, all recordings used in the study were of singers considered among the best of their time. It may be possible that participants do recognize the quality of past singers. Moreover, as observed by Day (2000), rhetorical and grand gestures in performance were stronger in early 20th century than in later 20th century performances. This may make earlier performances more emotionally affecting than later performances.

The judgments of perceived emotion (Exp3) are of interest, also because of an ambiguity in possible outcome: On the one hand, performing style changed considerably and therefore communication of a performer's intention to current day listeners may be difficult for older recordings. On the other hand, several authors and investigations have suggested that expression of emotion in singing and music performance has universal characteristics shared with expression of emotion in speech (Juslin and Laukka, 2003; Scherer, 1986; 1995; Sundberg, 1987). It would therefore be likely that communication of emotions is possible irrespective of recording date, as long as the relevant information is present. Indeed, analysis of the way singers express the different moods within Schubert songs showed high consistency between singers, over different time periods, despite evident changes in performing style (Timmers, 2007).

The judgments of perceived emotion were done using two rating scales: emotional valence and emotional activity. Valence and activity are two dimensions that distinguish well between different emotions (Russell, 1980). Emotions may

TABLE I. Overview of recordings used in the experiments.

Performers	Ref	Source
Susan Strong	SS 07	"Schubert Lieder on Record I, 1898-1939,"
Orchestra	Clean	EMI Classics 5 66150 2, 1997
Susan Strong	SS 07	HMV matrix 2004 f
Orchestra	Noisy	
Susan Metcalfe-Casals	SMC 37	"Schubert Lieder on Record II, 1929-1952,"
Gerald Moore	Clean	EMI Classics 5 66154 2, 1997
Susan Metcalfe-Casals	SMC 37	HMV matrix CTPX 3884-1
Gerald Moore	Noisy	
Lotte Lehmann	LL 41	"Lotte Lehmann: Schubert," LYS 231-234, 1997
Paul Ulanowsky	Clean	
Lotte Lehmann	LL 41	Columbia matrix XCO 30013-1
Paul Ulanowsky	Noisy	
Elisabeth Schwarzkopf	ES 52	"Schubert: 12 Lieder, 6 Moments musicaux,"
Edwin Fischer	Clean	EMI Classics 5 67494 2, 2000
	Noisy	
Elly Ameling	EA 75	"Schubert Liederen,"
Dalton Baldwin	Clean	Philips 464 334-2, 1999
	Noisy	
Gundula Janowitz	GJ 77	"Schubert Lieder,"
Irwin Gage	Clean	Deutsche Grammophon 453 082-2, n.d.
	Noisy	

have positive or negative valence, such as happiness compared to anger, and they may have high or low arousal, such as anger compared to sadness or depression. The use of these dimensions was preferred over the use of specific emotion words, because it allows for subtle distinctions between performers to come forward: Overall, a musical passage may be perceived to be negative. Within this overall tendency, one performance may be perceived to be more negative than another. These subtle differences are hard to express in words, and listeners tend to disagree on terminology when asked to characterize music in subcategories (Gabrielsson and Juslin, 2003).

Finally, the judgments of dynamics are of interest in two respects (Exp4). First, comparison between judgments of dynamics and measurements of amplitude is a test for the reliability of amplitude measurements. Second, it is of interest to compare the perceived range in dynamics for historical recordings with that of modern recordings. It is likely that historical recordings tend to have a smaller dynamic range than nowadays possible. Acoustical recordings were very noisy and needed a loud signal for a proper signal to noise ratio (Gelatt, 1956). On the other hand, an overload of the cutter due to too loud sounds had to be avoided as well. The situation improved with the introduction of microphones and amplification with electrical recording. Nevertheless, very soft and loud sounds remained problematic. Recording engineers started to control the recorded signal and often "tamed" the performed dynamic range to avoid overload and ensure audibility (Copeland, 1991).

It should be noted that this is an exploratory study that uses existing recorded material. This makes the study interesting for music research on recordings and ensures ecological validity. The drawback is, however, that the results are not entirely clear-cut: The effect of recording date on perception of performance combines the effect of performing style and recording conditions. The effect of singer similarly combines the effect of performer and recording conditions. Therefore, the three analyses are needed to come to a complete interpretation of the data.

II. METHOD

A. Musical material

Six performances of *Die junge Nonne* were selected from a database of recordings. *Die junge Nonne* is one of the songs by Schubert that has been recorded regularly throughout the 20th century. As mentioned earlier, characteristic of the song is that it contains a succession of moods.

The aim was to have a set of early performances, in a relatively unknown style, and a set of later performances in a more familiar style. This aim was counterbalanced with the aim of having performances spreading a time period more evenly. The result was the choice for three performances from before 1945, and three performances from after 1950, assuming a break in performing style around the second world war as was observed by Philip (1992). The restriction to six performances in total was made to limit the total number of stimuli to be used in the experiments. Details of the recordings used in the study are listed in Table I.

Four fragments from each performance were selected to serve as musical material. Each fragment has a specific mood: Fragment 1 (F1, bars 36–41) is high in activity and negative in mood. Fragment 2 (F2, bars 43–49) is low in activity and negative in mood. Fragment 3 (F3, bars 54–61) is high in activity and positive in mood, while Fragment 4 (F4, bars 71–74) is low in activity and positive in mood. The moods of each fragment were determined in a previous study (Timmers, 2007), based on the meaning of the text as well as structural aspects of the composition.

Two versions were used of each selected performance—one clean and the other noisy. The clean versions were taken from commercially issued CDs. To acquire an even cleaner version of the recording of Elisabeth Schwarzkopf, the few cracks and clicks in the digitized recording were diminished using declicker and denoiser functions of an audio editing program.

In this context, clean and noisy should not be interpreted in absolute terms, but relatively: Clean means relatively clean compared to its respective noisy version, and noisy means relatively noisy. A clean 78 is not as free from noise and clicks as a modern recording. Likewise, a clean acoustical 78 recording has a lower signal to noise ratio and is more limited in frequency range than a clean electrical 78 recording. The main point of having a clean and noisy version of a recording is to have two versions that differ in transfer of the original performance and that can be considered different in quality of reproduction.

The noisy versions of the 78 recordings were acquired by making a flat transfer of the recording: The original 78s were played back using a modern turntable. The analogue output from the turntable was led to an amplifier and into an analog-to-digital converter. The digital output was led into a personal computer and was recorded.

To obtain a noisy version of the recordings of Elly Ameling and Gundula Janowitz, noise had to be added to the recordings. Additionally, the recordings were modified to sound "older." First, the signal was compressed. Second, noise was added. Noise was acquired by recording the playback of a blank SP shell ac disc from 1950, which was used to add noise to the recording of Elly Ameling, and a blank SP shellac disc from 1935, which was used to add noise to the recording of Gundula Janowitz. Finally, the mixed audio track was bandpass filtered by reducing the amplitude gradually below 50 Hz (monotonically toward 30 Hz) and above 3000 Hz (monotonically toward 5000 Hz). This means that also the noise was deamplified at higher and lower frequencies. This enhanced the integration between the noise and the signal of the recorded performance. As a final modification, all audio files were saved as mono tracks and the resolution was set to 22000 Hz 16 bits.

B. Participants

All participants had more than ten years of formal musical training. Most of them were university music students (second year and higher). The others were advanced performers. Participants had a variety of nationalities (e.g., British, Dutch, American, Israeli, Greek, Japanese). Two were German native speakers. Most Dutch participants were able to understand German. All participants had a background in classical music.

Participants did three experiments in a row to limit the time per participant. Participants were randomly assigned to the experiments. In total, the number of participants was 22 for Exp1, 26 for Exp2, 40 for Exp3, and 32 for Exp4. The main reasons for the number of participants to vary over experiments were practical. Originally, two additional experiments were run.

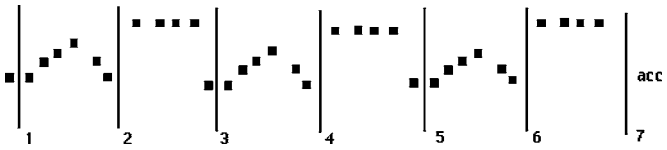


FIG. 1. Answer sheet for Fragment 1 of Exp4. The dots represent the melody of the first fragment. Bars are indicated by vertical lines and numbers.

C. General procedure

Participants were seated behind a laptop and read the instructions from a print out (instructions are given in full in the Appendix). After a general introduction to the experiments, the instruction was given for the first experiment. The participants started the experiment immediately without a practice trial. They put on headphones (Sony MDR-7506) and used a mouse to play a stimulus, to give the ratings, and to press the ok/save button to go to the next stimulus (all programmed in POCO, see Honing, 1990). Sound levels of the playback were set to a comfortable level and fixed throughout the experiments. The labels above the radio buttons of the rating scales changed with experiment. Either one or two rating scales were used depending on the experiment. The order of the two rating scales was counterbalanced between participants. The presentation order of the musical stimuli was randomized over participants.

Separate answer sheets were used for Exp4. For this experiment, the computer interface was only used to play stimuli. The answer sheets showed a representation of the sung melody of a musical fragment and bar lines were indicated. This was necessary, because the participants indicated the dynamics per bar. Figure 1 shows the representation of the melody of F1.

Because the stimuli were presented randomly to the participants, participants did not know beforehand which fragment would sound. For assistance, the fragment number was indicated before each stimulus in Exp4 using a computer voice mentioning the fragment number.

As mentioned before, all participants did three experiments in a row. The instruction for each experiment was given just before the start of an experiment. The instructions are given in the Appendix. For Exps 1, 2, and 3, all 48 stimuli were presented in random order after each other. These 48 stimuli included six performances of four fragments in two versions — clean and noisy. For Exp4, the 48 stimuli were split in half: half of the participants judged the noisy versions of the recordings and the other half judged the clean versions of the recordings. This was done to restrict the duration of this experiment. Each experiment took approximately 20 min, which resulted in an overall duration of about an hour per participant.

III. RESULTS

A. Effect of date and version

First, the effects of recording date and version on the judgments were examined. Average ratings over fragment were used for this analysis. Figure 2 shows the mean ratings per recording date and version for the variables of the differ-

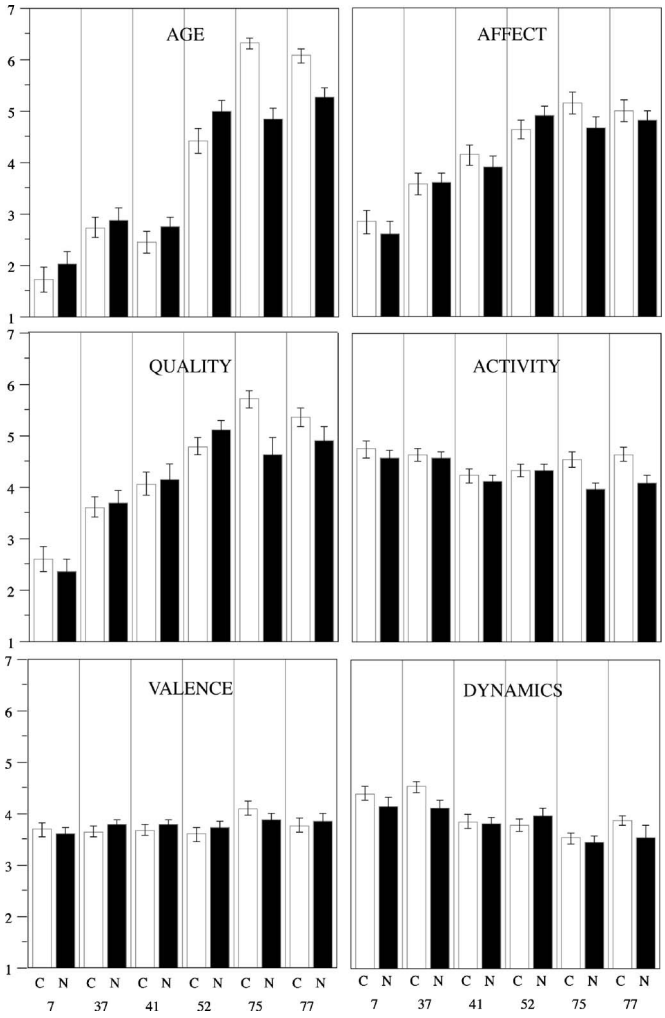


FIG. 2. Average and standard errors of ratings of Exps 1–4 per date and version.

ent experiments. It can be seen that the slope of the relationship between recording date and judgment is steep for age, and quality and almost flat for valence and activity. The effect of version is small for all judgments. It is generally larger for modern recordings that were made noisy than for older recordings, especially for the judgments of age, quality, and activity.

Table II reports the explained variances and significance values of the results of a series of mixed model ANOVAs. Each ANOVA had date (continuous variable) and version (nominal variable) as independent variables and one of the judgements as dependent variable. In a mixed model ANOVA, participants are treated as random effect and date and version as fixed effects. Date and version are within subject effects for all experiments except for Exp4, which varied version across participants. The recommended residual maximum likelihood method was used as estimation method.

The results confirm the observations made with respect to Fig. 2; the effects of date were considerably stronger than the effects of version or the interaction between date and version. The effect of date was small for judgments of activity and valence.

To examine, in addition, whether the range in responses changed systematically over time, the analyses were rerun using the standard deviation of responses over fragments within singers as data points. The hypothesis was that, for some judgments, the variation in judgments could be smaller for early singers than later singers. For example, the variation in dynamics was predicted to be more restricted for early recordings than for later recordings, because of an expected smaller range in volume differences between musical fragments.

The effect of date was significant, but small, for all judgments, except quality. The effect of date was relatively strong for judgments of dynamics ($p<0.0001$) and valence ($p<0.01$). Focusing on these stronger effects, the amount of variation increased over time suggesting restricted variation in perceived dynamics and valence for earlier performances (see Fig. 3). The effect of version was significant for judgments of quality only ($p<0.001$). Variation in perceived quality was more restrained for the noisy versions (all tended to be lower in quality) than the clean versions.

B. Effect of fragment and singer

The second analysis of the data examined the effect of fragment and singer and specifically the interaction between these effects on the judgments: Did singers communicate a personal interpretation of the musical fragments?

TABLE II. Summary of results of mixed model ANOVAs for Exp1 ($N=22$), Exp2 ($N=26$), Exp3 ($N=40$), and Exp4 ($N=20$) testing the effects of recording date (one level) and version (two levels). Partial explained variances (R^2), F ratios, and p values are given for significant effects.

	Date			Version			Date ×Version[a]		
	R^2	F	p	R^2	F	p	R^2	F	p
Exp1									
Age	0.60	472	<0.0001			n.s.	0.02	17.1	<0.0001
Exp2									
Affect	0.35	186	<0.0001			n.s.			n.s.
Quality	0.38	270	<0.0001			n.s.	0.01	4.02	<0.05
Exp3									
Activity	0.02	13.1	<0.0001	0.02	15.7	<0.01	0.01	6.36	<0.05
Valence	0.01	9.29	<0.01					n.s.	n.s.
Exp4									
Dyn.	0.23	76.5	<0.0001			n.s.			n.s.

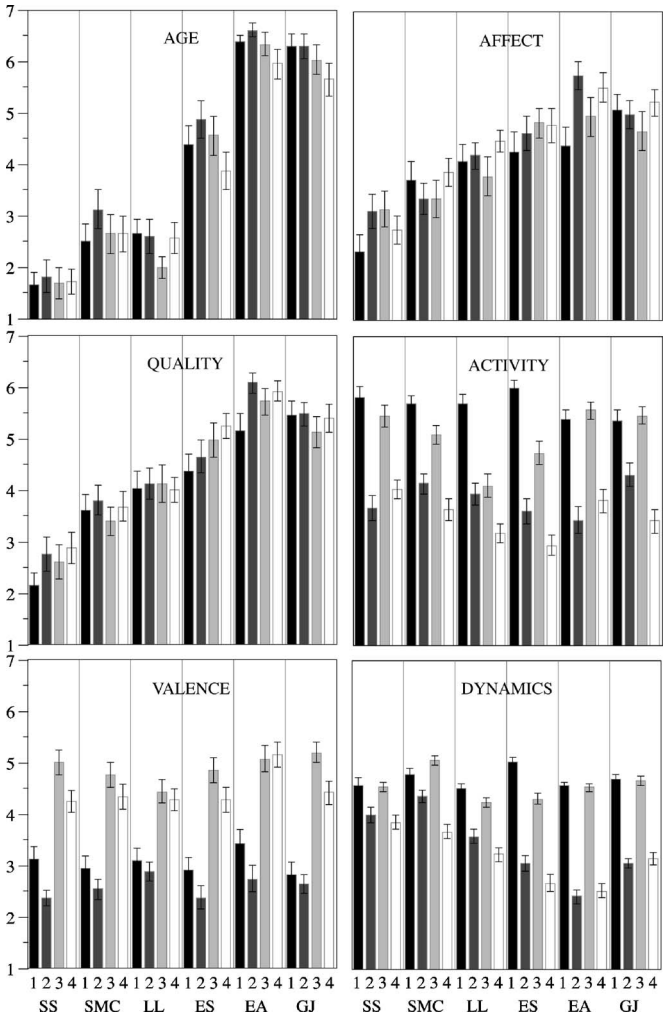


FIG. 3. Average and standard errors of ratings of Exps 1–4 per singer and fragment of the clean versions of the recordings. Singers are ordered according to recording date.

For this analysis, only the judgments of the clean versions of the recordings were used. A series of repeated measures ANOVAs were used to test the effects of fragment (nominal) and singer (nominal) and the interaction between them for each judgment separately. For Exp4, data consisted of the average rated dynamics per performance (averaged over bars).

Figure 3 and Table III show summaries of the results. For most of the judgments, there is only one variable that was significant or highly significant and contributed most to the explained variance. This is singer for the judgments of age, affect, and quality, and fragment for the judgments valence and activity. For dynamics, both effects of fragment and singer are highly significant. The interaction between fragment and singer is highly significant for the judgments of activity and dynamics. It is just significant for the judgments of valence.

These results confirm the division observed in the first analysis between the judgments of age, quality, and affect, on the one hand, and the judgments of emotional activity, valence, and dynamics, on the other hand. The first group of judgments vary strongly with recording date. They also intercorrelate strongly: The correlation is 0.86 for average perceived age and affect, 0.92 for average perceived age and quality, and 0.96 for average perceived affect and quality. The judgments of the second group depend, however, more strongly on the musical fragment and, especially for perceived emotional activity and dynamics, the performers' interpretation of the music. The judgments of dynamics and activity are strongly correlated (r=0.84).

C. Correlations with characteristics of the performances

The final analysis of the data examined the relationship between judgments and aspects of the performances. Ratings averaged over participants of the clean versions of the performances were correlated with measurements of the performances. These measurements were made in a previous study (Timmers, 2007). The measurements included duration of each bar in seconds, the average sound level of each bar, the average vibrato rate of a long note in each bar in cycles per second, the extent of a large vibrato cycle of a long note in

TABLE III. Summary of results of repeated measures ANOVA for Exp1 (N=22), Exp2 (N=26), Exp3 (N=40), and Exp4 (N=22) testing the effects of fragment (four levels) and singer (six levels) on the judgments of the clean recordings. Partial explained variances (R^2), F ratios, and p values are given for significant effects.[a]

	Fragment			Singer			Fragment × Singer		
	R^2	F	p	R^2	F	p	R^2	F	p
Exp1									
Age	0.01	3.46	<0.05	0.62	105	<0.0001			n.s.
Exp2									
Affect			n.s.	0.22	21.3	<0.0001			n.s.
Quality			n.s.	0.40	47.9	<0.0001			n.s.
Exp3									
Activity	0.33	88.4	<0.0001	0.02	3.43	<0.05	0.05	6.26	<0.0001
Valence	0.29	41.6	<0.0001	0.01	3.14	<0.05	0.01	2.02	<0.05
Exp4									
Dyn.	0.52	184	<0.0001	0.11	41.0	<0.0001	0.10	15.7	<0.0001

[a]Effects are significant using Greenhouse-Geisser epsilon for violations of sphericity for effects with larger number of levels.

TABLE IV. Significant correlations ($p<0.05$) between mean judgments (rows) and measurements (columns) of the clean recording of each fragment and each singer.

	Sound level	Bar duration	Vibrato extent	Vibrato rate	Up	Down
Age	−0.78	0.41		−0.62		
Affect	−0.89	0.62		−0.71		
Quality	−0.89	0.63		−0.75		
Activity			0.70		0.53	−0.41
Valence		0.50				0.45
Dynamics	0.63		0.57	0.44	0.55	

each bar in semitones, and the number of pitch glides up and down in each bar. Table IV shows the significant correlations.

All judgments show significant correlations with several aspects of the performances. Many of the correlations are high, except for valence, which shows only moderate correlations with performance aspects.

To interpret these correlations, it is useful to take the results of the measurement study (Timmers, 2007) into account. From the measurements, several systematic changes in performing style within the 20th century were observed. The amount of rubato tended to decrease over time, global tempi tended to decrease, later performances tended to be softer on average than early performances, vibrato rate decreased gradually over time, while vibrato extent increased over time, and the number of pitch glides was medium in the beginning of the 20th century, increased toward the 1930s, and decreased after the 1940s.

The correlations reported in Table IV partly reflect these changes over time. Judgments of age correlate negatively with average amplitude and vibrato rate and positively with average bar duration. Judgments of quality and affect also show these correlations. Notably, the correlations with quality are highest, suggesting that quality is strongly related to performing style.

Dynamics and emotional activity are, on the other hand, correlated with vibrato extent and number of pitch glides up or down. Perceived dynamics is correlated with measured amplitude, but the correlation is not very high. This suggests only limited reliability of the measurements of sound level to represent dynamics.

Significant correlations with perceived valence include medium correlations with average bar duration and number of downward pitch glides. These correlations confirm the relationship between valence and aspects of performances for this particular song as observed in Timmers (2007): Positive passages tended to be slower in tempo and had more downward pitch glides than negative passages. This can be understood if we interpret the positive passages to be a release of the negative tension rather than, e.g., a positive excitation or uplift.

Perceived dynamics was the only judgments that participants rated per bar. For this judgment, one more analysis was done and judgments per bar were correlated with measurements of sound level per bar for each singer individually. Table V shows the correlations that were significant. Notably, the correlations are now considerably higher than in Table IV for all modern recordings, starting from the recording of Lotte Lehman from 1941. The older recordings show lower correlations with an insignificant correlation between measured sound level and perceived dynamics for the recording from 1937.

A possible reason for the insignificance of the correlation for the recording of Susan Metcalfe-Casals from 1937 is the great difference in amplitude between voice and piano. If the singer sings only half a measure, the measured sound level drops considerably, while participants may rate the dynamics as forte based on the dynamics of the voice. Perceived loudness can be corrected for presence or absence of the voice by multiplying the judged dynamics with the fraction of the bar that the singer sings. After this rough correction of the judged dynamics, the correlation between measured sound level and perceived dynamics is significant ($r=0.60$, $p<0.01$).

The discrepancy between the correlation between perceived dynamics and measured sound level reported in Table IV and the correlations reported in Table V suggests that measurements of sound level capture relative variations in dynamics within a given recording more reliably than the relative loudness of different recordings. This is not necessarily a deficit of the measuring method, but may also be due to subjective perception of dynamics and the task of the participants in Exp4. The participants were instructed to write down the variations in dynamics within a musical fragment and used the scale for this. Their task was not to compare the

TABLE V. Significant correlations between mean judgments of dynamics per bar and measurements of sound level per bar, calculated for each singer separately.

	SS 07	SMC 37	LL 41	ES 52	EA 75	GJ 77
Sound level	0.52		0.83	0.88	0.92	0.88

relative loudness of different recordings. The indication of relative loudness of different recordings was only an indirect result of the task to indicate the dynamics within a performance.

IV. DISCUSSION AND CONCLUSION

The aim of the study was to examine the influence of the age of a recording and the quality of reproduction on the perception of recorded performances and to compare this to the influence of performance characteristics. This was done in an exploratory study using commercial historical and modern recordings.

Clear tendencies of judgments to systematically change with recording date were observed for all perceptual aspects. Judgments of age and quality changed most strongly with recording date followed by judgments of affect and dynamics. Judgments of perceived emotion were most independent of recording date. Additionally, the variation in communicated emotional valence and dynamics over musical fragments tended to be more restrained for older recordings than for modern recordings.

Tendencies of judgments to change with recording version were, in contrast, small for all judgments and significant for only a few perceptual aspects. This suggests that listeners were able to abstract relevant information from specific reproduction conditions.

The importance of performance characteristics was suggested by a significant interaction between the effects of fragment and singer for the judgments of perceived emotion and dynamics. This interaction highlights the influence of the performer on the perception of the musical fragments, which suggests communication of a personal interpretation of the music.

Finally, significant correlations between measured characteristics of the performances and perceptual judgments were observed for all judgments. This suggests that not only the judgments that showed an interaction between fragment and singer, but also the judgments that varied most strongly with recording date may have varied due to changes in performance characteristics.

In short, the main result of the study was the clear division between perceptual judgments that varied strongly with recording date and perceptual judgments that varied less strongly with recording date. Additionally, a limited effect of reproduction version was observed and strong correlations between perceptual judgments and measured performance variables, suggesting that the actual influence of the recording is limited compared to the influence of performance characteristics.

However, it should be noted that this conclusion is drawn tentatively. The study used existing recorded material. While this enhanced ecological validity, it limited the control over the experimental material. The effects of version and date are not single effects, but consist of different variables: Recording date implies differences in performing style as well as recording conditions and reproduction conditions, while reproduction version consists of differences in noisiness, frequency range, and possibly other aspects such as dynamic compression or boost. Each experiment could be refined and specific effects examined. Additionally, it might be useful to define sensitivity thresholds: For example, noisiness may influence perceptual judgments from a specific noise level onwards. The specific level may vary with perceptual aspect.

Nevertheless, the study generated interesting results, adding to a growing literature on perception of performance. For example, it showed a strong association between judgments of quality and affect, which emphasizes the possible importance of aesthetic experiences for emotional affect (e.g., Scherer, 2004). In contrast, emotional affect was not strongly related to judgments of emotional activity, which suggests that felt emotional impact is quite different from perceived emotional activity (Gabrielsson, 2001). Emotional affect was also negatively correlated with sound level, while emotional activity correlated positively with sound level. This further emphasizes the complexity of stimulus-response relationships for emotional arousal.

The high correlations observed between variations in judged dynamics and measured sound levels per bar are promising for research that uses measurements to assess performance characteristics. Nevertheless, the exact relationship between perception and measurement needs to be further examined in future research. Part of the complexity of perception of dynamics was highlighted by showing a possible focus of listeners on the voice when judging dynamics.

Part of the contribution of the study is to raise issues for further research. In being explorative, it addressed perception of recorded performances rather broadly. It distinguished between perceptual judgments that are highly sensitive to differences in recording date and judgments that are almost independent of recording date. It remains an interesting issue how listeners perceive old recordings. Listeners seem certainly able to listen through differences in manners of record reproduction and judgments are strongly associated with performance variables. Nevertheless, standards have clearly changed, which influence judgments of quality and age. Additionally, changes restrict to some extent the communication between early recorded performers and modern listeners.

ACKNOWLEDGMENTS

This research was funded by the Arts and Humanities Research Council's Research Centre for the History and Analysis of Recorded Music (CHARM). I would like to thank Daniel Leech-Wilkinson for his help during the preparation of the studies and for collecting the Schubert material. My thanks also go to Karsten Lehl for the flat transfers of the early 78 recordings, and to Roger Beardsley for the recordings of blank shellac 78s.

APPENDIX

1. Instruction experiment 1

In this experiment, you will hear short fragments of recordings of Schubert songs. Your task is to decide for each fragment if you think it is an historical recording from before 1945 or a modern recording from after 1950. You do this on

a scale from 1 to 7. 1 stands for certainly before 1945, and 7 for certainly after 1950. Try to use levels other than 4, and the extremes, as much as possible. Please note that the amount of noise or clicks is not a good criterion for the "age" of a recording, since historical recordings can be cleaned (noise is filtered out) and LP recordings can also have cracks and noise, and noise can be added digitally to modern recordings. Therefore pay most attention to the performance and try to base your answer on that. There will be 48 fragments in total.

2. Instruction experiment 2

In this experiment, you will hear short fragments of recordings of Schubert songs. Your task is to evaluate the quality of the performance and how much the performance affects you emotionally on a scale from 1 to 7. Please try to use the entire scale — both the extremes and the middle levels.

The reason to use these two rating scales is that you may consider a performance to be "good" and "well-performed" (the quality is high), but at the same time the performance may not affect you emotionally (affect is low). Other performances may be less "perfect" (quality is low), but may touch you much more (affect is high). There will be 48 fragments in total.

3. Instruction experiment 3

In this experiment, your task is to indicate the emotion you perceive "in" the performance. You do this by characterizing the perceived emotion along two dimensions—valence and activity. The term valence is used to indicate whether the perceived emotion is positive or negative. The term activity is used to indicate whether the perceived emotion is active or passive.

The dimensions of valence and activity were found by different researchers to give a suitable summary of emotions and relations between them. Some emotions have activity associated with them, such as joy and anger, while other emotions have passivity associated with them, such as sadness, and boredom. In addition, some emotions are seen as positive, while others are considered negative.

In this experiment, you indicate valence and activity on a scale from 1 to 7 for 48 fragments of recorded performances of Schubert songs. When rating valence, 1 stands for negative and 7 for positive. When rating activity, 1 stands for low and 7 for high. Please try to use the entire scale, so try to use both the extremes as well as the middle levels. Note that the emotion you feel may be different from the emotion you perceive in the performance. In the current task, we are interested in the communication from performer to listener. So we would like to know what intended emotion you perceive rather than how much the music affects you.

4. Instruction experiment 4

In this experiment, you notate the dynamics of a performance and you do this for 24 fragments of recorded performances of Schubert songs. Listen to the music and start notating the dynamics on the sheet, below the representation of the respective melodic line. Before each fragment, you will be told which of the four musical excerpts will sound. Indicate levels of dynamics using pp to ff (or 1–6 if you find that easier). Indicate changes in dynamics also by using pp to ff and not by writing crescendo or decrescendo. Please make one marking of dynamics per bar. Note that different musicians perform the same musical excerpts. We are interested in differences between performers — so even if they perform the same music, you may perceive that they use different dynamic levels. You may listen to a performance more than once if necessary.

Balkwill, L. L., and Thompson, W. F. (**1999**). "A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues," Music Percept. **17**, 43–64.

Copeland, P. (**1991**). *Sound Recordings* (British Library Board, London).

Day, T. (**2000**). *A Century of Recorded Music: Listening to Musical History* (Yale University Press, New Haven).

Gabrielsson, A. (**2001**). "Perceived emotion or felt emotion: Same or different?," Musicae Scientiae, Special issue, 2001/2002, 123–147.

Gabrielsson, A., and Juslin, P. N. (**2003**). "Emotional expression in music," in *Handbook of Affective Sciences* edited by R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (Oxford University Press, Oxford, pp. 503–534).

Gelatt, R. (**1956**). *The Fabulous Phonograph: The Story of the Gramophone from Tin Foil to High Fidelity* (Cassell & Company, London).

Honing, H. (**1990**). "POCO: An environment for analysing, modifying and generating expression in music," in *Proceedings of the 1990 International Computer Music Conference* (CMA, San Francisco), pp. 364–368.

Honing, H. (**2007**). "Is expressive timing relationally invariant under tempo transformation?," Psychol. Music. **35**, 1–10.

Juslin, P. N., and Laukka, P. (**2003**). "Communication of emotions in vocal expression and music performance: Different channels, same code?," Phys. Bl. **129**, 770–814.

Philip, R. (**1992**). *Early Recordings and Musical Style: Changing Tastes in Instrumental Performance, 1900-1950* (Cambridge University Press, Cambridge).

Repp, B. H. (**1995**). "Detectability of duration and intensity increments in melody tones: A partial connection between music perception and performance," Percept. Psychophys. **57**, 1217–1232.

Russell, J. A. (**1980**). "A circumplex model of affect," J. Pers Soc. Psychol. **39**, 1161–1178.

Scherer, K. R. (**1986**). "Vocal affect expression: A review and a model for future research," Polymer **99**, 143–165.

Scherer, K. R. (**1995**). "Expression of emotion in voice and music," J. Voice **9**, 235–248.

Scherer, K. R. (**2004**). "Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them?," J. New Music Res. **33**, 239–251.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, DeKalb).

Timmers, R., Marolt, M., Camurri, A., and Volpe, G. (**2006**). "Listeners' emotional engagement with performances of a Scriabin étude: An explorative case study," Psychol. Music **44**, 481–510.

Timmers, R. (**2007**). "Vocal expression in performances of Schubert songs on record." Musicae Scientiae **XI**, 73-101.

Stream ambient noise, spectrum and propagation of sounds in the goby *Padogobius martensii*: Sound pressure and particle velocity[a)]

Marco Lugli[b)]
Dipartimento di Biologia Evolutiva e Funzionale, Universitá di Parma, Italy

Michael L. Fine
Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284-2012

(Received 31 May 2006; revised 31 July 2007; accepted 10 August 2007)

The most sensitive hearing and peak frequencies of courtship calls of the stream goby, *Padogobius martensii*, fall within a quiet window at around 100 Hz in the ambient noise spectrum. Acoustic pressure was previously measured although *Padogobius* likely responds to particle motion. In this study a combination pressure (p) and particle velocity (u) detector was utilized to describe ambient noise of the habitat, the characteristics of the goby's sounds and their attenuation with distance. The ambient noise (AN) spectrum is generally similar for p and u (including the quiet window at noisy locations), although the energy distribution of u spectrum is shifted up by 50–100 Hz. The energy distribution of the goby's sounds is similar for p and u spectra of the Tonal sound, whereas the pulse-train sound exhibits larger $p-u$ differences. Transmission loss was high for sound p and u: energy decays 6–10 dB/10 cm, and sound p/u ratio does not change with distance from the source in the nearfield. The measurement of particle velocity of stream AN and *P. martensii* sounds indicates that this species is well adapted to communicate acoustically in a complex noisy shallow-water environment. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2783113]

PACS number(s): 43.80.Ev, 43.80.Ka, 43.30.Nb [WWA] Pages: 2881–2892

I. INTRODUCTION

Acoustic pressure and particle velocity are physically related components of sound (Morse Uno Ingard, 1968; Michelsen, 1983). Pressure p is a scalar, i.e., its magnitude at a given point is the same regardless of the orientation of an (omnidirectional) receiver to the sound source. Particle displacement and its derivatives, particle velocity u and acceleration, are vectors, and their magnitude is maximal along the propagation axis of the sound.

In the far field several lengths from the source and in absence of reflecting boundaries, the front of an acoustic wave is approximately planar, and p and u are in phase; their ratio (p/u) is proportional to the product of the water density (ρ) and sound velocity (c) in the medium (Michelsen, 1983). In the near field well below one wave length to the source, particle velocity predominates over pressure, and the phase difference between the two quantities changes with distance. Further, the magnitude of p/u increases with propagation distance for simple sound sources, i.e., a pure tone produced by a monopole, dipole or quadrupole source (Kalmijn, 1988). Velocity depends on factors such as the geometry and frequency response of the source and the presence of nearby reflecting boundaries (Michelsen, 1983; Rogers and Cox, 1988). Excluding simple sound sources under ideal conditions, the relationship between the two quantities becomes quite complex and unpredictable in the near field. To our knowledge there are no theoretical or empirical studies describing the p/u amplitude and phase changes with propagation distance for acoustic frequencies below the cutoff frequency, i.e., under very shallow water conditions.

Many sonic fishes live in shallow waters (near the shore of the sea or lakes, in small rivers, ponds, etc.) and emit low frequency acoustic signals (i.e., with long wavelengths) intended for receivers at short distances. For example the wavelength of a 100 Hz sound would be approximately 15 m. Thus, acoustic communication in many teleosts occurs mainly or exclusively in the near field, and many fishes that communicate at close distance are primarily or exclusively particle-motion sensitive species (reviewed in Popper and Fay, 1973).

Sensing particle velocity in water for military purposes such as submarine detection has long been performed using a geophone in sonobuoys. However, particle displacement of fish sounds has been rarely reported in the fish literature. Horch and Salmon (1973) measured both acoustic pressure and particle displacement components of the sound in the squirrelfish, *Myripristis violaceus*. Using separate sensors, an omnidirectional hydrophone for p and a bidirectional geophone for u, they found similar spectral energy distribution and attenuation levels for both components. Simultaneous characterization of the scalar and vectorial quantities of the same sound was not performed. The recent development of a $p-u$ probe for underwater measurements allows simulta-

[a)]The species' scientific name has recently changed to *Padogobius bonelli*. However, *P. martensii* is retained in this paper to facilitate the tracking of the subject species across related papers.
[b)]Electronic mail: Marco.Lugli@Unipr.it

neous measurement of both parameters along a single axis at the same point in the acoustic field. Bastyr *et al.* (1999) discussed the advantages of a $p-u$ probe over other methods that employ the finite differencing technique (i.e., $p-p$ or $u-u$ probes) for accurate intensity measurements in complex acoustic fields.

The Italian freshwater goby, *Padogobius martensii*, is a bottom-dwelling species whose acoustic behavior has been investigated thoroughly (Torricelli and Romani, 1986; Lugli *et al.*, 1997). This small goby (maximum size: 9 cm total length (TL)) lives in small stony streams and rivers. Individual adult males defend a hollow under a large stone. During the breeding season (March–July), males court ripe females to attract them to the nest hollow for mating (Torricelli *et al.*, 1986; Lugli *et al.*, 1992), after which the male cares for the eggs. The nest hollow has one or, most often, two openings.

Courting males produce two sound types (Lugli *et al.*, 1995). The tonal sound, typical of early courtship, is made of rapidly repeated pulses (sound duration: <0.5 s, pulse rate: 150–200 pulses/s (pps); Torricelli *et al.*, 1990). The pulse-train sound, emitted only when the female enters the nest hollow, is a short train of low-frequency pulses repeated at a lower rate (around 50 pps, Lugli *et al.*, 1995). Pulse-train sounds are often combined with a short tonal sound to form complex "spawning sounds" (Lugli *et al.*, 1995). Female *P. martensii* are silent during courtship and spawning (Torricelli *et al.*, 1986). The sound-producing mechanism is unknown.

Lugli and Fine (2003) measured ambient noise (AN) and sound propagation in a small stony stream inhabited by this species. They found a quiet window around 100 Hz in the noise spectrum at many noisy sites (e.g., near small waterfalls, riffles, etc.). The window lies between low-frequency turbulent noise and higher-frequency noise from bubbles associated with breaking water (see Franz, 1959). The quiet window is also present in another stony river inhabited by the goby *Gobius nigricans* (Lugli and Fine, 2003), and similar windows are a recurrent feature of the AN spectrum also of Austrian rivers and streams (Amoser, 2007). Lugli and Fine (2003) found sound propagation in the stream was limited to only a few decimeters from the source because of low water depths (15–50 cm). Lugli *et al.* (2003) also found a match between stream AN, sound frequency spectrum and hearing sensitivity of the goby. The main frequencies of the sound, and the most sensitive frequencies of the goby audiogram, fit within the low-frequency window/notch of the noise spectrum.These studies utilized acoustic pressure, but the goby is likely sensitive to particle motion. Its small oval-shaped gas bladder is not involved in sound production or hearing, i.e., bladder deflation does not affect sound level and spectrum or hearing range and sensitivity (Lugli *et al.*, 2003). Therefore, measurement of particle velocity for both the stream AN and *P. martensii* sounds is essential to obtain a realistic picture of their relationship and to validate conclusions about communication.

Here we measure pressure and particle velocity of stream AN and *P. martensii* sounds with the $p-u$ sensor to reexamine the relationship between the noise spectrum of the stream and the communication frequencies of the goby. Additionally, we measure propagation of *P. martensii* sounds to investigate the relationship between acoustic pressure and particle velocity close to the sound source in a complex near-field environment. Results will be relevant to the understanding of the fish acoustic communication in the presence of high AN levels and short-range sound propagation.

II. MATERIALS AND METHODS

The freshwater goby, *P. martensii*, inhabits Stream Stirone, a small hill stream in Northern Italy (Lugli *et al.*, 1992). The study site is a 2.5 km stretch of the stream with a stony bottom, water depths usually <30 cm, and low water current (under non-flood conditions). Quiet areas alternate with small waterfalls, rapids, riffles, places with high water turbulence, where water breaks the surface and clouds of air bubbles form underwater. The goby population is widespread at both quiet and noisy areas (Lugli *et al.*, 1992).

A. Ambient noise measurements

AN was measured at five representative noisy sites near small waterfalls and rapids. Although adjacent to noise sources, the probe was placed in locations with no or modest underwater flow to minimize flow induced noise (McConnell, 2003) and self-induced turbulent pressure fluctuations by the probe and mounting support (Morse Uno Ingard, 1968). AN was also measured at two quiet sites away from the waterfalls for comparison. A single noise recording was made at each location for approximately 1 min. The AN spectrum shows only minor changes with sample duration (Lugli and Fine, 2003), being remarkably stable over time (M. Lugli, 2003).

AN was measured with an underwater acoustic pressure-velocity probe (Mk.2, Acoustech©; outer ∅: 4.3 cm) containing two built-in units: a piezoelectric, omni-directional hydrophone (sensitivity: −203.1 dB re:1 V/μPa) to measure acoustic pressure (p), and a bidirectional geophone (sensitivity: 10.5 V/m/s) to measure particle velocity (u) along one axis. The geophone has resonance peaks at 17 and 28 Hz. The response of both sensors is almost flat (hydrophone: ±1 dB; geophone: ±1 V) between 60 and 800 Hz. The probe was connected to an external, battery powered two-channel preamplifier (Acoustech@). Separate output jacks of the preamplifier for pressure and velocity were connected respectively to a portable Digital Audio Tape (DAT) recorder (Casio DA-7, sampling rate: 48 kHz). The probe was mounted on an iron support anchored to the stream bottom (Fig. 1). The support holds the outside of the unit rigidly, and the inner core containing the sensors is neutrally buoyant and free to register acoustic pressure and particle velocity. The distance between the top of the probe (mounted on its supporting harware) and the bottom substrate was about 9 cm. The probe's main axis was aligned toward the noise source like waterfalls or parallel to the stream axis at quiet sites. Additionally, three orthogonal measurements were made at two sites with minimal underwater flow that were adjacent to noisy sites: facing the stream axis (x), perpendicular (y), and vertical (z). Location 1 was a quiet pool below a small waterfall (water depth 28 cm). Location 2 was a riffle with low

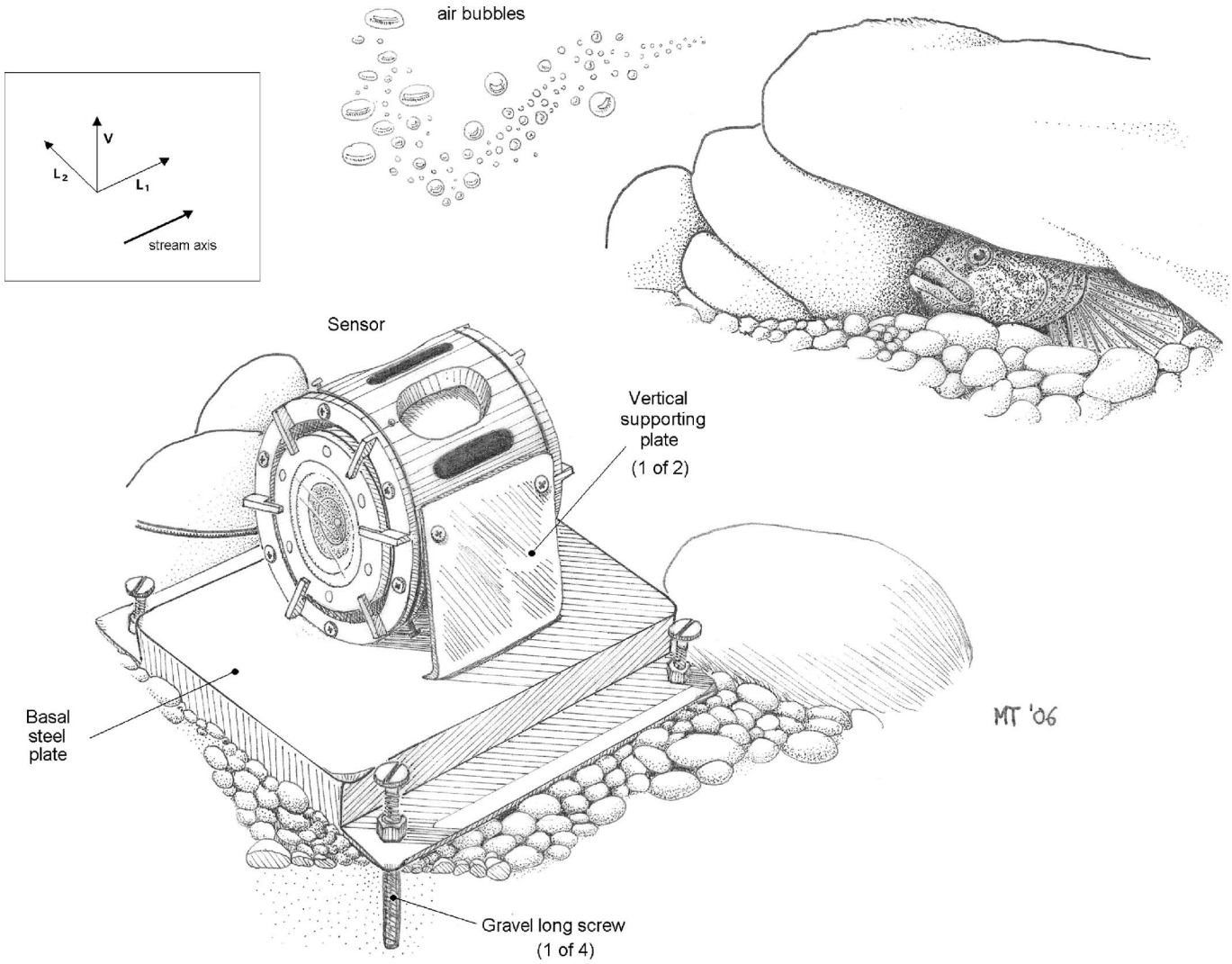


FIG. 1. Schematic of the $p-u$ probe and the mounting hardware, anchored to the stream bottom. For the measurements along three orthogonal directions (L_1, L_2, and V) (respectively x, y, and z) at two nonturbulent locations (see also Fig. 2) the mounting harware and the probe were oriented parallel to the stream axis (L_1), perpendicular to it (L_2), and vertical axis (V).

current and places of gas-bubble release (water depth of 26 cm). Therefore, these data exhibit differences from AN at noisy sites and are treated separately (see below).

Noise recordings were stored on a PC (sampling rate 44,100 Hz), and analyzed with AVISOFT© software. Recordings were band-pass filtered between 50 Hz - 1 kHz to avoid the geophone's two resonance peaks and capture frequencies important for acoustic communication (Lugli *et al.*, 1995; Lugli and Fine, 2003; Lugli *et al.*, 2003). The AN spectrum was determined for each location following Lugli and Fine (2003). Three noise segments of approximately 700 ms were randomly selected from the 1-min recording. Each segment was analyzed for the pressure spectrum level (the sound energy in 1 Hz bands of noise, dB re: 1 $\mu Pa^2/Hz$) and particle velocity spectrum level [the sound energy in 1 Hz bands of noise, dB re: 1 $(cm/s)^2$]. Levels from the continuous spectra were measured at 50 Hz intervals from 50 to 900 Hz, using the power spectrum function of AVISOFT©. Decibel values referenced to 1 V were converted to absolute measurements using the calibration factors for all components of the measuring system (i.e., probe sensitivity, gain of the preamplifier, gain of the DAT recorder). Data from the three noise segments were averaged to compute the AN pressure and particle velocity spectrum level curves.

B. Sound measurements

Spectrum shape, level and propagation of *P. martensii* sounds were also measured using the $p-u$ probe. To avoid corruption of sound measurements by AN, recordings utilized nests far from sources of elevated AN levels in places with low water current. Sound production was elicited by presenting territorial males with a ripe, conspecific female inside a small plastic-mesh cage, in front of the nest entrance. Males produced courtship sounds [i.e., tonal, pulse-train, and complex sounds, (Lugli *et al.*, 1995)]. The sensor was anchored to the substrate facing the nest entrance and the female, and multiple sounds were recorded. Five males were recorded with the probe approximately 10 cm from the nest entrance, and three were also recorded at 20, and 40 cm (the exact distance depended upon positioning the sensor because of bottom complexity). The male was then netted and

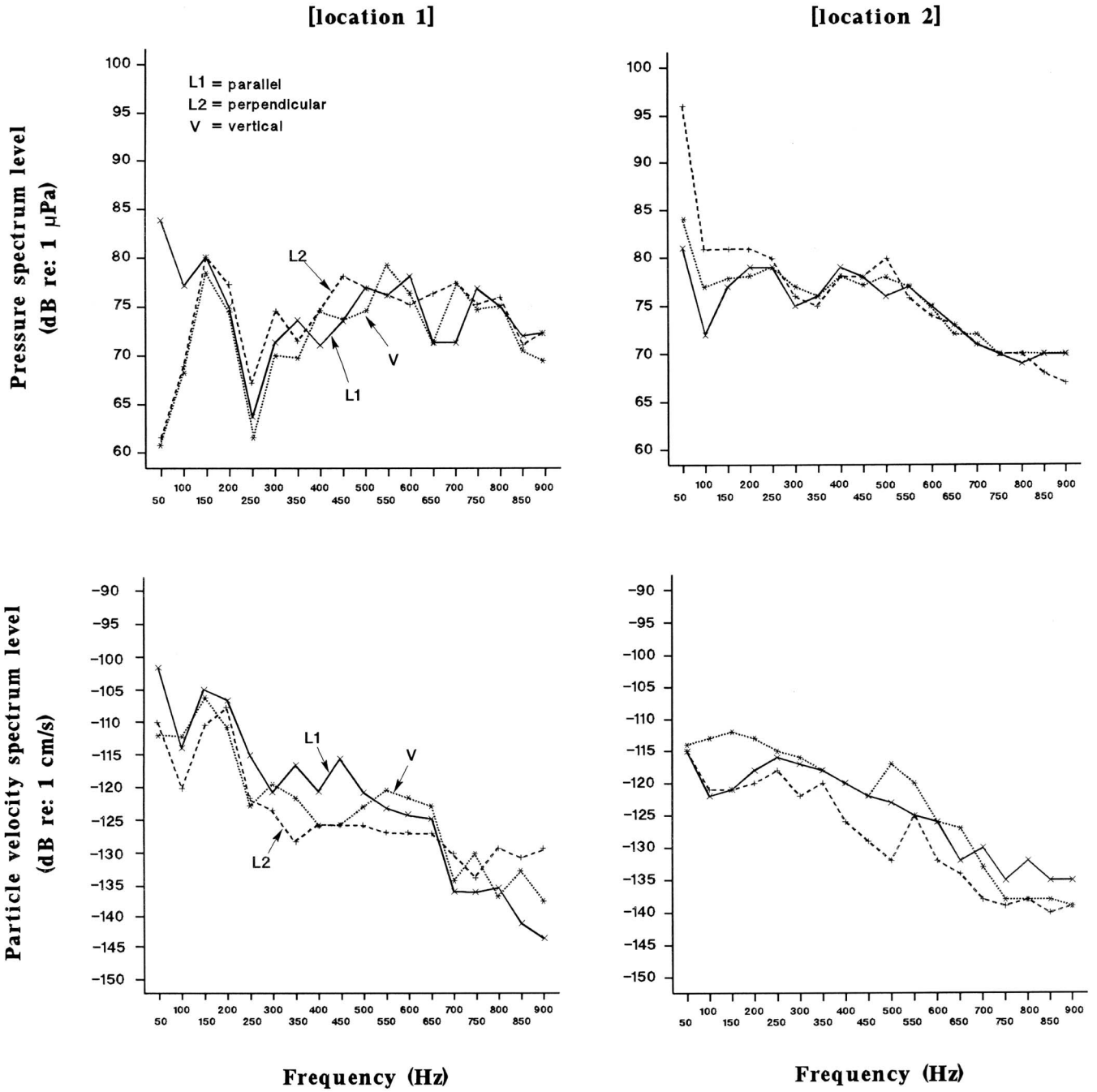


FIG. 2. Pressure (top) and particle velocity spectra (bottom) of the ambient noise at two nonturbulent locations close to sites of gas-bubble formation, measured with the sensor positioned in each of the three orthogonal directions: L_1, L_2, and V (respectively x, y, and z). Pressure levels above 100 Hz are similar in the $x-y-z$ plane, whereas velocity levels are more variable. However, in the lower part of the spectrum (i.e., at 100 Hz and below), noise level variability in the three directions is larger for pressure.

measured for total length (mm), and water temperature was measured.

Pressure and velocity spectra and amplitudes were determined for 5 sounds with the best signal-to-noise (S/N) ratio in the pressure modality for each male at 10 cm from the nest. Twenty-five sounds (see the following) were high-pass filtered at 50 Hz and analyzed (FFT length: 512 Hz, bandwidth: 15 Hz, Hamming window). The frequency spectrum of the *P. martensii* sounds resolves into few (tonal sound) to many (pulse-train sound) harmonic bands between 50 and 900 Hz (Lugli *et al.*, 1995, 2003). To compare the level and frequency distribution of the *P. martensii* sound spectrum with stream AN spectrum, the continuous p and u spectra of each sound were partitioned into seventeen 50 Hz frequency bands. Each band was examined for the presence of harmonics. Amplitude and peak (=center) frequency of all harmonics were measured, i.e., a harmonic band with peak frequency at 320 Hz was assigned to the 300–350 Hz class. Amplitudes were converted to absolute dB_{rms} (re:1 μPa, pressure mode; re:1 cm/s, velocity mode) (bandwidth: 10 Hz, Hamming window) using the calibration factors for all components of the measuring system. Sound spectra were

determined for 25 sounds (12 tonal and 13 pulse-train sounds), and mean p and u spectra were obtained by averaging the frequency and amplitude within each frequency band.

Sound propagation was examined by measuring the amplitude of the fundamental frequency (i.e., the pulse repetition rate of the sound; Watkins, 1967; Torricelli *et al.*, 1990) and higher harmonics, as the logarithmic root mean square pressure (SPL re: 1 μPa) or particle velocity (SL re: 1 cm/s) converted to absolute dB_{rms} (bandwidth: 10 Hz, Hamming window). Only tonal sounds were used for the analysis as pulse-train sounds were not consistently present at each distance. Levels of at least five sounds from each male at each distance were averaged to determine the transmission loss of p and u. Our measurements are an estimate of transmission loss as they do not control for variation in call level. However, call levels in a single recording typically vary by 1–4 dB standard deviations (s.d.), which is considerably less than the propagation losses observed in the field (around 20 dB/20 cm, Lugli and Fine, 2003; Fig. 9, present paper).

C. Statistical analysis

Noise spectra from the five locations, tested with Kendall's concordance test (Siegel and Castellan, 1988), demonstrated significant concordance (see Lugli and Fine, 2003 for details) and therefore averaged within each frequency band to compute mean AN spectra for p and u. Similarly, the mean spectrum of sounds of individual males was calculated to determine if differences occurred in spectral shape. Finally, the magnitude (i.e., the real part) of the p/u ratio was calculated for frequencies of both AN spectra and sounds (see Sec. III for details). A preliminary analysis of the phase relationships between p and u showed the $p-u$ shift was around −90° (the theoretical near-field value) for most sound and noise frequencies. These relationships will be considered in a forthcoming paper.

Mean p and u spectra for each individual were calculated by measuring the amplitude of the peak frequency (if present) in each 50 Hz band and averaging regardless of sound type. Not every frequency band contained significant sound energy (i.e., >3 dB above the background noise), and therefore, the number of sounds used to calculate means was usually less than the number of sounds analyzed. Mean sound spectra for p and u across males were concordant, and therefore all sounds recorded in a given modality were pooled, and the mean sound spectra for p and u were calculated regardless of sound type. The *P. martensii* sound spectrum is not affected by male size (Torricelli *et al.*, 1990).

III. RESULTS

A. Spectral characteristics of the stream ambient noise

AN spectra in the x–y–z plane for acoustic pressure (p) and particle velocity (u) exhibit similarities in shape and level at two nonturbulent locations (Fig. 2). Above 100 Hz, p levels are relatively similar in the three orthogonal directions, whereas corresponding u levels are more variable. At 50 and 100 Hz however, pressure is more variable. The p and u spectra vary considerably at location 1 but are more similar at location 2. At location 2 p and u spectra of bubble noise (i.e., frequencies above 100 Hz) peak around 250 Hz, except for the vertical u spectrum, which peaks at 150 Hz. Thereafter, noise levels fall off with increasing frequency for both spectra although the u slope is greater (~9–13 dB for p and u, respectively, between 400 and 800 Hz), i.e., the magnitude of p/u increases with frequency (Fig. 5, top graph).

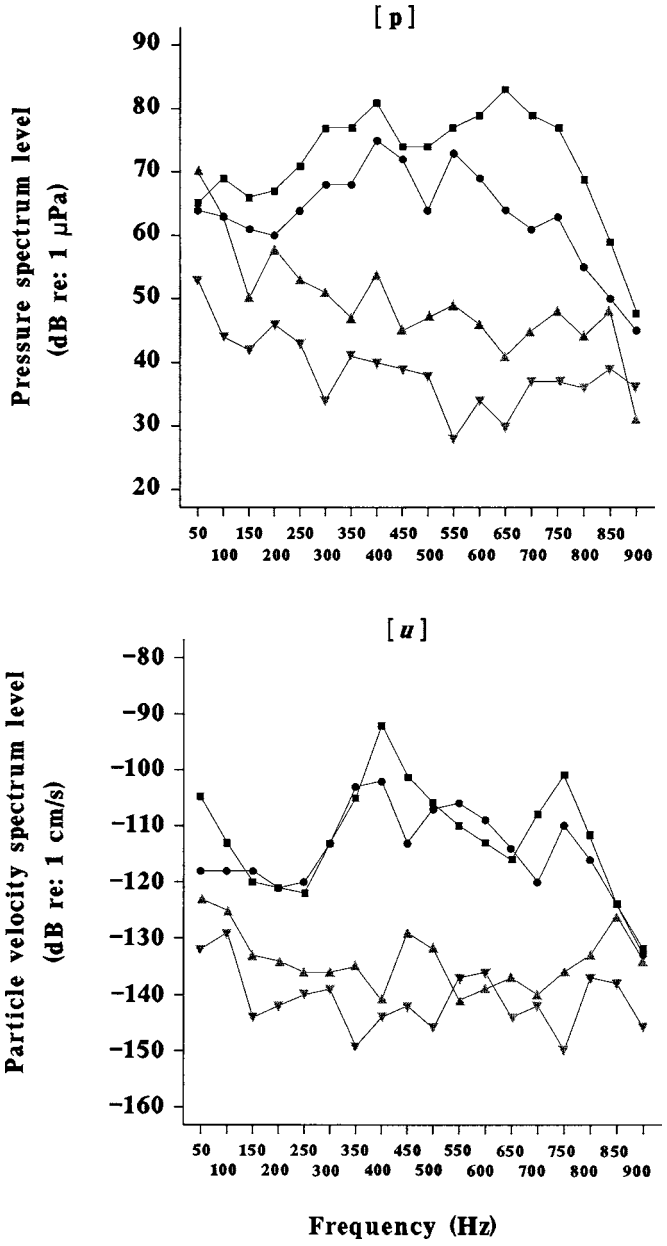


FIG. 3. The pressure spectrum (top) and the particle velocity spectrum (bottom) of the ambient noise at two noisy locations and two quiet locations (a given location is identified by the same symbol on the pressure and velocity spectra; triangles: quiet locations). Notice the region of lower noise levels at frequencies below 250 Hz is more marked on the velocity than on the pressure spectrum at the two noisy locations (i.e., the quiet window, Lugli and Fine, 2003, see also Fig. 4).

At quiet locations, AN levels for both p and u are higher at low frequencies and decrease slowly and irregularly with frequency although the u curve is somewhat flatter (Fig. 3). At noisy locations, both spectra feature a quiet window (see also Fig. 2) below 300 Hz, increased levels in the

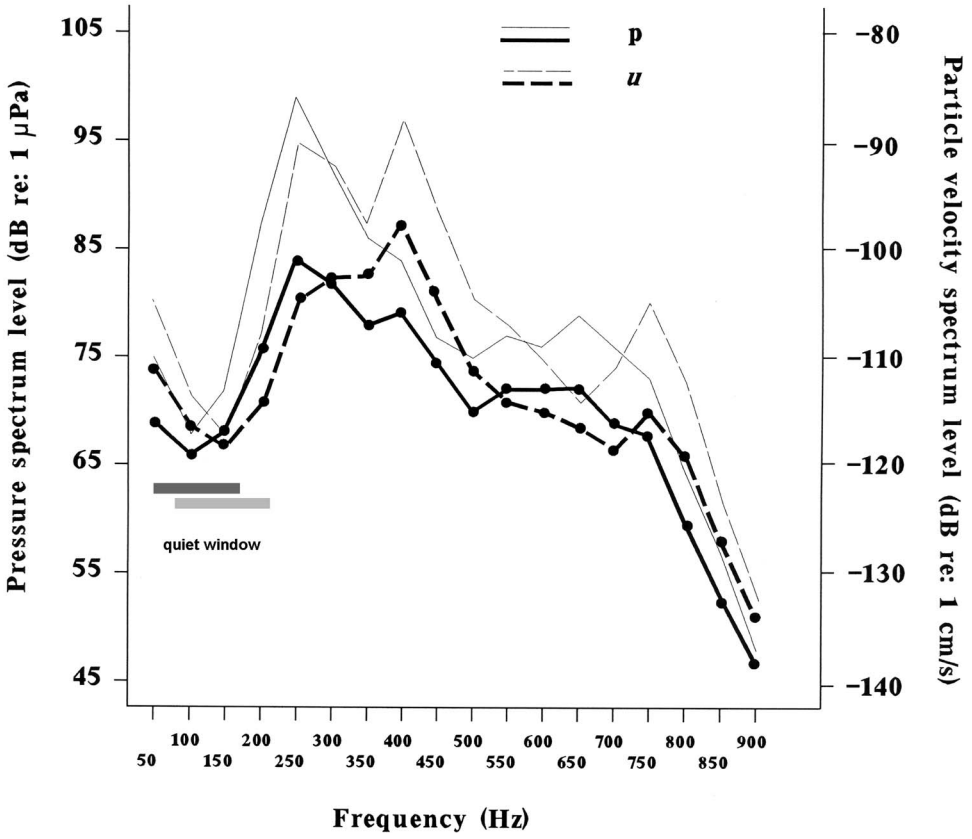


FIG. 4. Mean (thick lines) +1 s.d. (thin lines) pressure spectrum levels (left *Y* axis) and particle velocity spectrum levels (right *Y* axis) of Stream Stirone AN (noisy locations, $n=5$). The presence of lower spectrum levels at around 100 Hz (i.e. the quiet window, Lugli and Fine, 2003) is clear in both pressure (dark-gray horizontal bar) and velocity (light-gray horizontal bar) spectra. However the quiet window for particle velocity is displaced by 50 Hz, and maximal levels about 150 Hz toward higher frequencies.

300–700 Hz band, and decreasing levels at higher frequencies. *U* spectra however, have a wider window that is 10–15 dB deeper. Although *p* and *u* spectra appear similar, some peaks exhibit differences in amplitude and frequency. For instance, in Fig. 3 (square symbols), both exhibit peaks at 400 Hz, but a second peak occurs at 650 Hz for *p* and 750 for *u*.

Shifts between the *p* and *u* components are also apparent in the mean AN spectrum (Fig. 4). The quiet window exhibits minimal variability in both *p* and *u* spectra, although the *u* window is displaced from 150 to 200 Hz.

The magnitude of p/u at noisy locations (Fig. 5, bottom graph) exhibits peaks between 200 and 300 Hz and between 600 and 700 Hz, and the mean value (Fig. 5, bottom graph: thick line) peaks at 200 and 650 Hz, the two peaks reflecting the $p-u$ shift of individual AN spectra (Fig. 3) and of the mean spectrum (Fig. 4), respectively.

B. Spectral characteristics of *P. martensii* sounds

Tonal ($n=12$) and pulse-train sounds ($n=13$) have similar levels and generally similar energy distributions for *p* and *u*. In the tonal sound from male #4 (Fig. 6, top) both the *p* and *u* spectra have four harmonic bands between 180 and 800 Hz, with the largest band at the fundamental frequency of 180 Hz. The pressure peaks are distributed relatively equally between bands, but the second harmonic of the *u* spectrum is almost absent. The *u* waveform of the tonal sound in Fig. 6 builds up more slowly and decays more quickly, although this response was not present in all signals.

Differences between *p* and *u* spectra are greater in the multiharmonic pulse-train sound than in the tonal sound (Fig. 6, bottom). The fundamental frequency (around 100 Hz) is the dominant *p* band, but the second harmonic (220 Hz) is larger in the *u* spectrum. Pressure exhibits a harmonic series of three decreasing bands and a second set of peaks with a maximum at about 500 Hz, which then exponentially decays to around 800 Hz (Fig. 6, bottom). Most of the energy in the *u* spectrum is concentrated in the lower half of the spectrum (i.e., below 400 Hz), with a secondary component between 700 and 900 Hz. The greater energy at lower frequencies on the *u* spectrum of the pulse-train sound is determined by the longer period components of *u* sound pulses (Fig. 7). The pulse waveform differs considerably between *p* and *u* oscillograms, with the single pulses barely distinguishable in the *u* sound (Fig. 7: male #4).

Mean spectrum shape of the tonal sound (Fig. 8, top) is similar for *p* and *u*: energy is concentrated within a narrow band around 150 Hz and a wider band around 500 Hz. However, the position of the peak frequency within these bands differs between the spectra. The *u* spectrum also contains more energy at higher frequencies (>600 Hz), where the *p* spectrum drops off. The mean *p* and *u* spectra of the pulse-train sound exhibit greater differences than for the tonal sound (Fig. 8, bottom). As in the tonal sound spectra, both spectra of the pulse-train sound exhibit a low frequency maximum around 100 Hz, and a secondary peak at middle frequencies. However, both peaks are shifted toward lower frequency bands in the *u* spectrum, particularly the secondary one. In the *p* spectrum of the pulse-train sound, energy is concentrated within a single low frequency band (from 100 to 150 Hz), and the secondary peak is distributed over a broader frequency band, than in the *u* spectrum. Finally,

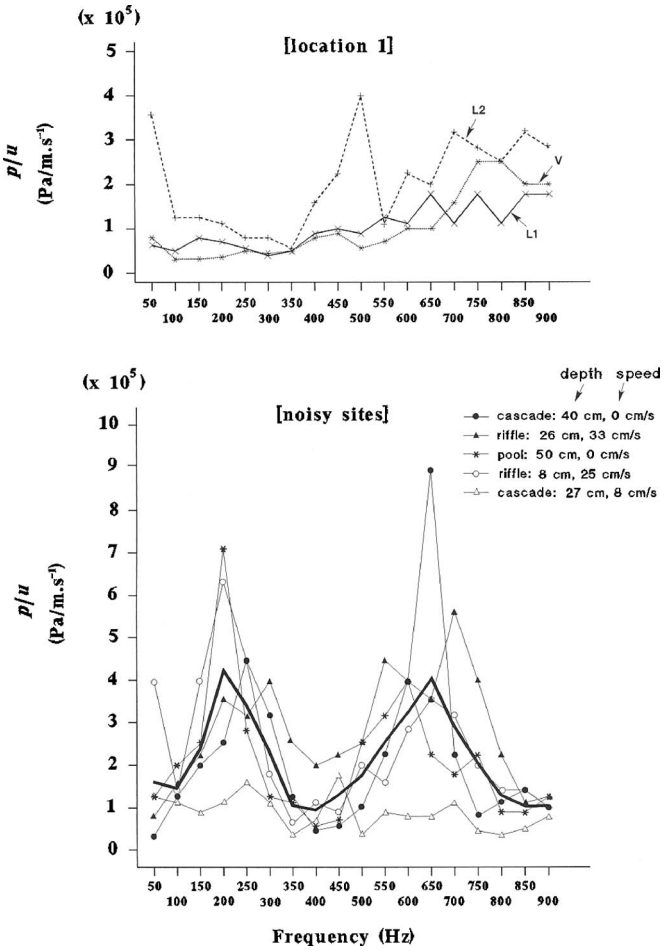


FIG. 5. Relationship of the magnitude of p/u to frequency in the three orthogonal directions at location 1 (top graph), and along an axis of AN propagation at the five noisy locations (bottom graph; CASCADE: probe in a riffle close to a cascade; RIFFLE: probe in a riffle close to places of bubble noise release; POOL: probe in a pool located below a small waterfall). The mean p/u ratio curve of the five locations (bottom graph: thick line) was computed by averaging the five values of p/u at each frequency. P/u ratios were computed by converting decibel levels of pressure and velocity into original units and computing the ratio as Pascals per meter per second.

pressure falls off above 600 Hz, whereas velocity varies by only about 3 dB from 500 to 900 Hz. Notice spectra shape of tonal and pulse-train sounds are more similar for p than u.

C. Propagation of sound pressure and particle velocity in the stream

The two nests were located over a stony bottom with low current speed and maximum water depth of 11 cm (male #1, total length=80 mm, water temperature: 27.8 °C) and 21 cm (male #4, total length=88 mm, water temperature: 26.7 °C) respectively. The tonal sound generally exhibits a large transmission loss at all frequencies for p and u (Fig. 9; exception: 400 Hz for u sounds recorded at 10 cm from male #1). Attenuation of the fundamental frequency was 6–7 dB/10 cm for p, and 7–10 dB/10 cm for u. P losses are similar to those previously recorded for this species [8 dB/10 cm from 5 to 45 cm (Lugli and Fine, 2003)]. Attenuation of p and u for other frequencies is similar to those determined for the fundamental frequency. Notice in male #1 an increase of 4 dB between 10 and 20 cm for 400 Hz for u, whereas p decreases. The magnitude of p/u for the fundamental frequency (ranging from 100 000 to 150 000 Pa/m/s for male #1 and 50 000 to 100 000 Pa/m/s for male #4) is independent of distance from the nest (Kruskal-Wallis tests: $t=3.44$, NS, male #1; $t=0.09$, NS, male #4).

D. Relationship between sound and ambient noise spectrum

The dominant frequencies of *P. martensii* sound occur within the quiet window for both p and u (Fig. 10). Both p and u levels of the sound mean spectrum peak between 100 and 150 Hz, the band with lowest AN levels for p and u in the fish's hearing range, which extends to about 700 Hz (Lugli and Fine, 2003; Lugli *et al.*, 2003). Maximal S/N ratio within the quiet window is similar (i.e., about 10 dB) for pressure and velocity. A secondary, broad p peak around 500 Hz simultaneously occurs with a drop off in the AN spectrum, which peaks at 250 Hz. The presence of a secondary peak around 450 Hz in the call's u spectrum is less clear. In addition, this peak is only 50 Hz higher than the noise peak at 400 Hz. However, this peak is considerably above the noise in the quiet u spectrum (Fig. 3) and could provide information under quiet conditions.

IV. DISCUSSION

A. Noise

Previous work demonstrates that acoustic pressure in the AN spectrum is concentrated between 300 and 700 Hz, and a quiet region exists around 100 Hz in AN spectra from most noisy sites, i.e., pools below small waterfalls, small rapids, riffles, etc. (Lugli and Fine, 2003). The window is a recurrent and stable feature of the stream AN at such places and has recently been reported in Austrian rivers and streams (Amoser, 2007). The current study demonstrates the quiet window is also present in velocity spectra at noisy but not quiet locations, and in some cases can be deeper for the u than the p spectrum (Fig. 3). Noisy sites are characterized by high water turbulence and formation of air bubbles (Lugli and Fine, 2003). Water flowing around submerged objects (and the sensor) generates noise at the lowest frequencies (i.e., below 10–20 Hz; Strasberg, 1979), and this flow-induced noise may corrupt noise measurements at higher frequencies relevant to goby communication. We minimized this effect by placing the sensor in areas with low or no water current. Therefore, most of the noise energy between 50 Hz and 1 kHz is from nearby turbulence and bubble sources. Lugli and Fine (2003) argued that the low-frequency notch/window in the AN pressure spectrum results from the combined effect of the turbulence noise falling off above infrasonic frequencies and the presence of bubble noise above 200 Hz.

Shapes of the mean p and u AN spectra at noisy sites are generally similar although the u spectrum is shifted up 50–100 Hz (i.e., a minor $p-u$ mismatch). The magnitude of the p/u ratio of the AN spectrum at noisy sites exhibits a frequency-dependent pattern consistent with the $p-u$ mismatch of the mean spectrum. The reason for the shift at turbulent sites is unclear. This pattern is not present in the x,

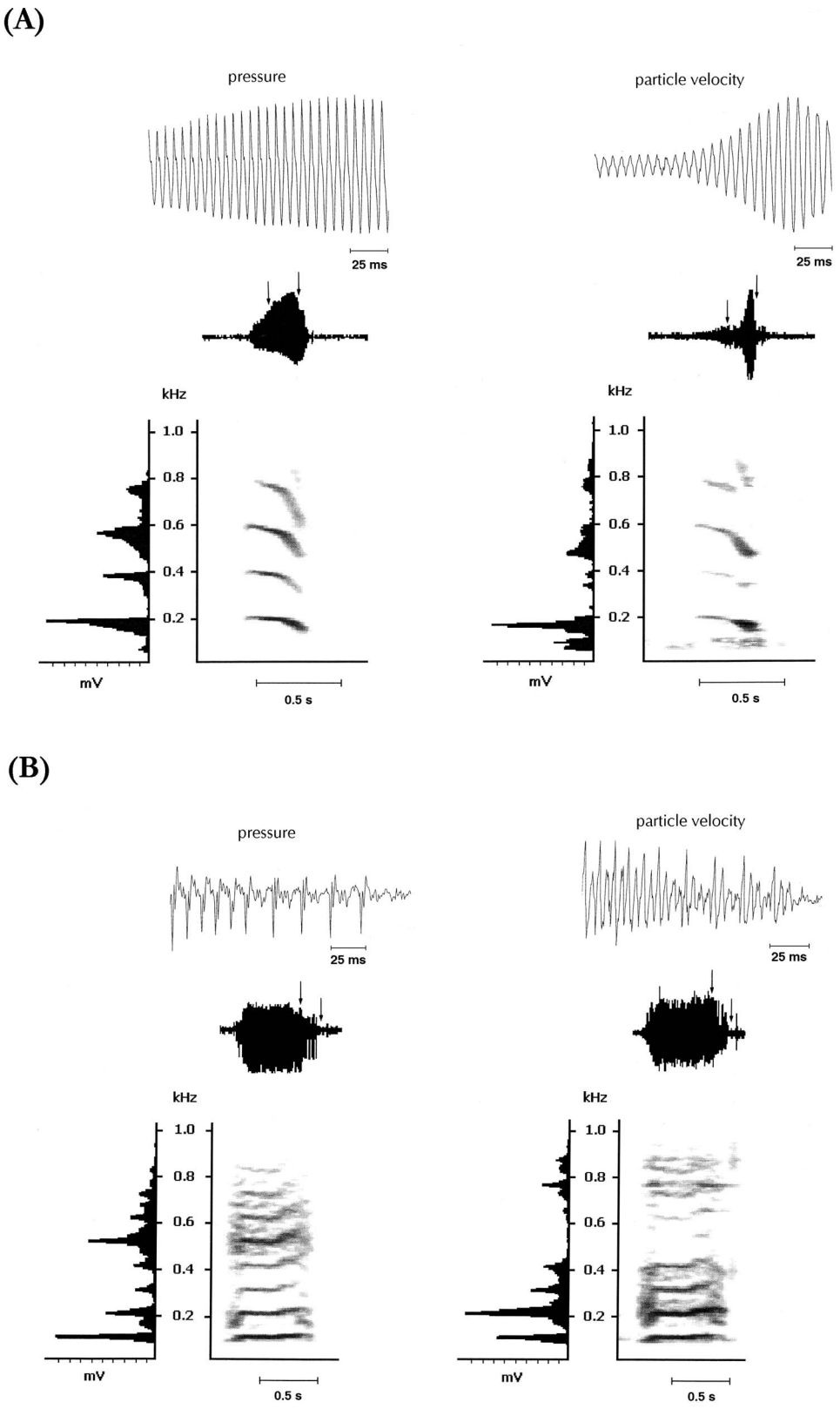


FIG. 6. Spectrogram (bottom-right, FFT=512 points, Hamming window), power spectrum (bottom-left half of panel, amplitude on a linear scale of 100 mV/division, arbitrary units), envelope (trace above the spectrogram), and oscillogram of a 150 ms section of the sound (top trace), of a representative (A) tonal sound (male #4) and (B) pulse-train sound (male #2) of *P. martensii*, recorded as pressure (left) and particle velocity (right). The expanded waveforms illustrate the portion of the sound between the two arrows. (A) Tonal sound: the pressure and particle velocity envelopes differ, but the tonal structure is clear in both cases. Notice sound energy is concentrated at the fundamental harmonic (about 200 Hz) in both spectra. The signal at about 100 Hz in the velocity sonogram is background noise. (B) Pulse-train sound: the oscillograms illustrate the end of the sound. The pulse-train waveform differs for pressure and particle velocity. Like the tonal sound, the velocity spectrum of the pulse-train sound has relatively more energy concentrated at the lower harmonics.

y, and z spectra at two nonturbulent sites (locations 1 and 2). At location 1 the p/u ratio increases with frequency regardless of sensor orientation. Theoretically, pressure–velocity ratios should increase with frequency in the nearfield of simple noise sources (Siler, 1969; Kalmijn, 1988). Simple theoretical predictions, however, may not apply to complex sources of underwater bubble noise which may interact destructively, as well as constructively, and whose energy at lower frequencies decays rapidly because of the very shallow depths (see the following). The ratio measured at noisy locations does not increase linearly with frequency but exhibits a remarkably similar pattern suggesting the presence of environment-specific relationships between p and u ambient noise close to the source.

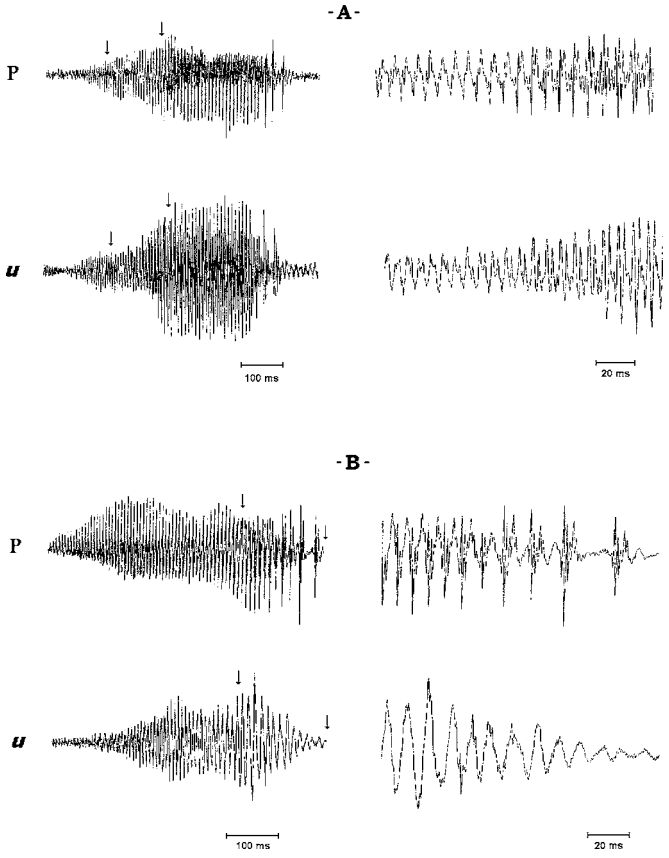


FIG. 7. Two examples of a pulse-train sound (A: male #2, recording distance of 10 cm; B: male #4, recording distance of 8 cm) illustrating the waveform of the sound (left) and an expanded portion of it (right) between the two arrows for acoustic pressure (p) and particle velocity (u). Notice the sound by male #2 is different from the selection in Fig. 6 (bottom).

B. Sounds

All male *P. martensii* emitted tonal and pulse-train sounds to a caged ripe female near the nest opening. Sounds were recorded while the male was inside the nest (a hollow underneath a flat stone), a situation typical for courtship sound production in this species (Lugli *et al.*, 1997). Spectral characteristics of pressure for tonal and pulse-train sounds were similar to those determined in earlier studies (e.g., Lugli *et al.*, 1997; Lugli and Fine, 2003), i.e., the sound energy is concentrated at the fundamental frequency [=pulse repetition rate (Watkins, 1967; Torricelli *et al.*, 1990)] around 200 Hz (tonal sound), or at the lower harmonics at around 100 Hz (pulse-train sound). Laboratory recordings of *P. martensii* courtship sounds indicate little or no energy above 0.5 kHz in either sound (e.g., Lugli *et al.*, 1997). In this study however, both types of sound had notable energy above 500 Hz (likely due to the higher recording temperature in the stream; M. Lugli, private communication), allowing us to compare pressure and particle velocity over a broader frequency range. Most energy in p and u spectra of both sounds is still below 200 Hz. The frequency distribution is similar for p and u spectra of the tonal sound, but the pulse-train sound exhibits larger differences between spectra. In several pulse-train sounds, some frequencies in the p spectrum are missing from the u spectrum, and vice-versa (Fig. 6).

Because of the complexity of the acoustic environment, reasons for these differences are speculative. Position of the emitter might affect radiation of directional and nondirectional components of the sound field. Additionally, the acoustics of the nest, a hollow under a flat stone, might affect sound propagation. Laboratory experiments indicate the nest may amplify pressure (M. Lugli, private communication). Perhaps sound is reflected from the back of the nest hollow resulting in constructive interference. Additionally, some of the sound energy could propagate through the substrate as in other bottom-dwelling species such as the mottled sculpin (A. Whang and J. Janssen, 1994) and other gobiids (Janssen, private communication). This effect however, is likely minimized in a rock-lined stream compared to a sandy or muddy bottom.

A gas bladder may change the directional properties of the sound through scattering, even if it is not involved in sound generation. Barimo and Fine (1998) found a directional field that matched the heart-shaped swim bladder in the oyster toadfish in which the swim bladder is the acoustic radiator. For the goby passive effects of the swim bladder would likely dominate at the resonant frequency of the bladder's gas cavity, and it is unlikely that the small bladder in *P. martensii* would exert large effects at various frequencies. Possible filtering effects of stones and gravel surrounding the fish nest (rocks: density approximately 2–3 times the water density) could attenuate and/or change directionality of some frequencies, thereby accounting for differences in the p and u spectra. The amplitude and frequency spectrum of croaker sounds (acoustic pressure) was not affected by enclosing a hydrophone within a terra cotta drainage tile (Barimo and Fine, 1998). However, particle velocity is more likely to be affected by stone barriers. Still, given the long wavelength of goby sounds, it is unlikely that rock positions would dramatically affect low frequency propagation, and we suggest that p and u are affected separately by destructive and constructive interference.

The tonal sound has a simpler acoustic structure than the pulse-train sound, and acoustic differences in waveform between p and u spectra are less pronounced in the tonal sound. Although both sounds attenuate rapidly, the tonal sounds has features that may allow longer distance propagation than the pulse-train sound, which is only emitted when the female is in the male's nest. For instance, the tonal sound is frequency (=pulse-rate) modulated, a feature common to acoustic signals used for long-range communication among birds and mammals (Wiley and Richards, 1982).

C. Transmission loss

Far from the sound source (i.e., in the far-field) pressure and particle velocity are in phase and decay slowly with distance from the source; furthermore, the p/u ratio is constant for all frequencies and noise-source types, being equal to the product of the density of the medium times the sound velocity (Michelsen, 1983; Rogers and Cox, 1988), i.e., the acoustic impedance of the medium, which is about 1.4×10^6 Pa/m/s in freshwater (sound velocity: 1430 m/s;

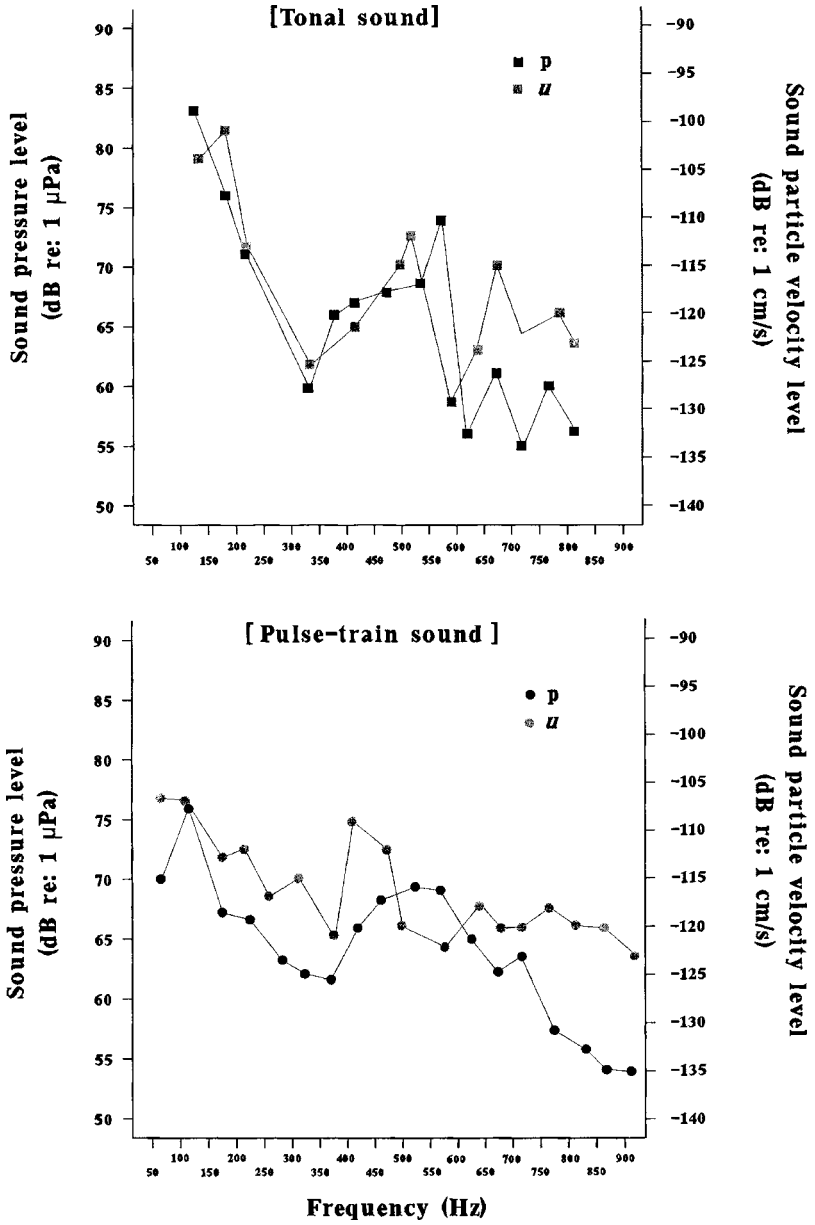


FIG. 8. Mean spectrum of the tonal sound (top, $n=12$) and the pulse-train sound (bottom, $n=13$) of male *P. martensii*. Acoustic pressure (black symbols), particle velocity (light symbols).

water density: 1000 kg/m^3). In this study recordings were conducted close to the sound source (i.e., in the extreme near field) and near the air–water–bottom interfaces at shallow depths (<40 cm). Therefore biologically relevant sound frequencies (50–700 Hz, Lugli *et al.*, 2003) will not propagate (Urik, 1983). Transmission loss is high for pressure and velocity at all frequencies, and the observed values (i.e., 6–10 dB/10 cm) are much larger than expected from spherical spreading in an acoustic free field, 6 dB loss per distance doubled, equivalent to a decay of $1/r$, r=source distance >1 m (Urik, 1983). Considering the shallow depths and cylindrical spreading of 3 dB/DD, attenuation is even more striking.

There is a tendency for p and u transmission loss to decrease at higher frequencies. The high, frequency-dependent attenuation can be predicted using the waveguide model analogy for sound transmission in small tanks with low water depths (Akamatsu *et al.*, 2002): a sound frequency below the cutoff within a circular tank whose diameter is larger than its depth. The energy of, e.g., a 200 Hz tone will be reduced by 20 dB every 15 cm for a water depth of 20 cm and 22 dB for a water depth of 30 cm. Slightly longer distances are obtained for higher harmonics (e.g., a 400 Hz tone). Thus, our propagation measurements are consistent with theoretical predictions for frequencies traveling below the cutoff in very shallow bounded bodies of water.

The magnitude of the p/u ratio for the fundamental frequency of the tonal sound (around 200 Hz) does not change with distance from the source in the near field, and the ratios are 10-fold lower than the theoretical value for the free-field situation (see Fig. 9). Again, this result is reasonable considering that sound energy decreases rapidly below the cutoff, and particle displacement predominates over sound pressure in the near field. Note Rogers and Cox (1988) caution that the p/u ratio is not a meaningful quantity in nonfree-field situations, and we use it here to permit simple comparisons

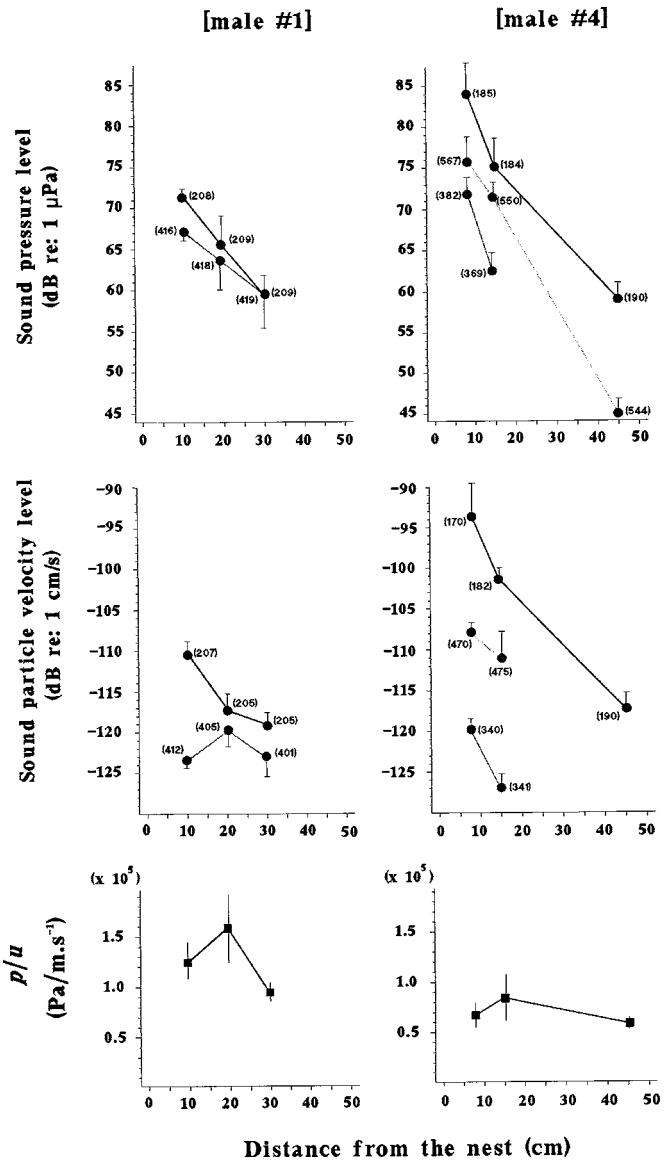


FIG. 9. Transmission loss (mean +1 s.d.) with distance (cm) for the fundamental frequency and higher harmonics of sounds by two males, recorded as pressure (top) and particle velocity (middle), and p/u ratios of the fundamental frequency (bottom). Sounds were emitted by the male at the nest entrance. Mean levels for pressure (dB re:1 μPa) and velocity (dB re:1 cm/s) of each harmonic were calculated from five sounds at each distance. Differences in fundamental frequency (and higher harmonics) between the two males are explained by water temperature (Torricelli *et al.*, 1990). Changes in the mean value of the p/u ratio with distance were determined for the fundamental frequency of sounds of the two males. P/u ratios were computed by converting decibel levels of pressure and velocity into original units and computing the ratio as Pascals per meters per second.

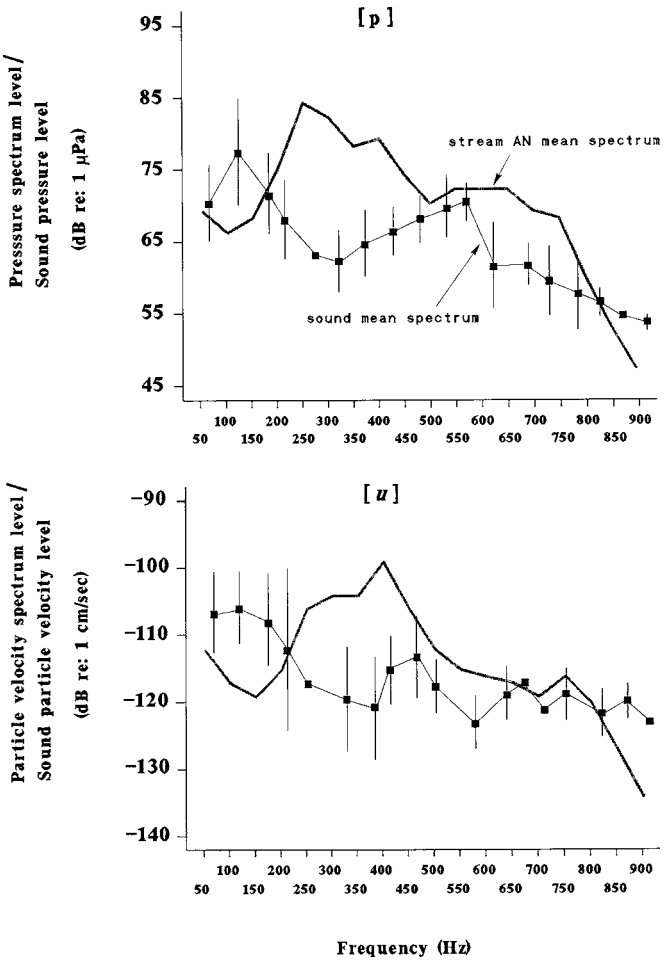


FIG. 10. The relationship between *P. martensii* mean sound spectrum (10 cm from the calling male) and the stream mean AN spectrum for sound pressure (top) and particle velocity (bottom). Mean spectrum was calculated using 30 sounds by 5 males recorded at quiet sites of the stream (i.e., far from sources of elevated AN levels, see Sec. II). Square symbols without standard error bar indicate mean values computed using $n<3$ sounds. The AN mean spectra for pressure and particle velocity are those reported in Fig. 4.

D. Relationships between sound and AN spectrum: Implications for acoustic communication and hearing in noise

By fitting the main sound frequencies into the quiet region, *P. martensii* increases the S/N ratio for the perception of sound under high AN levels, regardless of the physical stimulus used for detection (i.e., pressure or velocity). Maximum auditory sensitivity of *P. martensii* occurs between 70 and 150 Hz, i.e., within the quiet window, and thresholds are above 105 dB (re:1 μPa). Thus SPL of *P. martensii* sounds recorded 10 cm from the nest (see Fig. 10, top) should be audible only within several centimeters of the emitter. Yet, laboratory experiments indicate both sexes respond to courtship sound playback at typical levels through a speaker placed 30 cm from the test fish (Lugli *et al.*, 1996, 2004). The stream goby is likely a velocity sensitive species, and a complete understanding of signal/AN matching must also consider velocity sensitivity.

Noisy sites will impair sound detection. In the quiet region, levels at 10 cm from the nest exceed spectrum levels by 10 dB, or less, for both pressure and velocity. A 10 dB S/N ratio for p is below the minimum (generally 15–25 dB for frequencies below 1 kHz) required by teleosts for detection of tone signals under masking conditions (e.g., Popper and Fay, 1973; Fay and Megela-Simmons, 1999). Goby sounds however, are composed of multiple frequencies and thus will excite more auditory neurons than a tone of a single frequency. Torricelli *et al.* (1986) found that courting males leave the nest and approach ripe females using visual displays and courtship sounds to entice them back to the nest. Communication over truly short distances may be all that is required in this species although communication will be fa-

cilitated under noisy conditions by the quiet window. Detection distances would of course increase in quiet regions.

E. Pressure versus particle motion measurements: suggestions for future studies on fish bioacoustics

P and *u* components of stream AN and for fish sounds may, or may not, track each other for single measurements although mean spectra are more similar. The presence of the low-frequency quiet window is clear in both *p* and *u* AN spectra as is the concentration of the energy at the fundamental in *p* and *u* spectra of the tonal sound. Therefore SPL pressure measurements, either for environmental noise or sounds emitted by a particle-sensitive teleost are likely relevant for characterization of the dominant frequencies used for communication in the *u* dimension. However, the peak values of particle velocity may be under- or overestimated on the pressure spectrum. These conclusions are drawn for frequencies below the cutoff (i.e., under conditions of non propagation of the sound wave). Further investigations are needed to validate them in the case of a propagating sound wave.

ACKNOWLEDGMENTS

This study was supported by grants from University of Parma (FIL 2005). The authors are grateful to Patrizia Torricelli (University of Venice) for providing the sensor, Elisa Sbarbada, for help in data collection and analysis, Mario Toledo for the drawing in Fig. 1, and Tim Cameron for acoustic expertise and comments on the results.

Akamatsu, T., Okumura, T., Novarini, N., and Yan, H. Y. (**2002**). "Empirical refinements applicable to the recording of fish sounds in small tanks," J. Acoust. Soc. Am. **112**, 3073–3082.

Amoser, S. (**2007**). "Acoustic communication in fishes: Influence of the habitat on sound detection," Thesis Dissertation, University of Vienna, Vienna, Austria.

Barimo, J. F., and Fine, M. L. (**1998**). "Relationship of swim-bladder shape to the directionality pattern of underwater sound in the oyster toadfish," Can. J. Zool. **76**, 134–143.

Bastyr, K. J., Lauchle, G. C., and Mc Connell, J. A. (**1999**). "Development of a velocity underwater acoustic intensity sensor," J. Acoust. Soc. Am. **106**, 3178–3188.

Fay, R. R., and Megela-Simmons, A. (**1999**). "The sense of hearing in fishes and anphibians," in *Comparative Hearing: Fish and Anphibians*, edited by R. R. Fay and A. N. Popper (Springer, New York), pp. 269–318.

Franz, G. J. (**1959**). "Splashes as sources of sound in liquids," J. Acoust. Soc. Am. **31**, 1080–1096.

Horch, K., and Salmon, M. (**1973**). "Adaptations to the acoustic environment by the Squirrelfishes *Myripristis violaceus* and *M. Pralinius*," Mar. Behav. and Physiol. **2**, 121–139.

Kalmijn, Ad. J. (**1988**). "Hydrodynamic and acoustic field detection," in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer, New York), pp. 83–130.

Lugli, M., Bobbio, L., Torricelli, P., and Gandolfi, G. (**1992**). "Breeding ecology and male spawning success in two hill-stream populations of the freshwater goby, *Padogobius martensii*," Environ. Biol. Fishes **35**, 37–48.

Lugli, M., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: ambient noise and short-range propagation in shallow streams," J. Acoust. Soc. Am. **114**, 512–521.

Lugli, M., Pavan, G., and Torricelli, P. (**1996**). "The importance of breeding vocalizations for mate attraction in a freshwater goby with composite sound repertoire," Ethol. Ecol. Evol. **8**, 343–351.

Lugli, M., Pavan, G., and Torricelli, P. (**2004**). "The response of the male freshwater goby to natural and synthetic male courtship sound playback following exposure to different female sexual stimuli," Ethol. Ecol. Evol. **16**, 55–70.

Lugli, M., Pavan, G., Torricelli, P., and Bobbio, L. (**1995**). "Spawning vocalizations in male freshwater gobies (Pisces, Gobiidae)," Environ. Biol. Fishes **43**, 219–231.

Lugli, M., Torricelli, P., Pavan, G., and Mainardi, D. (**1997**). "Sound production during courtship and spawning in freshwater gobies (Pisces, Gobiidae)," Mar. Freshw. Behav. Physiol. **29**, 109–126.

Lugli, M., Yan, H. Y., and Fine, M. (**2003**). "Acoustic communication in two freshwater gobies: The relationship between ambient noise, hearing thresholds and sound spectrum," J. Comp. Physiol. [A] **189**, 309–320.

McConnell, A. J. (**2003**). "Analysis of a compliantly suspended acoustic velocity sensor," J. Acoust. Soc. Am. **113**, 1395–1405.

Michelsen, A. (**1983**). "Biophysical basis of sound communication," in *Bioacoustics: A Comparative Approach*, edited by B. Lewis (Academic, London), pp. 3–38.

Morse, P. M., and Uno Ingard, K. (**1968**). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ).

Popper, A. N., and Fay, R. R. (**1973**). "Sound detection and processing by teleost fishes: A critical review," J. Acoust. Soc. Am. **53**, 1515–1529.

Rogers, P. H., and Cox, M. (**1988**). "Underwater sounds as a biological stimulus," in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer, New York), pp. 131–149.

Siegel, S. and Castellan, N. J., Jr. (**1988**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw–Hill, New York).

Siler, W. (**1969**). "Near- and farfields in a marine environment," J. Acoust. Soc. Am. **46**, 483–484.

Strasberg, M. (**1979**). "Nonacoustic noise interface in measurements of infrasonic ambient noise," J. Acoust. Soc. Am. **66**, 1487–1493.

Torricelli, P., and Romani, R. (**1986**). "Sound production in the Italian freshwater goby, *Padogobius martensii*," Copeia **1**, 213–216.

Torricelli, P., Lugli, M., and Gandolfi, G. (**1986**). "A quantitative analysis of the occurrence of visual and acoustic displays during the courtship in the freshwater goby, *Padogobius martensi* (Gunther, 1861) (Pisces, Gobiidae)," Boll. Zool. **53**, 85–89.

Torricelli, P., Lugli, M., and Pavan, G. (**1990**). "Analysis of sounds produced by male *Padogobius martensi* (Pisces, Gobiidae) and factors affecting their structural properties," Bioacoustics **2**, 261–275.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula Publishing, Los Altos, CA).

Watkins, W. A. (**1967**). "The harmonic interval: fact or artifact in spectral analysis of pulse trains," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, New York), vol. **2**, pp. 15–43.

Whang, A., and Janssen, J. (**1994**). "Sound production through the substrate during reproduction in the mottled sculpin, *Cottus bairdi* (Cottidae)," Envir. Biol. Fish. **40**, 141–148.

Wiley, R. H., and Richards, D. G. (**1982**). "Adaptations for acoustic communication in birds: sound transmission and signal detection," in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), Vol. **1**, pp. 131–278.

The social vocalization repertoire of east Australian migrating humpback whales (*Megaptera novaeangliae*)

Rebecca A. Dunlop[a] and Michael J. Noad
School of Veterinary Science, University of Queensland, St Lucia, Qld 4072, Australia

Douglas H. Cato
Defense Science and Technology Organization, PO Box 44, Pyrmont, NSW 2009, Australia, and University of Sydney Institute of Marine Science, Sydney, NSW 2006, Australia

Dale Stokes
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92037, USA

(Received 18 May 2007; revised 1 August 2007; accepted 10 August 2007)

Although the songs of humpback whales have been extensively studied, other vocalizations and percussive sounds, referred to as "social sounds," have received little attention. This study presents the social vocalization repertoire of migrating east Australian humpback whales from a sample of 660 sounds recorded from 61 groups of varying composition, over three years. The social vocalization repertoire of humpback whales was much larger than previously described with a total of 34 separate call types classified aurally and by spectrographic analysis as well as statistically. Of these, 21 call types were the same as units of the song current at the time of recording but used individually instead of as part of the song sequence, while the other 13 calls were stable over the three years of the study and were not part of the song. This study provides a catalog of sounds that can be used as a basis for future studies. It is an essential first step in determining the function, contextual use and cultural transmission of humpback social vocalizations.
 [DOI: 10.1121/1.2783115]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.30.Nb [WWA] Pages: 2893–2905

I. INTRODUCTION

Acoustic signaling is often significantly more effective than using only visual cues in the marine environment because sound is attenuated far less than light in that medium. Humpback whales (*Megaptera novaeangliae*) in particular use many acoustic signals of which the long, complex, stereotyped, repetitive "song" is the best studied (Payne and McVay, 1971). Song units are highly variable in frequency range and usually lie between 100 Hz and 4 kHz (Tyack and Clark, 2000). However, other studies of humpback whale song have found fundamental frequencies of units as low as 30 Hz (Payne and Payne, 1985), and others with harmonics extending beyond 24 kHz (Au *et al.*, 2006).

In addition to song, humpback whales produce another set of sounds known as "social sounds" which include surface-generated percussive sounds (e.g., breaches, pectoral flipper slaps, tail slaps) and social vocalizations. For the purpose of this study, "vocalizations" are any sounds that are produced internally by the animal and include sounds that may be generated from the blow hole. These vocalizations are not produced in the highly complex structural format of the humpback song and many of the social vocalization units are not heard in the characteristic breeding population song. Song is only produced by males, whereas social vocalizations are produced by both males and females (Winn *et al.*, 1979; Mobley *et al.*, 1988). However, other reports have suggested that social vocalizations are rarely heard from lone humpback whales or mother and calf pairs (Tyack 1983; Silber 1986). Social sounds are commonly produced on the feeding grounds (Thompson *et al.*, 1977; Jurasz and Jurasz, 1979; D'Vincent *et al.* 1985; Mobley *et al.* 1988; Sharpe *et al.* 1998) and breeding grounds (Silber 1986) though these sounds are yet to be reported in migrating humpback whales.

Quantification of a species' acoustic repertoire is an essential first step before investigating its function, contextual use, geographic variation and cultural transmission. However, a comprehensive catalogue of humpback whale social vocalizations has not yet been produced. Categorizing and cataloguing the vocalization repertoire of any marine mammal is difficult. Studies on dolphin vocalizations have attempted to statistically quantify the acoustic repertoire using multivariate techniques such as principal components analysis (PCA) and discriminant function analysis (DFA) (e.g. Biosseau 2005) to eliminate subjectivity. However, the wide variation both within each sound type and between different sounds made classification difficult. Clark (1982) proposed a repertoire description for southern right whales (*Eubalaena australis*), which included vocal and nonvocal calls. In it, sounds were divided into eight broad categories (one of which was "slaps") using spectrographic and aural properties, but there was no individual call analysis or description in this study. Humpback calls are also highly variable and acoustically complex making them difficult to classify objectively. Chabot (1988) proposed a quantitative technique to classify humpback song sounds based on a clustering technique on binary data, in an attempt to eliminate the subjec-

[a]Author to whom correspondence should be addressed. Electronic mail: r.dunlop@uq.edu.au

tivity of aural classification. Other studies have attempted to quantitatively compare humpback whale song units over time and measure the variability within these units (e.g., Mercado *et al.*, 2005). In this study, specific frequency and temporal variables within the song units were selected and measured to obtain an objective characterization of their acoustic features. One of the few studies on humpback social vocalizations associated with feeding aurally categorized these sounds into "moans," "grunts," "pulse-trains," "blowhole-associated sounds" and "surface impacts" (Thompson *et al.*, 1986), making a total of five different sound types. However, no statistical methods were applied to categorize them.

In all of the above studies, broad categories such as "tonal," "noisy" and "pulsive" were used to classify sounds. Placing sounds into broad categories such as these has its problems, for example, sounds may form a structured continuum rather than fall into discrete categories resulting in a misinterpretation of the animal's communication system (Tyack, 1997). Clark (1983) found that southern right whale calls did indeed form an acoustic continuum consisting of two functional subdivisions; a set of discrete frequency-modulated calls and a set of highly variable intergraded signals. Therefore, placing sounds into such broad categories may not be a biologically relevant way of classifying sounds; rather sounds should be analyzed as discrete and separate call types.

The east Australian population of humpback whales (part of the Group E metapopulation; IWC 2005) migrates annually along the east Australian coastline between feeding grounds in the Antarctic and breeding grounds inside the Great Barrier Reef off central and northern Queensland (Chittleborough, 1965; Dawbin, 1966). Humpbacks were acoustically recorded for three years during the southward migration, which occurs from August to November, with a peak in early October at the latitude of our study area. The aims of this paper are to (i) describe and quantify the whales' nonsong social vocalizations and (ii) classify and categorize these using multivariate techniques to propose a social vocalization repertoire of east Australian migrating humpback whales.

II. METHODS

A. Visual and acoustic data

1. Visual data collection

Data were collected during the Humpback Whale Acoustic Research Collaboration (HARC) Project, in September/October 2002–2004 during the humpback whale southward migration (for detailed methods see Noad *et al.*, 2004). The study site was located at Peregian Beach, 150 km north of Brisbane, on the east coast of Australia (26°29'S, 153°06'E) and about 400 km south of the Great Barrier Reef (Fig. 1(a)). Land-based behavioral observations were collected on a daily basis (7 am–5 pm, weather permitting) from an elevated survey point, Emu Mountain (73 m), adjacent to the coast (Fig. 1(b)). A theodolite (Leica TM 1100) was connected to a notebook computer running CYCLOPES software (E. Kniest, Univ. Newcastle, Australia) and used to track and observe passing whales. CYCLOPES recorded the positions of whales from the theodolite elevation and azimuth in real time and these were annotated with observed behaviors and group compositions (e.g., adult and calf, two adults, etc.) out to the 10 km limit of the study area.

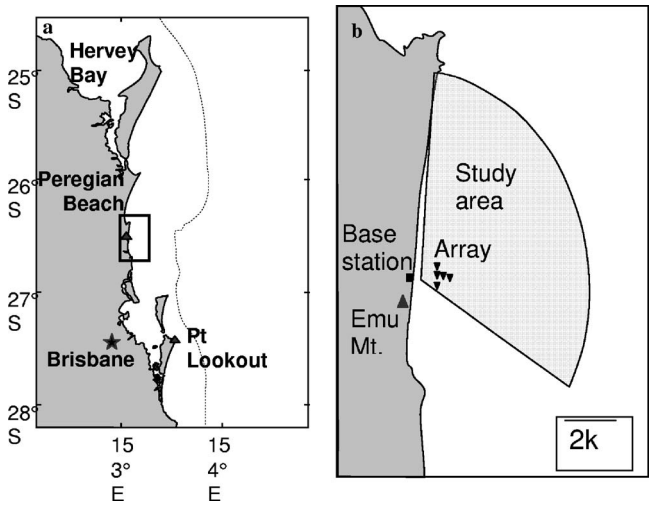


FIG. 1. Map showing the position of the study site at Peregian Beach, southeast Queensland, Australia (Fig. 1(a)) and (Fig. 1(b)) the arrangement of the base station, visual station (Emu Mountain) and hydrophone array within the study site (see inset, Fig. 1(a)).

2. Acoustic data collection

Acoustic recordings of both humpback song and humpback social sounds were made from five hydrophone-buoy systems anchored in 18–28 m of water and arranged in a T shape. Each hydrophone buoy consisted of a surface buoy with attached solar panel, a preamplifier (+20 dB) and VHF radio transmitter. At the seabed, each buoy was held in place by a concrete mooring. A High Tech HTI-96-MIN hydrophone with built-in +40 dB preamplifier was attached to a separate subsurface float a meter off the bottom and cabled to the surface buoy. Buoys 1–3 were 1.5 km from the beach and parallel to the shoreline and approximately 0.7 km apart. Buoys 4 and 5 extended seaward from buoy 2 in a line perpendicular to the shore and were approximately 0.5 km apart (Fig. 1(b)). Accurate positions of hydrophones were obtained by a careful shore-based theodolite survey of the surface buoys (Noad *et al.*, 2004).

3. Acoustic tracking

Radio transmissions from the buoys are received at a base station just behind the beach (Fig. 1(b)) on a large, vertically orientated Yagi antenna attached to a four channel, low noise, VHF receiver (type 8101) and (in 2003) a Winradio receiver. Two computers equipped with National Instruments E-series data acquisition cards and using ISHMAEL software (Mellinger, 2001) were used to record the acoustic signals received from the hydrophone buoys (one computer) and track the sound sources in real time using the arrival time differences of the sounds as they are received at the various hydrophones (the other computer). On each channel digital recordings were made at 22.05 kHz sampling rate. Tracking was performed either in the field in real time, si-

multaneously with the theodolite tracking, or during postfield analysis. Although the bearings to the sound sources calculated by ISHMAEL were accurate, there were small errors in the range estimates. At a range of 2 km from the array, the range error for a single sound position was less than 50 m. At 10 km from the array, the error was less than 1 km. However, sound location accuracy was significantly improved by using the mean position of several estimates calculated over a brief period. The array was validated by comparing acoustically calculated positions with theodolite positions of visually identified singers within the study area (for further validation experiments see Noad *et al.*, 2004).

Acoustic tracks of vocalizing whales were overlaid on the visual tracking map in CYCLOPES and the combined acoustic/visual data was shared between the base and hilltop stations using a wireless network. This provided almost real time superposition of acoustic and visual tracks out to the 10 km limit of the study area. There were rarely more than six groups migrating through the 10-km-radius study area at any one time, and these were usually widely dispersed, unless a joining interaction between two groups was occurring. Given the accuracy of the system and the way in which groups could be simultaneously visually and acoustically tracked in real time, there was no doubt as to which groups were vocalizing at any time. Within groups, however, it was not possible to determine which animal was vocalizing.

B. Data analysis

1. Spectrographic analysis

Humpback whale vocal sounds (surface-generated sounds are not included in this analysis) with average/good signal-to-noise ratios (vocalizations in which the signal was at least 10 dB higher than the background noise) were extracted from recordings for further analysis. Every pod analyzed was represented by a minimum of five vocalizations. Spectrograms of vocalizations were produced either using Spectrogram 14 (R. Horn, Visualization Software) with 4096 point fast Fourier transforms (FFTs) and 5.4 Hz frequency resolution or Raven 1.2 (Cornell Lab of Ornothology) with 1024 point FFTs, Hamming window, 21.5 Hz resolution and 75% overlap. The higher frequency resolution spectrograms tended to be used for the lower frequency sounds.

2. Vocalization classification and statistical analysis of vocalization type

Vocalizations were classified into a series of vocalization types based on aural and spectrographic characteristics. A series of variables, listed in Table I, were measured from the spectrographic samples of each vocalization for statistical analysis. The variables measured included the vocalization duration, the number of vocalizations in a bout for repetitive sounds, the number of inflections, the frequency of the spectral peak and frequency properties. Measurements of the lowest frequency component of the vocalization (the fundamental in harmonic sounds) were: start and end frequencies, minimum and maximum frequencies, ratio of start to end frequency (frequency trend ratio), ratio of maximum to minimum frequency (frequency range ratio), and the percentage of duration to the maximum frequency. Ratios of frequencies were measured rather than the differences since ratios better match mammal perception of frequency differences (Richardson *et al.*, 1995). All frequency measurements were initially made on a linear scale (shown in Table II and Table III). They were then converted to a logarithmic scale for analysis, as this better matches the perception of pitch (Richardson *et al.*, 1995). The measured variables were used in a principal component analysis (PCA) which can describe the variation of a set of multivariate data in terms of sets of uncorrelated factors, each of which is a particular linear combination of the original variables (Pearson 1901). This reduced the variables into a number of factors that were used in place of the original variables during some subsequent analyses. A varimax rotation was applied to maximize loading scores and facilitate interpretation of the data. These factors and the original measured variables were used in discriminant function analysis (DFA) with cross validation to determine the probability of sounds being correctly classified to each of the possible groups determined by aural and spectrographic characteristics.

TABLE I. Measurements and a description of the measurements made on all vocalizations. Some of the measurements are illustrated in Fig. 2.

Measurement	Abbreviation	Description
Variables measured on lowest frequency component F_0		
Duration (s)	Dur F_0	Vocalization length
Minimum frequency (Hz)	Min F_0	Minimum frequency
Maximum frequency (Hz)	Max F_0	Maximum frequency
Start frequency (Hz)	Start F_0	Start frequency
End frequency (Hz)	End F_0	End frequency
Percentage to maximum (%)	%	Percentage of duration to the maximum frequency
Frequency range (as ratio)	Range F_0	Max freq/min freq
Frequency trend (as ratio)	Trend F_0	Start freq/end freq
Inflections	Inflec	Number of reversals in slope
Other variables measured		
Peak frequency (Hz)	Peak F	Frequency of the spectral peak
Number of units per bout	No./bout	Equals one for nonrepetitive sounds
Pulse rate (/s)	Pulse rate	for pulsative sounds

III. RESULTS

Acoustic recordings of 61 vocal pods were made between 2002 and 2004 and from these 660 social vocalizations were extracted for analysis. Group compositions from which sounds were recorded included singletons ($n=15$), mother and calf pairs ($n=12$), pairs of adults ($n=8$), mother, calf and escort ($n=18$), three adults ($n=3$), four adults ($n=1$) and mother, calf and three escorts ($n=4$). All escorts were assumed to be males (Glockner and Venus, 1983; Clapham *et al.*, 1992). The sexes of adults other than moth-

TABLE II. Mean (SD) spectrogram parameters (see Table I) of measured low-frequency, amplitude-modulated, noisy, complex and repetitive sounds, the number measured of each vocalization and the number of groups in which they were heard (see Figs. 2–4 for associated spectrograms). All frequency measurements are shown in the linear scale (Hz). Social vocalizations which were also part of the song structure are highlighted along with the song year of which they were part.

	Wop	Thwop	Snort	Grumble	Sigh	Siren	Short moan	Horn	Violin	Groan	Ascend moan	Mod moan	Cry	Mod cry	Ascend cry	Trumpet	Ascend shriek	Des shriek	Squeak
N (sounds)	64	67	43	73	2	5	7	18	12	12	4	13	8	2	8	27	2	7	6
N (groups)	24	15	17	20	1	1	5	5	4	5	2	5	2	1	1	6	1	3	3
Dur F_0	0.748 (0.196)	0.949 (0.244)	0.284 (0.107)	1.509 (1.362)	2.533 (0.901)	2.309 (0.584)	0.353 (0.060)	0.365 (0.126)	0.312 (0.057)	1.262 (0.630)	1.854 (0.510)	1.212 (0.418)	1.623 (0.194)	2.204 (0.517)	1.773 (0.119)	0.732 (0.345)	1.252 (0.021)	2.297 (1.192)	0.213 (0.132)
Min F_0	43 (15)	42.2 (13)	51 (25)	43.6 (12)	30 (10)	95 (33)	161 (34)	180 (51)	548 (185)	139 (39)	162 (24)	201 (24)	338 (138)	524 (193)	171 (46)	231 (80)	714 (0)	809 (84)	1431 (302)
Max F_0	73 (16)	83.1 (20)	53 (25)	48.5 (15)	327 (23)	337 (48)	206 (60)	430 (116)	908 (239)	172.6 (50)	271 (29)	618 (182)	709 (202)	1415 (32)	437 (68)	854 (124)	2268 (20)	1373 (138)	2420 (746)
Start F_0	52 (34)	43.6 (18)	53 (26)	44.2 (14)	183 (226)	162 (79)	168 (37)	180 (51)	884 (246)	156.7 (54)	162 (24)	206 (31)	352 (163)	534 (179)	259 (41)	234 (77)	714 (0)	889 (107)	1649 (506)
End F_0	73 (16)	83.0 (20)	52 (25)	46.7 (15)	174 (193)	272 (78)	199 (64)	401 (133)	616 (256)	160.0 (41)	233 (36)	610 (188)	627 (226)	1415 (32)	406 (79)	836 (135)	2268 (20)	932 (90)	2227 (897)
% duration	100	100	7 (25)	26.0 (44)	0	9.5 (2)	59 (43)	100 (0)	8 (29)	53.1 (31)	100 (0)	64 (35)	21 (10)	100 (0)	13 (1)	77 (35)	17 (3)	16 (15)	55 (44)
Peak F	80 (19)	94.6 (27)	84 (41)	62 (24)	86 (7)	168 (18)	178 (55)	302 (32)	700 (152)	156.4 (41)	229 (38)	364 (50)	516 (89)	588 (458)	354 (29)	471 (140)	953 (8)	1083 (239)	1478 (265)
Inflect	1	2	0	0	0	6	0	1	0	0	0	2	0	2	3	0	0	1	0
No./ bout																			
Pulse rate (/s)																			
Song unit yr	No	No	No	No	02	02	02	02	02	02, 03, 04	02, 03, 04	02, 03, 04	No	No	No	02	02, 03, 04	02	02, 03, 04

TABLE III. Mean (SD) spectrogram parameters (see Table I) of measured repetitive and harmonic sounds, the number measured of each vocalization type and the number of pods they were heard in (see Figs. 5–7 for associated spectrograms). All frequency measurements are shown are in the linear scale (Hz). Social vocalizations which were also part of the song structure are highlighted along with the song year in which they were heard.

	Purr	Growl	Trill	Blow	Bark	Bellow	Creek	Screech	Scream	Grunts	Croaks	Yap	Yelps	Pulses	Low yap
N (sounds)	16	14	6	6	50	4	4	4	8	90	17	10	21	25	5
N (groups)	4	5	3	3	13	4	1	1	3	9	3	4	1	2	1
Dur F_0	1.566 (0.575)	2.253 (1.268)	2.194 (0.468)	0.281 (0.152)	0.205 (0.119)	0.939 (0.415)	1.723 (0.614)	1.366 (0.336)	1.179 (0.288)	0.221 (0.110)	0.522 (0.255)	0.372 (0.101)	0.251 (0.074)	0.062 (0.027)	0.129 (0.037)
Min F_0	56 (43)	60 (18)	245 (83)	45.5 (15)	139 (98)	382 (117)	29 (7)	119 (42)	678 (156)	43 (20)	75 (45)	230 (94)	95 (42)	139 (11)	243 (128)
Max F_0	59 (41.8)	73 (17)	427 (92)	45.5 (15)	346.9 (204)	403 (100)	102 (26)	221 (39)	1436 (270)	45 (23)	120 (61)	2480 (875)	372 (158)	294 (152)	820 (310)
Start F_0	56 (43.9)	62 (18)	261 (72)	45.5 (15)	153.4 (104)	379 (114)	29 (7)	125 (51)	684 (149)	43 (20)	75 (45)	2480 (875)	236 (118)	191 (102)	712 (85)
EndF_0	59 (41.8)	70 (18)	342 (127)	46.3 (20)	350.4 (201)	404 (100)	102 (26)	216.0 (45)	1157 (248)	43 (20)	120 (61)	250 (94)	173 (111)	188 (39)	268 (383)
% Duration	0	14 (36)	34 (36)	0	100	75 (50)	100	57 (15)	38 (18)	0	100	0	100	0	10 (22)
Peak F	206 (62)	73 (36)	360 (42)	191 (57)	225.7 (106)	437 (97)	243 (15)	365 (38)	1177 (166)	74.8 (39)	193 (62)	1325 (611)	170 (46)	171 (63)	488 (330)
Inflects	0	0	0	0	1	0	1	1	0	0	0	1	1	0	0
No./ bout										5	4 or 5 (3 - 8)	6 - 8	3 or 4	7 - 15	6
Pulse rate	21.6 (6.1)	67.3 (3.9)	40.5 (3.4)												
Song unit yr	02	02, 03, 04	02	No	02	No	02	02	No	03, 04	02	02	No	No	No

ers and escorts were usually not known unless they had been heard singing, in which case they were males (Winn *et al.*, 1973; Glockner-Ferrari and Ferrari, 1981).

Initial subjective aural and spectral analysis classified the social vocalizations as 34 separate types. The humpback whale song structure for the three study years was also aurally and visually inspected. Song structure was similar in 2003 and 2004 with the same units being used in the song in both years, but the 2002 song was very different from the 2003/2004 song. A sample of each song unit was extracted from the song recordings and used as a template to subjectively aurally and visually compare each social vocalization type (a match between a song unit and a social vocalization was obvious from very similar aural and spectrogram properties). Of the 34 different social vocalizations, 21 were similar to units of the song current at the time of recording, but used individually or in a small bout (of less than six vocalizations) instead of part of the song sequence (usually lasting more than 5 min). Some of these were observed over the three years of recording while others were evident only for one or two years. The remaining 13 vocalizations were not part of the song and were observed in three years.

A. Initial vocalization classification and individual call description

Most of the calls are harmonic and have been separated arbitrarily into three frequency bands for ease of presentation (Figs. 2–4). Low-frequency harmonic sounds had fundamentals below 100 Hz, mid-frequency sounds 160–550 Hz and high-frequency sounds above 700 Hz. There are also amplitude modulated calls (Fig. 5), calls that have been classified as "noisy and complex," either broadband or harmonic with additional noise like (broadband) characteristics or broadband (Fig. 6), and repetitive calls that always occur as a repetition of individual sounds (Fig. 7).

1. Low-frequency sounds

Vocalizations were aurally and visually categorized into separate call types and measured. The most common vocalizations heard were "wops" and "thwops" (Figs. 2(a) and 2(b)), which were brief harmonic upsweeps at a low frequency (fundamental generally below 60 Hz) compared to other humpback vocalizations. These two vocalizations were similar; the primary difference being that the "thwop" was broken into two parts ("ther-wop") and had two inflection points, whereas the "wop" was one sound and had one inflection point (Table II). "Snorts" (Fig. 2(c)) and 'grumbles' (Fig. 2(d)) were also relatively common sounds with a fundamental frequency generally below 60 Hz (Table II). Where the fundamental frequency could not be distinguished from the background noise, the frequency recorded was that of the lowest detectable frequency above noise limit. These vocalizations were not similar to any units heard in the song for the three study years (Table II). The "sigh" (Fig. 2(e)), though heard only twice in one pod, was part of the 2002 song structure (Table II).

2. Mid-frequency harmonic sounds

"Sirens" (Fig. 3(a)), "short moans" (Fig. 3(b)), "horns" (Fig. 3(c)) and "violins" (Fig. 3(d)) were similar to sounds in

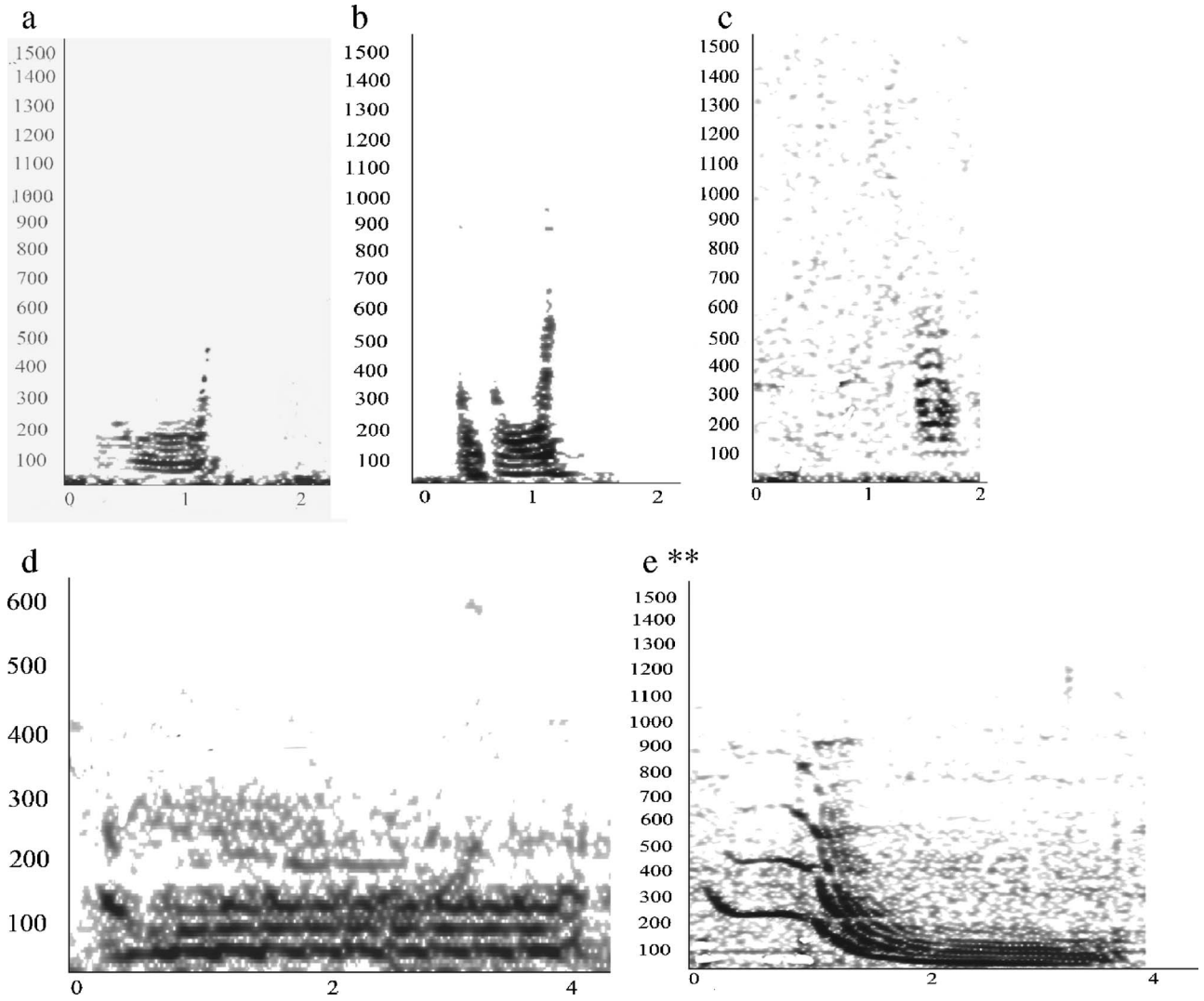


FIG. 2. Spectrogram (x=time (s), y=frequency (Hz)) of sounds in the nonsong, low-frequency, narrowband category; "wop" (a), "thwop" (b), "snort" (c) "grumble" (d) and "sigh" (e). Spectrograms were generated using a FFT of 4096 and frequency resolution of 5.4 Hz. Social vocalizations which were also part of the song structure are identified by**.

the 2002 song, though only heard in later years (2003 and 2004) as part of the social vocalization repertoire (Table II). Longer harmonic sounds such as "groans" (Fig. 3(e)), "ascending moans" (Fig. 3(f)) and "modulated moans" (Fig. 3(g)) were a relatively common part of the humpback song and social vocalization repertoire during all three study years (Table II). "Modulated moans" were particularly common in the 2003/2004 song and were one of the more common song-unit social vocalizations. Higher frequency harmonic sounds such as "cries" (Fig. 3(h)), "modulated cries" (Fig. 3(i)) and "ascending cries" (Fig. 3(j)) were not similar to any units in the song structure for the three study years (Table II).

The most common sound in this group was the relatively loud harmonic upsweep "trumpet," similar to a major unit found in the 2002 humpback song structure (Fig. 3(k)) (Table II), though heard in all three years as part of social vocalization repertoire.

3. High-frequency harmonic sounds

"Shrieks," the highest frequency long, harmonic sound, were either ascending (Fig. 4(a)) (similar to units used in the song structure in all three study years) or descending (Fig. 4(b)) (similar to units in the 2002 song) (Table II). "Squeaks" (Fig. 4(c)) were the shortest and highest frequency sounds heard in the social vocalization repertoire and were part of the song structure in all three years. All of these social vocalizations were heard in all three study years (Table II).

4. Amplitude-modulated sounds

"Growls" (Fig. 5(a)) were heard in all three analyzed years, in both the song and as part of the social vocalization repertoire. These vocalizations were a combination of harmonic and amplitude-modulated components, the modulation being close to the fundamental frequency (Table III). "Purrs" (Fig. 5(b)) were low-frequency amplitude-modulated sounds in which the sound energy was distributed over a broad spectrum. These vocalizations were heard in all three years as part of the social vocalization repertoire, but were heard only in the 2002 song (Table III). "Trills" (Fig. 5(c)) were long, highly amplitude-modulated sounds also similar to units heard in the 2002 song, though were heard in other years as part of the social vocalization repertoire (Table III).

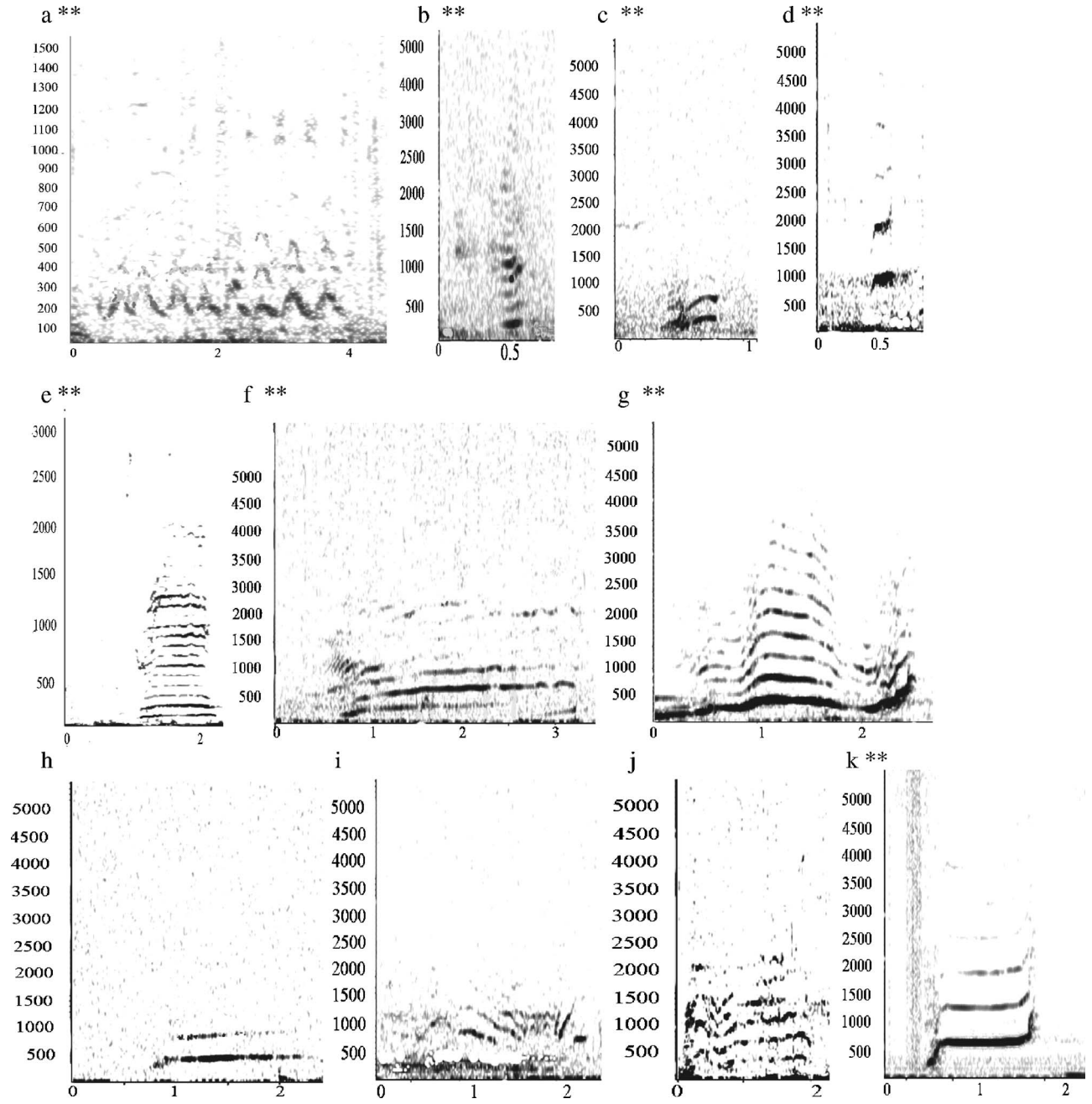


FIG. 3. Spectrogram (x=time (s), y=frequency (KHz)) of mid-frequency harmonic sounds; "groan" (a), "siren" (b), "short moan" (c), "ascending moan" (d), "modulated moan" (e), "horn" (f), "cry" (g) "modulated cry" (h), "ascending cry" (i), "violin" (j) and "trumpet" (k). Spectrograms were generated using a FFT of 1024 samples and frequency resolution of 21.5 Hz. Social vocalizations which were also part of the song are identified by **.

5. Broadband, "noisy" and complex sounds

Sounds which were presumed to be underwater blows were brief, high level, broadband sounds with energy spread over a wide frequency range (45–>10,000 Hz) (Fig. 6(a), Table III). Other sounds in which energy was distributed over a broad spectrum, though with a harmonic component, (complex sounds) were "barks," "bellows," "creaks" and "screeches" (Figs. 6(b)–6(e)). "Bellows," "creaks" and "screeches" had a harmonic structure but with considerable bandwidth so energy distributed over a wide frequency range creating "smearing" around the frequency bands. All of these sounds were similar to units heard in the 2002 song (Table III), but were heard in the social vocalization repertoire throughout the three analyzed years; "barks" being particularly common. The "scream" (Fig. 6(f)), heard in multiple pods, was a unique sound consisting of a high-frequency carrier which shifted upward in frequency over the duration (from about 800 to 2500 Hz) with some overlying a broadband component (Table III). It was not heard in the song over the three years.

6. Repetitive sounds

Repetitive sounds were short, low-frequency, discrete sounds that occurred in groups or bouts. "Grunts" (Fig. 7(a), Table III) were one of the most commonly heard vocalizations in the repertoire throughout the three years and oc-

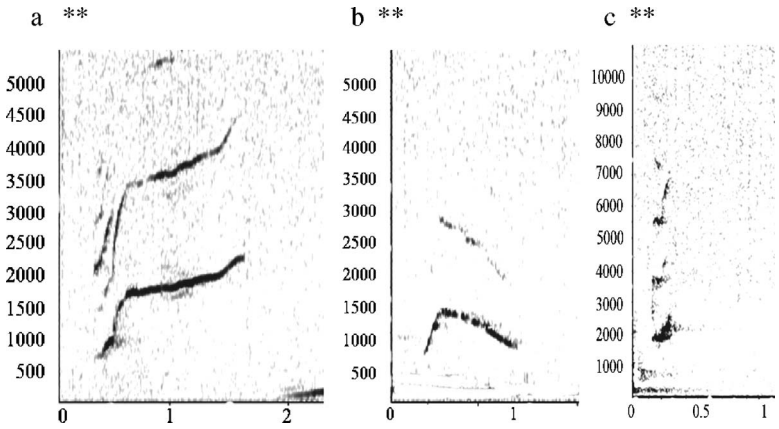


FIG. 4. Spectrogram (x=time (s) v y=frequency (KHz) of high-frequency harmonic sounds; "ascending shriek" (a), "descending shriek" (b) and "squeak." Spectrograms were generated using a fast Fourier transformation size of 1024 samples and a frequency resolution of 21.5 Hz. Social vocalizations which were also part of the song are identified by **.

curred in the 2003 and 2004 song. "Croaks" (Fig. 7(b)) and "yaps" (Fig. 7(c)) were part of the 2002 song (Table III) though were only heard in 2003 and 2004 as part of the social vocalization repertoire. "Yelps" (Fig. 7(d)), "pulses" (Fig. 7(e)) and "low yaps" (Fig. 7(f)) were not heard in the song during the three years (Table III) and were comparatively uncommon vocalizations (Table III).

B. Statistical analysis of vocalization type

1. Comparison of signal types

A one-factor analysis of variance (ANOVA) was conducted on all of the measured parameters to compare between subjectively categorized signals. The means of all measured parameters were significantly different between the 34 proposed signal types ($P<0.001$). Of the measured frequency parameters (start and end frequencies, minimum and maximum frequencies, frequency trend ratio, frequency range ratio, frequency of the spectral peak) the highest F value was for the maximum frequency of the lowest frequency component of the sound, suggesting this frequency parameter varied most between signal types (Fig. 8). The maximum frequency of the lowest frequency component of the sounds ranged from 40 to 2500 Hz. High F values (>100) were also obtained for the end frequency of the lowest frequency component (range of 40–2500 Hz), which also suggests a high degree of variability in this parameter among signal types. Of all the sounds parameters, F values for the number of inflections was highest ($F=1126$) reflecting the large range of number of inflection points between different vocalization types (range 0–6).

2. Principal components analysis

Principal component analysis using the measured parameters in Table I generated four factors accounting for 80.3% of the variation, eigenvalues for the first four factors being greater than 1. Therefore, using these four factors resulted in 19.7% of the information being lost. Factor loading scores showed that factor 1 was most closely correlated with minimum, maximum, start and end fundamental frequency and peak frequency, reflecting the frequency characteristics of the signal. Factor 2 was most closely correlated with percentage of duration to the maximum frequency, the number of inflections and frequency trend ratio, reflecting the frequency modulation characteristics. The number of sounds per bout was also included in this factor. Factor 3 correlated with the duration of the sound and pulse rate of amplitude-modulated sounds. Factor 4 contained the frequency range ratio only. The mean values for the mean components differed significantly between signal types (ANOVA, $F>155$, $p<0.001$).

3. Discriminant function analysis

The DFA process correctly classified 78.6% of calls ($n=660$) correctly when using the factor scores generated by

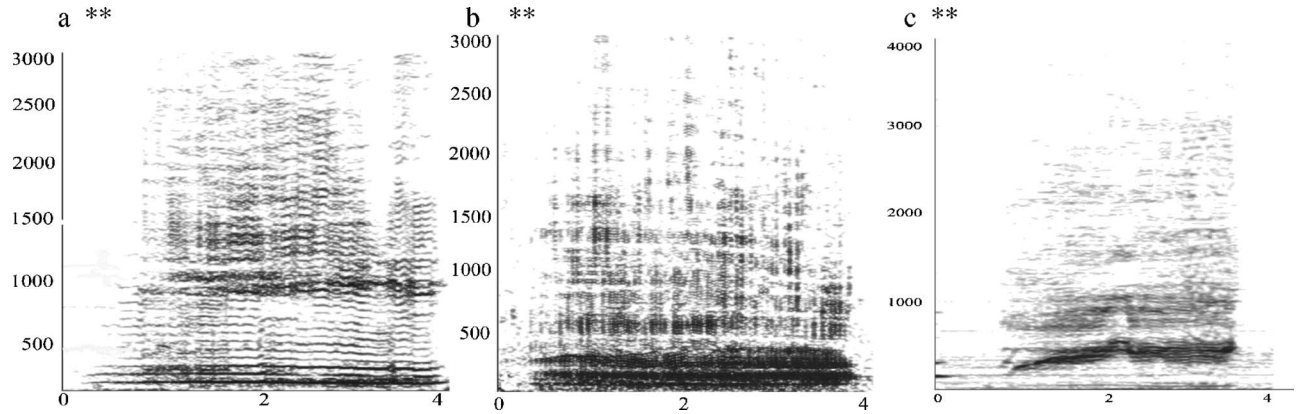


FIG. 5. Spectrogram (x=time (s), y=frequency (Hz)) of amplitude-modulated sounds; "growl" (a), "purr" (b) and "trill" (c). Spectrograms were generated using a FFT of 4096 and frequency resolution of 5.4 Hz (Figs. 4(a)–4(c)) and FFT of 1096 and frequency resolution of 21.5 Hz (Fig. 4(d)). Social vocalizations which were also part of the song are identified by **.

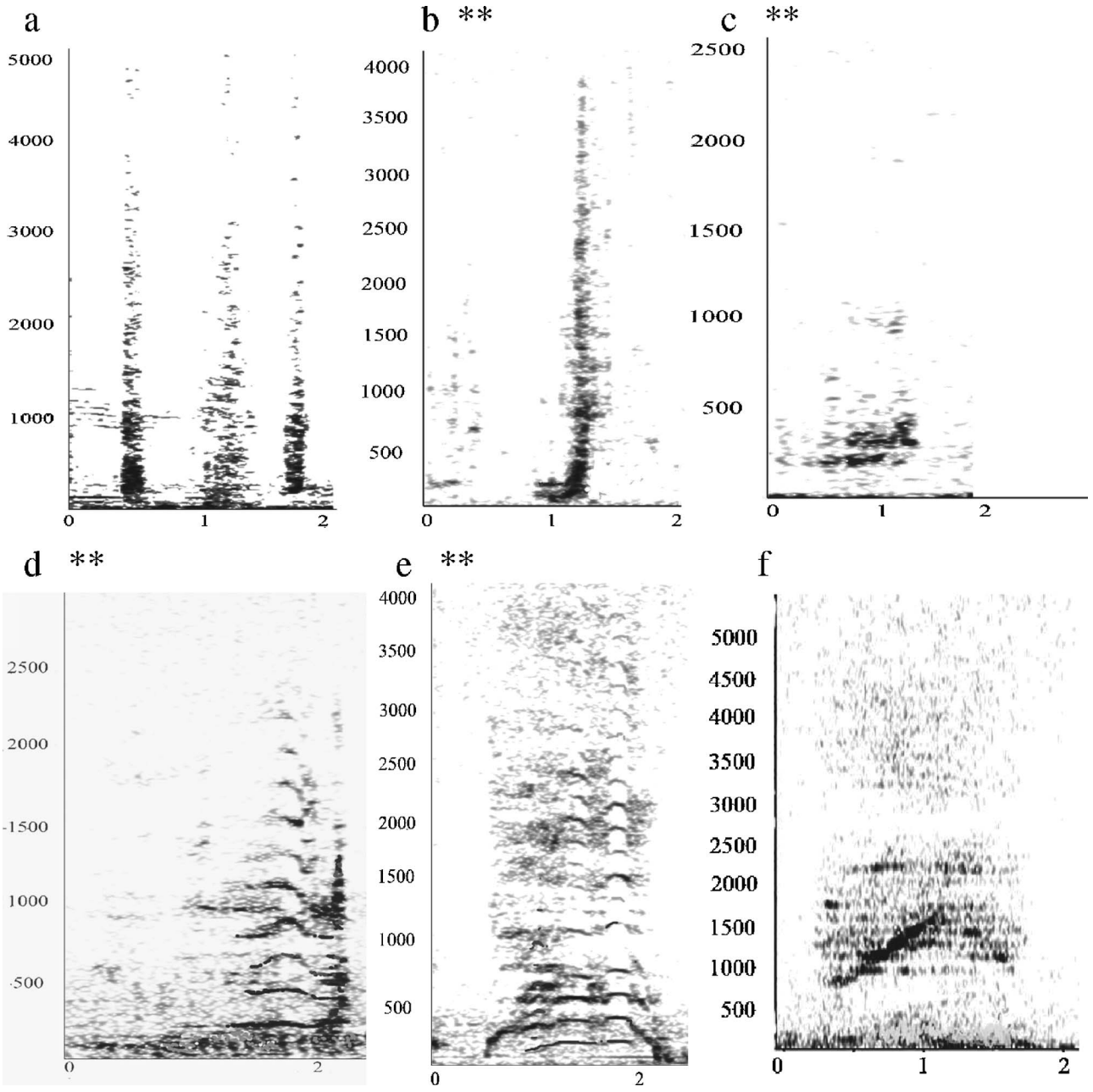


FIG. 6. Spectrograms (x=time (s), y=frequency (Hz)) of "broadband," "noisy" and complex sounds; "presumed underwater blow" (a), "bark" (b), "bellow" (c), "creek" (d) and "screech" (e) and "scream" (f). Spectrograms were generated using a FFT of 4096 and frequency resolution of 5.4 Hz. Social vocalizations which were also part of the song are identified by **.

the PCA. Predicted classifications were over 67% correct in all but five of the sound groups (presumed underwater blows were classified as "snorts," "creeks" were classified as "thwops," "horns" were classified as "barks," "screeches" were classified as "barks" or "groans" many of the "short moans" were classified as "barks"). However, generating factor scores from sound measurements may lose important information about that sound. When using the sound measurements as independent variables, the DFA process correctly classified 89.4% of the sounds and predicted classifications were over 60% correct for all groups. The DFA output showed that all measurements were significantly different between sound type ($P<0.0001$). In this analysis, the number of inflections ($F=720$) and the pulse rate ($F=1484$) discriminated most between vocalizations (according to the F values), followed by the number in the bout ($F=76$). Of the fundamental frequency parameters, the end frequency had a slightly higher F value than the rest of the measurements ($F=36$).

IV. DISCUSSION

Humpback whales are infamous for their varied and diverse song repertoire. This study shows that humpback social vocalizations are also extremely diverse in structure. Frequencies ranged from less than 30 Hz to 2.5 kHz and initial subjective qualitative visual and acoustic assessment of the number of different vocalizations produced 34 different vo-

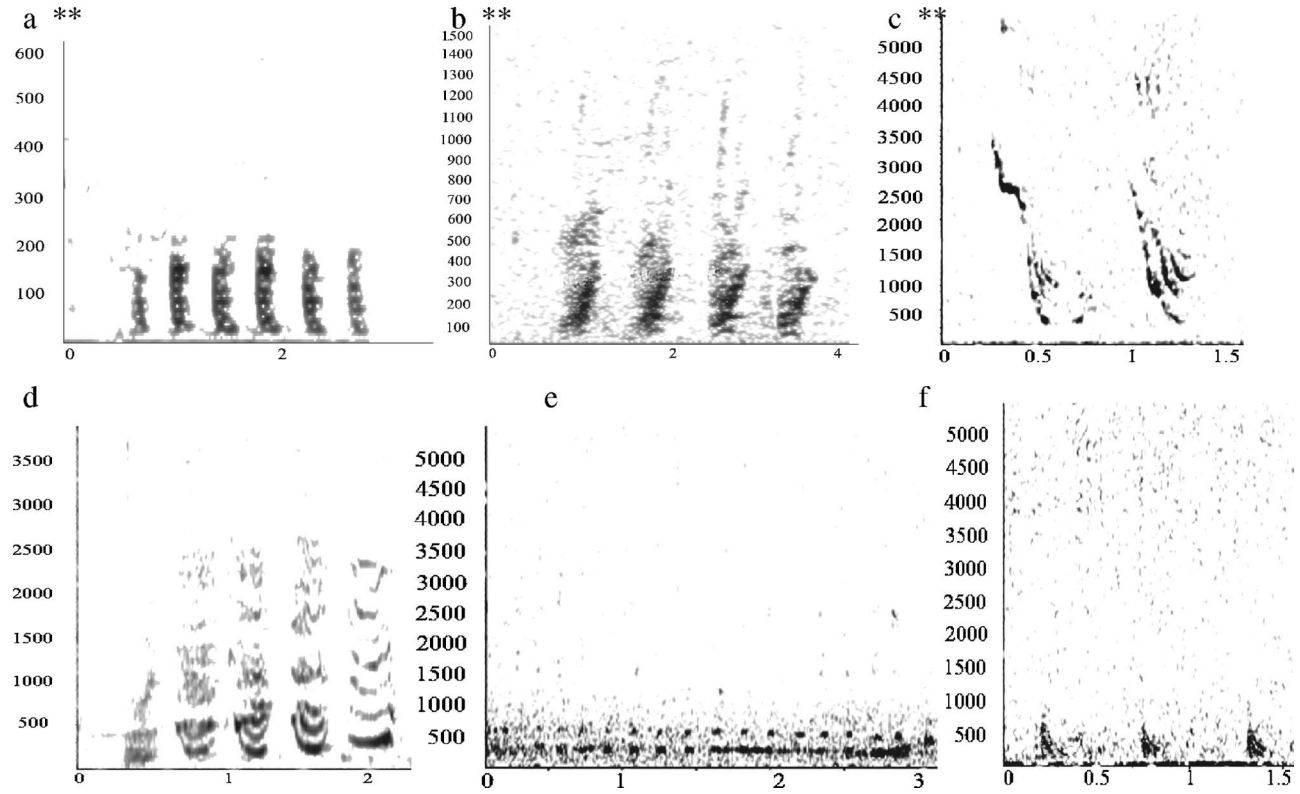


FIG. 7. Spectrograms (x=time (s), y=frequency (Hz)) of repeated sounds; "grunts" (a), "croaks" (b) and "yelps" (c). Spectrograms were generated using a FFT of 4096 and frequency resolution of 5.4 Hz. "Pulses" (d), "low yaps" (e) and "yaps" (f) spectrograms were generated using a FFT of 1024 samples and frequency resolution of 21.5 Hz. Social vocalizations which were also part of the song are identified by **.

calization types, of which 27 were heard from more than one group. This catalogue is much larger than previously described social vocalization repertoires (e.g. Thompson *et al.*, 1986). Contrary to previous suggestions (e.g. Tyack 1983; Silber 1986), they were commonly heard in mother and calf pairs and single animals as well as in other social groups such as mother-calf and escorts, pairs of adults, groups of three and groups of four adults. The use of social vocalizations within these different group compositions will be the subject of a further study. However, there are problems with subjectively analyzing sounds. There is a degree of both individual and between-group variation in similar vocalizations and it may be difficult to decide if similar calls are classified as one vocalization, or two separate vocalization types. Measuring sound parameters and statistically analyzing them helps in making objective decisions.

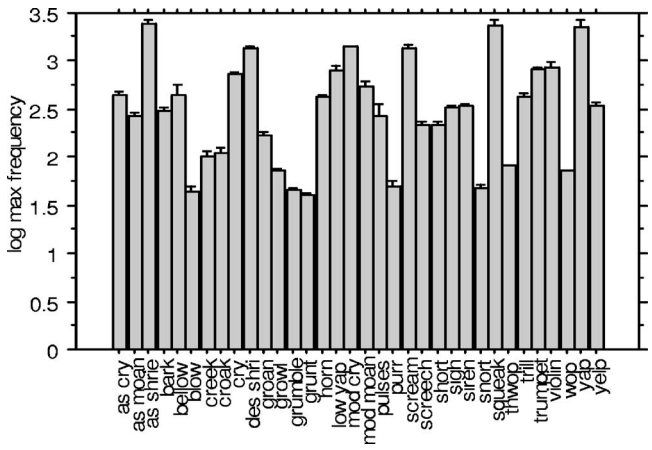


FIG. 8. Mean (± scanning electron microscropy) plot of log maximum frequency of the 34 proposed different vocalization types. The maximum frequency was significantly different between vocalization type (ANOVA: $P < 0.0001$).

The PCA arranged vocalizations primarily by the frequency characteristics and secondly by the frequency modulation characteristics and the number of vocalizations within a bout. A third factor contained the duration of the vocalization and pulse rate (where applicable in amplitude-modulated vocalizations) and the fourth factor contained only the frequency range ratio. The highest source of variation between different vocalization groups came from the pulse rate, which separated amplitude-modulated vocalizations (e.g., "purrs," "growls" and "trills") from harmonic, complex and "noisy" vocalizations. Also, the number of inflection points in the sound was a source of high variation, separating highly modulated vocalizations (e.g., the "siren," "modulated moan," "ascending cry" and "modulated cry") from unmodulated, upsweep or downsweep vocalizations. Another source of variation was the number of vocalizations within a bout, which separated vocalizations that were heard as single sounds from those that were heard in bouts (i.e., repetitive sounds, e.g., "grunts," "croaks" and "yelps"). The most variable of the fundamental frequency characteristics was the maximum frequency (highlighted in the ANOVA) and end frequency of the sound (highlighted in the DFA). These features are obvious and easily recognizable and it is likely they are relevant in the contextual use of different acoustic signals. Fundamental frequency characteristics and frequency

TABLE IV. Synopsis of humpback whale nonsong vocalizations recorded from previous studies (including the fundamental frequency and duration measurments if stated). Visual comparisons of spectrograms for vocalizations recorded in the following studies were made to spectrograms generated for vocalizations recorded in this study.

	Signal type	Fund freq (Hz)	Duration (s)	Spectrogram similar to
Watkins 1967	Wheezing blow	Complex	1.5–2.0	Scream
			0.72–3.48	One similar to a wop
Winn *et al.*, 1979	Pulsed moans		(variable)	
			2.1	Short moan
	Moans		(variable)	As moan
	Yups		0.14–0.2	
	Short cry	1625–1775	0.23	Squeak
	Chirps	1923–5520	0.05–0.08	
	Clicks		0.006	
D'Vincent *et al.*, 1985	Feeding cry	400–800	3	
Thompson *et al.*, 1986	Simple moans	25–30	0.8	Wop
	Complex moans	35–360	0.8	
	Grunt trains	25–80	0.3–0.4	Grunts
	Broadband pulse trains	400–5000	0.001	
	Shrieks	555–2000	0.42	Scream
	Trumpets	410–420	0.55	
Cherchio and Dahlheim 2001	Feeding cry	360–988	>0.5	As cry

modulation characteristics probably play important roles in humpback signal perception, though to test this, further behavioral studies should be carried out.

Although most of the sounds were separated into discrete groups with a DFA, using PCA generated factor scores does not accurately capture the entire acoustic repertoire. Also, some of the vocal repertoire formed a continuum rather than falling into discrete clusters of sounds, as was found previously in the southern right whale calls (Clark, 1982). For example, long, amplitude-modulated "growls" sometimes merged into amplitude-modulated "purrs." "Growls" were sometimes very similar to "grumbles" and "snorts" were arbitrarily divided from "grumbles" based on the duration of the vocalization. What also must be remembered from this kind of statistical analysis is that measured sound parameters are selected subjectively. These parameters, believed to be important to human observers, may not necessarily be of importance to the whales.

Measured variables in a DFA are "independents," that is, are a set of uncorrelated variables and PCA scores are usually used in lieu of raw data in a DFA (e.g., Boisseau 2005). However, the condensation of all sound measurements into four factors may not be appropriate given that the forced correlations in PCA may lose important acoustic aspects of the sound. The factor scores assume that sound components are correlated, for example, factor 1 assumes the fundamental frequency components (start, end, maximum, minimum frequency) are all highly correlated. Though this may be true in many sounds (the start and end frequencies must lie between the maximum and minimum frequencies), one frequency measurement is not necessarily the predictor of another frequency measurement in all sounds. A majority of the sounds were upsweeps or unmodulated, accounting for the high correlation between the start and minimum frequency (or the end and maximum frequency). However, this correlation is not true in downsweep sounds, therefore an important acoustic property of these sounds was eliminated when using factor scores in lieu of raw data. Using all sound measurements as a set of uncorrelated independent variables in a DFA produced a set of 34 discrete calls and seemed to accurately support the aural and spectrographic categorization of sounds. Clark (1983) tested southern right whale sound categories to determine their biological significance. The same testing should be carried out on both the proposed humpback sound categories and separate calls to validate this proposed acoustic repertoire.

Vocalizations were either heard multiple times in multiple groups (the most common being the "thwop" and "wop"), or heard in only one group (though always multiple times within that group). Single males joining other groups were found to switch from low-frequency sounds to higher frequency song-like harmonic sounds, often in short, repeated sequences with similarity both in structure and sound type to song phrases, but lacking the continuity and highly structured themes of true song. Winn *et al.* (1979) referred to sounds which were similar to song units recorded from two entrapped humpback whales, a male and female. In the three study years reported here, 21 of the 34 vocalization types were part of the song repertoire at some time during the three study years; 14 of these 21 vocalization types were part of the 2002 song and only one vocalization was part of the 2003/04 song (the song changed little between 2003 and 2004) with six vocalization types heard in the song during all three study years. However, the social vocalization song units were not only part of the social vocalization repertoire in the year they were heard in the song. For example, units that were only part of the 2002 song were heard in the social vocalization repertoire during 2002, 2003 and 2004. Therefore, song units and social sounds are not discrete categories but seem to be interchangeable, and it is possible that all social vocalization units were present in a song at some

stage, or *vice versa*. This will be the subject of a further and more detailed investigation.

Some of the vocalizations recorded in this population of humpbacks also seemed to be similar to vocalizations recorded from other humpback populations. For example, the "wop" has similar frequency characteristics and spectrogram characteristics to the "simple moans" recorded from humpbacks in southeast Alaska (Thompson *et al.*, 1986) (Table IV). "Grunt trains" were also present in these recordings and "blowhole-associated shrieks" seem to be structurally similar (based on similarities in the spectrogram) to the "screams" recorded from the east Australian humpbacks (Watkins, 1967) (Table IV). A particular rhythmic series of eight repeated cries (named "ascending cries") recorded in one particularly vocal mother-calf and escort group, were of a similar structure (though at a lower frequency) to a series of cries described in feeding humpbacks in Alaska (Cerchio and Dahlheim, 2001) (Table IV). These are initial and tentative comparisons and further comparative work should be carried out to investigate the similarity of these calls between populations.

This study provides the first comprehensive catalogue of social vocalization calls in the migrating humpback whales of eastern Australia. The catalogue includes 34 different vocalization types which are extremely varied in terms of structure. Most of the vocalization types were repeatable (i.e. heard in multiple groups), stereotyped among groups and were commonly heard in all social groups. This catalogue provides a necessary basis for subsequent analysis, for example, the changes of the social vocalization repertoire over time, and how frequently, if at all, new sound units are incorporated into and eliminated from the repertoire. Comparisons can also be made with the repertoire of other humpback populations, especially those on the feeding or breeding grounds. The significance of song units being used as social vocalizations is not known and should also be the subject of future studies.

ACKNOWLEDGMENTS

The authors thank everyone involved in the Humpback Acoustic Research Collaboration (HARC), in particular Patrick Miller, Nicoletta Biassoni, John Noad, Josh Smith and the numerous volunteers who donated their time and energy to this project, especially Ceri Morris, Ben Chicoski and Alice Pope. The U.S. Office of Naval Research and the Australian Defense Science and Technology Organization provided funding. The authors also thank David Paton for his invaluable field expertise and Eric Kniest for his continued support in the development of *Cyclopes*. Finally the authors would like to thank Emeritus Professor Michael Bryden for his valued comments.

Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M. H., and Andrews, K. (**2006**). "Acoustic properties of humpback whale songs," J. Acoust. Soc. Am. **120**(2), 1103–1110.

Boisseau, O. (**2005**). "Quantifying the acoustic repertoire of a population: The vocalizations of free-ranging bottlenose dolphins in Fiordland, New Zealand," J. Acoust. Soc. Am. **117**(4 Pt.1), 2318–2329.

Cerchio, S., and Dahlheim, M. (**2001**). "Variations in feeding vocalizations of humpback whales *Megaptera novaeangliae* from southeast Alaska," Bioacoustics **11**, 277–295.

Chabot, D. (**1988**). "A Quantitative technique to compare and classify humpback whale (*Megaptera novaeangliae*) sounds," Ethology **77**(2), 89–102.

Chittleborough, R. G. (**1965**). "Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski)," N.Z.J. Mar. Freshwater Res. **16**, 33–128.

Clapham, P. J., Palsbøll, P. J., Mattila, D. K., and Vasquez, O. (**1992**). "Composition and dynamics of humpback whale competitive groups in the West Indies," Behaviour **122**, 182–194.

Clark, C. W. (**1982**). "The acoustic repertoire of the southern right whale: A quantitative analysis," Anim. Behav. **30**, 1060–1071.

Clark, C. W. (**1983**). "Acoustic communication and behaviour of the southern right whales, *Eubalaena australis*," in *Communication and Behaviour of Whales*, edited by R. Payne (Westview, Boulder, CO) pp. 163–198.

Dawbin, W. H. (**1966**). "The seasonal migratory cycle of Humpback whales," in *Whales, Dolphins and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley) pp. 145–170.

D'Vincent, C. G., Nilson, R. M., and Hanna, R. E. (**1985**). "Vocalization and coordinated feeding behaviour of the humpback whale in southeastern Alaska," Sci. Rep. Whales Res. Inst. **36**, 41–47.

Glockner, D. A., and Venus, S. C. (**1983**). "Identification, growth rate and behaviour of humpback whale (*Megaptera novaeangliae*) cows and calves in the waters off Maui, Hawaii, 1977–79," in *Communication and Behavior of Whales*, edited by R. Payne AAAS Selected Symposia Series (Westview, Boulder, CO), pp. 223–258.

Glockner-Ferrari, D., and Ferrari, M. (**1981**). "Correlation of the sex and behavior of individual humpback whales, *Megaptera novaeangliae*, to their role in the breeding population," paper presented at the *Fourth Biennial Conference on the Biology of Marine Mammals*, Abstract.

I.W.C. (2005). "Intersessional working group on southern hemisphere humpback whales: Revised table by breeding stock (as at of 1 May 2005." Document SC/57/SH11 presented to the *Scientific Committee of the International Whaling Commission*, Ulsan, South Korea, May 2005.

Jurasz, C. M., and Jurasz, C. M. (**1979**). "Feeding modes of the humpback whales, *Megaptera novaeangliae*, in southeast Alaska," Sci. Rep. Whales Res. Inst. **31**, 69–83.

Mellinger, D. K. (**2001**). "Ishmael 1.0 User's Guide." NOAA Technical Memorandum OAR PMEL-120, 2001.

Mercado III, E., Herman, L. M., and Pack, A. A. (**2005**). "Song copying by humpback whales: Themes and variations," Anim. Cogn. **8**, 93–102.

Mobley, J. R., Herman, L. M., and Frankel, A. S. (**1988**). "Responses of wintering humpback whales (*Megaptera novaeangliae*) to playback recordings of winter and summer vocalizations and of synthetic sounds," Behav. Ecol. Sociobiol. **23**, 211–223.

Noad, M. J., Cato, D. H., and Stokes, M. D. (**2004**). "Acoustic tracking of humpback whales: Measuring interactions with the acoustic environment," *Proc. Acoustics 2004, Annual Conference of the Australian Acoustical Society, Gold Coast*, 3–5 November 2004, pp. 353–358.

Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 585–597.

Payne, K., and Payne, R. (**1985**). "Large scale changes over 19 years in songs of humpback whales in Bermuda," Z. Tierpsychol **6**, 89–114.

Pearson, K. (**1901**). "On lines and planes of closest fit to systems of points in space," Philos. Mag. **2**, 559–572.

Richardson, W., Greene Jr., C. R., Malme, C. I., and Thomso, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Sharpe, F. A., Dill, L. M., Beaver, V., and Spellman, B. (**1998**). "Killing me softly: Feeding calls of the Alaskan humpback whale," *Abstracts of the World Marine Mammal Science Conference*, Monaco, January, pp. 20–24.

Silber, G. K. (**1986**). "The relationship of social vocalizations to surface behaviour and aggression in the Hawaiian humpback whale (*Megaptera novaeangliae*)," Can. J. Zool. **64**, 2075–2080.

Thompson, P. O. W., Cummings, W. C., and Ha, S. J. (**1986**). "Sounds, source levels and associated behavior of humpback whales, Southeast Alaska," J. Acoust. Soc. Am. **80**(3), 735–740.

Thompson, P. O. W., Cummings, W. C., and Kennison, S. J. (**1977**). "Sound production of humpback whales," J. Acoust. Soc. Am. **62**, S89.

Tyack, P. L. (**1997**). "Acoustic communication under the sea," in *Animal Acoustic Communication: Sound Analysis and Research Methods*, edited

by S. C. Hopp, M. J. Owren and C. S. Evans (Springer-Verlag, Berlin).
Tyack, P. L. (**1983**). "Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds," Behav. Ecol. Sociobiol. **13**, 49–55.
Tyack, P. L., and Clark, C. W. (**2000**). *Communication and Acoustic Behaviour of Dolphins and Whales*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 156–224.
Watkins, W. A. (**1967**). "Air-borne sounds of the humpback whale (*Megaptera novaeanglia*)," J. Mammal. **48**(4), 573–578.
Winn, H. E., Beamish, P., and Perkins, P. J. (**1979**). "Sounds of two entrapped humpback whales (*Megaptera novaeangliae*) in Newfoundland," Mar. Biol. (Berlin) **55**, 151–155.
Winn, H. E., Bischoff, W. L., and Taruski, A. G. (**1973**). "Cytological sexing of cetaceans," Mar. Biol. (Berlin) **23**, 343–346.

Estimated source intensity and active space of the American alligator (*Alligator Mississippiensis*) vocal display

Neil P. McAngus Todd[a)]
Faculty of Life Science, Jackson's Mill, University of Manchester, Manchester M60 1QD, United Kingdom

(Received 3 August 2006; revised 21 August 2007; accepted 27 August 2007)

In this article the results are reported of a study to measure the intensity of the vocal displays of a population of American alligators (*Alligator mississippiensis*). It was found that the dominant frequencies in air range between 20 and 250 Hz with a source sound pressure level (SPL) of 91–94 dB at 1 m. The active space for the air-borne component is defined by the background and was estimated to be in a range up to 159 m in the 125–200 Hz band. For the water-borne component the dominant frequency range was 20–100 Hz with a source SPL of 121–125 dB at 1 m. The active space in water is defined by hearing thresholds and was estimated to range up to 1.5 km in the 63–100 Hz band. In the lowest frequency bands, i.e., 16–50 Hz, the estimated active space for otolith detection of near-field particle motion in water ranged to 80 m, which compared significantly with far-field detection for these frequencies. It is suggested that alligator vocal communication may involve two distinct sensory mechanisms which may subserve the functions of scene analysis and reproduction, respectively. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2785811]

PACS number(s): 43.80.Lb, 43.80.Ka [JAS] Pages: 2906–2915

I. INTRODUCTION

In animal acoustic communication the active space is the radial distance from source over which a signal remains above detection threshold by a receiver and is dependent on four factors: the source intensity, the transmission characteristics of the medium, the background noise, and the receiver sensitivity (Brenowitz, 1982). The several active space studies that have been carried out have shown a very considerable range of estimates from less than 100 m in air for the elephant seal attraction call (Southall *et al.*, 2003) up to tens of kilometers for high-frequency water-borne cetacean calls (Janik, 2000; Miller, 2006). Primate loud calls are known to be of high intensity (Todd and Merker, 2004) and may have an active space in air of a few kilometers (Brown, 1989). This extensive range reflects the variable function, source, and transmission characteristics as well as the environment in which the vocalizations are made. For vertebrate acoustic communication active space considerations are further complicated by the change in the nature of medium (water to air), mode of transmission (particle motion versus acoustic pressure), and mechanism for reception, perilymphatic, or otolithic (Fritzsch, 1999).

In anamniotes the otolith organs, the sacculus, utriculus, and lagena, are important auditory structures for near-field (acoustic motion) sensing (Popper and Fay, 1999; Lewis and Narins, 1999). During the course of evolution the amniote ear developed new perilymphatic structures for far-field (sound pressure) hearing in air, e.g., the mammalian cochlea (Lewis *et al.*, 1985). It has been established, however, that the otolith organs in species from all vertebrate classes, including humans, have conserved a particular sensitivity to substrate or bone conducted sound (Lewis and Narins, 1985; Cuthoys *et al.*, 2006; Welgampola *et al.*, 2003; Todd *et al.*, 2003), consistent with their function as near-field sound sensors in anamniotes (Fritzsch, 1999). A question remains however as to what function, if any, might be served by a near-field otolithic sound reception in amniotes (Todd and Merker, 2004). In order to go some way towards answering this question it is necessary to estimate the relative size of active space for far-field (cochlear) versus near-field (otolith) reception.

Crocodilians are interesting animals to study with regard to the above question as they are aquatic, highly vocal, and possess an anatomically large sacculus (Lewis *et al.*, 1985). Alligators in particular are one of the most vocal species. They produce a loud roaring call referred to as a bellow (Vliet 1989) which is performed from a "head oblique tail arched" (HOTA) posture and produced within the body, analogous to fish vocal production. Alligators spontaneously bellow after dawn throughout the year but are most active during the spring courtship season. Both males and females bellow and often do so together in choruses. Male vocalizations though tend to include an infrasound component, referred to as subaudible vibration (SAV), just prior to the audible bellow (see Fig. 2). The acoustic near field, characterized by incompressive bulk flow, for such sounds is extensive in water, as the near field dominates within about 1/3 a wavelength for a monopole source and the wavelength in water is five times that of air. Thus for a 20 Hz sound in water the near-field dominates within about 25 m under ideal conditions. This means that a fish-like otolithic acoustic sensor, specialized for the detection of particle motion, may be more optimally placed to mediate the water-borne component of alligator vocal displays at short distances. In order to estimate the active space for near-versus far-field reception,

[a)]Electronic mail: neil.todd@manchester.ac.uk

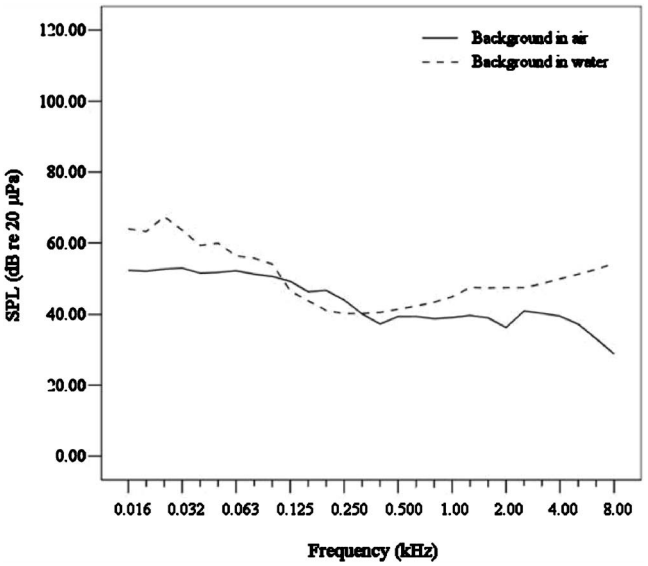


FIG. 1. A 1/3 octave spectral analysis of the background noise at the Alligator pond in air (continuous line) and water (broken line). All measurements are maximum values of SPL with fast linear weighting (L_{LF}) and ~20 μPa.

however, it is necessary to measure the source intensity and transmission characteristics of alligator vocal displays, as well as the background noise, in air and water, and this was the goal of the study reported here.

II. METHODS

A. Subjects

The subjects for the study were a population of American alligators at the Australian Reptile Park, NSW, Australia. The population consisted of about 30 adults including four large males. The alligators were maintained in an enclosed pond of about 50 by 120 m, with maximum depth of about 2 m.

B. Apparatus

The vocal displays were recorded simultaneously in air and water on a Marrantz PMD670 digital sound recorder using a B&K free-field microphone (sensitivity −26.4 dB ~1 V/Pa) amplified by a B&K 2260 and a HIT-96-MIN hydrophone with built in preamplifier (sensitivity −43.6 dB ~1 V/Pa). Playback was carried out with an adapted car stereo system and a subwoofer speaker.

C. Procedures and measurements

The study was conducted in 12 sessions from March to May. While bellowing is a spontaneous activity, study of the animals out of season is facilitated by the fact that they can be easily induced to vocalize by use of playback of recorded alligator calls (Beach, 1944). The playback speaker was placed close to the edge of the pond. The microphone was suspended on a boom over the pond and the hydrophone was thrown into the pond close to the microphone. During a recording session notes were taken on the sex and approximate distance of the animals. From the 12 sessions about 100 edited extracts were obtained and analyzed off-line using the B&K 2260 analyzer. For each extract, measurements were made of peak SPL (L_{Lpk}), rms SPL (L_{Ll}), and the 1/3 octave frequency and pressure of the dominant mode for air and water. For each of the sessions recordings were also made of both broadband and 1/3 octave measures of the background noise. Although it is normal to use a pressure reference of 1 μPa for measurements in water, all pressure measures were made ~20 μPa for convenience of making direct comparisons of the same source.

D. Source and transmission estimation

In order to estimate source intensity and transmission characteristics linear and quadratic regression analyses were performed on each of the pressure measures in air and in water as a function of estimated distance. The source intensity at 1 m was then taken as the intercept on the log-log plot and the transmission characteristic defined by the slope. In the ideal cases slopes of −0.5 and −1 in the log-log plot correspond, respectively, to cylindrical and spherical spreading. A nonlinear slope would indicate an additional process coming into play with increasing distance. In those instances where there was a significant nonlinearity in slope the quadratic regression was used to estimate the source intensity. An estimate range was obtained from the standard error of the regression parameters.

E. Model for near-field transmission in water

In order to estimate the effective range for an otolithic ear we assume that the source is a pulsating sphere, i.e., monopole. This assumption is reasonable as the radius of the source is small compared with the wavelength for low frequencies. For example, with a radius $a=0.25$ m (approximate alligator girth about 0.5 m) then $ka<1$, where k is the wave number. The source in this case is located in shallow water with a muddy bottom and pressure release at the surface. A theoretical modeling of this situation is not a trivial matter and it was decided therefore to allow the empirically determined transmission characteristics to guide the estimates. Given the concentration in the low-frequency region an analysis was carried out independently for each 1/3 octave band between 20 and 40 Hz for which there was sufficient data and for each of the bands the source sound pressure at 1 m and transmission slope were estimated by regression. For a monopole source in the free field the corresponding peak particle velocity at radial distance r is given by

$$v(r) = v_{far}(r) + v_{near}(r) = \frac{p(r)}{\rho c}\left(1 + \frac{1}{jkr}\right)$$

(Rossing and Fletcher, 1995), where $p(r)$ is the acoustic pressure and ρc is the acoustic impedence (for water 1.48×10^6), which gives the near-field velocity at 1 m as $v_{near}(1) = p(1)/jk\rho c$. In logarithmic terms the calculation for the near-field particle velocity level (PVL) at 1 m reduces to

$$\mathrm{PVL}(1) = \mathrm{SPL}(1) - 34 - 20\log_{10}(\omega),$$

where ω is angular frequency and where the reference velocity and pressure are 1 μm s^{-1} and 20 μPa, respectively. The

particle acceleration is related to velocity by $a=\omega v$ so that the calculation for particle acceleration level (PAL) is simply given by SPL(1)−34, where the reference acceleration is 1 μm s^{-2}. As the near-field velocity is related to the far-field by $v_{near}=v_{far}/jkr$, then we can estimate the decay slope of the near field in the log-log plot as that of the far field minus 1. Thus for ideal spherical spreading from a monopole source the near-field slope is −2.

F. Threshold values

For the final estimate of active space data on detection thresholds are required. Several studies have measured crocodilian hearing sensitivity in air using a variety of methods, including behavioral responses (Beach, 1944), cochlear potentials (Wever, 1971), single unit recordings from primary afferents (Kinke and Pause, 1980), and the auditory brain stem response (ABR) (Strain *et al.*, 1987; Higgs *et al.*, 2002). Studies which have provided threshold values for the American alligator specifically differ quite markedly (Wever, 1971; Higgs *et al.*, 2002) however. The most likely cause for the differences is that the ABR is dominated by activity from neurons sensitive to higher frequencies from about 1 to 4 kHz (Kraus and McGee, 1992) and although quite good at estimating behavioral thresholds for higher frequencies is not as good for lower frequencies. For this reason active space estimates are made separately for each threshold obtained by the two methods.

At this time only the Higgs *et al.* (2002) study has measured alligator hearing in both air and water. This has indicated that alligator ABR responses in water compares favorably with those of the goldfish, which is classified as a specialist for hearing. Although the ABR is less reliable for estimating behavioral thresholds for low frequencies, it can allow us to determine the relative threshold between air and water. For 100–200 Hz this is about 37 dB, which is close to the 36 dB impedance mismatch between air and water. This can then be added to the threshold values for air obtained by cochlear potentials in order to provide a second estimate of hearing thresholds, and hence active space, in water.

There are currently no data available for otolith acoustic sensitivity of the alligator. However, given the remarkable sensitivity of the vertebrate otolith to bone-conducted sound, including in mammals and humans (Cuthoys *et al.*, 2006; Todd *et al.*, 2003), and that alligator hearing in water is comparable to that of fish we use the data that are available for near-field detection in fish for our estimates. It is appropriate that we use nonspecialist fish for comparison, however, as we are interested in the sensitivity of near-field detection. Specialist fish hearers make use of adaptations for the detection of the far-field sound pressure and hence audiograms derived from such species would not give a conservative estimate of near-field detection. Several studies have determined behavioral particle motion audiograms for nonspecialist species of fish, e.g., the cod, plaice, and dab (Fay and Simmons, 1999). These have indicated that the region of best sensitivity lies between 40 and 120 Hz with threshold PVL values as low as −30 dB 1 μm s^{-1} (corresponding PAL values of 20 dB ~ 1 μm s^{-2}) at 80 Hz. As an additional estimate for near-field detection in the alligator we also use data from the seismic sensitivity of the amphibian saccule which has been determined in several species, including *Rana temporaria* (Christensen-Dalsgaard and Jorgensen, 1988), *Rana catesbeiana* (Koyama *et al.*, 1982), and *Leptodactylus albilabris* (Narins and Lewis, 1984). Some species show a fish-like band-pass response to acceleration with best frequencies between 20 and 160 Hz and thresholds as low as 0.001 cm/s^2 or 20 dB ~ 1 μm s^{-2} while *Rana temporaria* shows a low-pass response with best frequencies at 10–20 Hz (Christensen-Dalsgaard and Jorgensen, 1988).

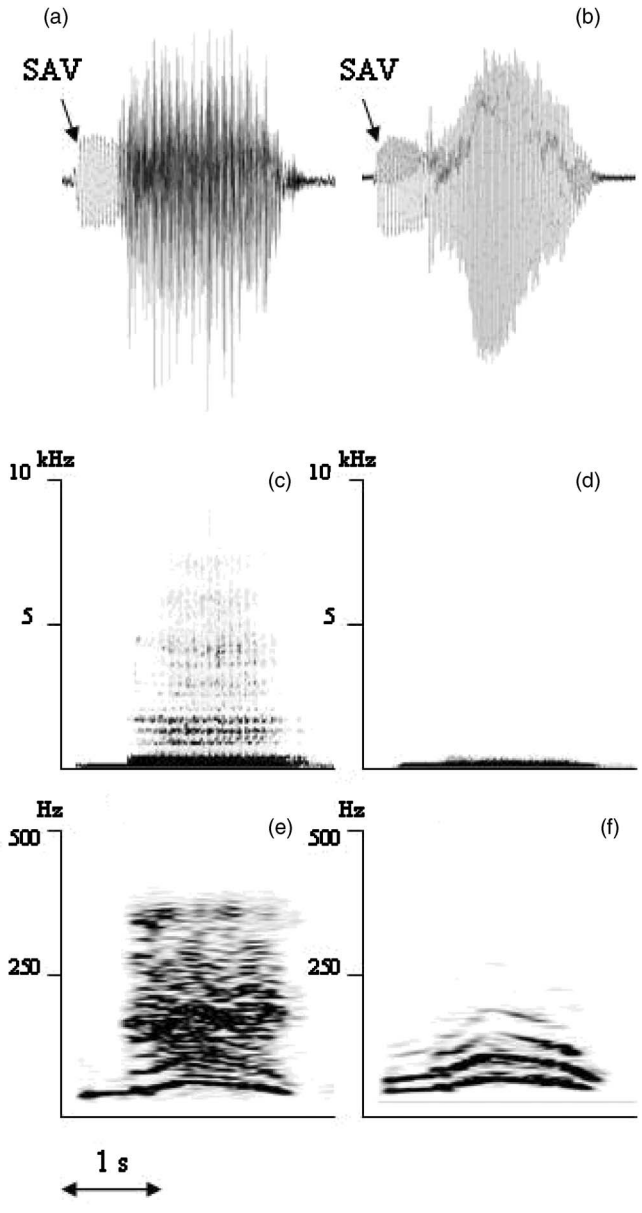


FIG. 2. A typical male alligator vocalization in air (left) and water (right). Wave form in air: (a) and water (b). Broadband spectrogram over 10 kHz in air (c) and water (d). Narrowband spectrogram concentrating on frequencies below 500 Hz for air (e) and water (f). The recording was made a distance of about 3 m in air and 2 m in water. The call examples are available at ⟨http://www.acoustics.org/press/151st/Todd.html⟩.

TABLE I. Descriptive statistics for background noise.

	Air				Water			
	L_{Lpk}	L_{LI}	L_{LF}	$L_{LF1/3}$	L_{Lpk}	L_{LI}	L_{LF}	$L_{LF1/3}$
Mean	75.2	66.7	65	43.6	87.6	79.4	76.9	50.5
SD	2.4	3.1	3.1	6.9	3.4	4.1	4.1	8.1
Max	77.9	70.6	68.9	52.9	92	86.4	84.4	67.4
Min	71.7	63.9	61.8	28.8	82.1	74.4	72.6	40.2

III. RESULTS

A. Background noise

Typically each recording session was made just after dawn and coincided with the chorusing of a number of Australian species of bird and other animals at the Reptile Park. On occasions the wind was quite strong and its interaction with the surrounding trees produced a fairly broadband noise. In addition traffic noise was also present to varying degrees from a distant freeway. Both broadband and spectral (1/3 octave) characteristics were determined for background noise in air and water and these are presented in Table I and Fig. 1. The 1/3 octave analysis shows that for both air and water the noise distribution exhibited a low-frequency maximum (16–100 Hz) dropping about 15–20 dB to a high-frequency plateau (400 Hz–5 kHz). The low-frequency feature for water though dropped away more rapidly at 100–200 Hz before rising again in the high frequencies (above 5 kHz the water spectrum is dominated by internal noise). Both air and water spectral profiles show features between 1 and 3 kHz, most likely due to bird calls. The broadband measures are largely influenced by the low-frequency features, which show an air-water difference of about 12 dB, reflecting the difference in the low-frequency region of the spectrum. The mean value of the 1/3 octave measures across the spectrum though are within 6 dB which reflects the fact that between 100 Hz and 1 kHz the air and water background is broadly equivalent in sound pressure.

B. Source characteristics

The wave form and narrowband spectrogram of a typical male vocalization is shown in Fig. 2, illustrating the air/water differences. For the male vocalization the first portion comprises the SAVs which are below the range of human hearing and are associated with a "water dance" above the Alligator's back. The audible part of the call, superimposed on the SAV, is almost entirely carried by the air-borne component which includes a formant like structure extending up to 10 kHz. In contrast the water-borne component shows little energy above 100 Hz.

The air–water differences above are also a feature of the 1/3 octave band modal analysis (Table II) which shows that the dominant frequencies were significantly higher in air than in water, by about one octave. There were also differences in the distribution of frequencies. The distribution for air was strongly skewed towards higher frequencies, hence the large difference between mode and mean, 40 and 94 Hz respectively, whereas for water the distribution was tightly grouped around the mode 31.5 Hz. This is also illustrated in Fig. 3 which shows the 1/3 octave analysis of the signal used above overlaid on the background.

C. Source intensity and transmission characteristics

Descriptive statistics for the acoustical measurements (Table III) derived from the samples indicate that the mean SPL was lower in air than in water, between 23 and 27 dB. From the regression analyses (Table III) estimates of source intensity (at 1 m) are made from the value of the constant parameter b_0 and these indicate air–water differences consistent with the mean. For air these range from 78 to 104 dB and for water a range of 105–129 dB, depending on the intensity measure and frequency band.

There are also air–water differences in the slope of the regressions which indicate differences in the transmission characteristics of the medium (illustrated in Fig. 4). For air linear regression produces slopes consistently less than −1 ($0.49 > b_1 > 0.74$) while that for water gave slopes close to −1 ($-0.93 > b_1 > -1.15$). The slopes could be taken to indicate that air is subject to cylindrical and water to spherical spreading. However, this seems counterintuitive on theoretical grounds, i.e., it would be expected to be the reverse of the observed. The most likely reason for the low values of the slopes for air is that with increasing distances the noise back-

TABLE II. Descriptive statistics for acoustical measurements of alligator vocal displays.

	Air					Water				
	L_{Lpk} (dB)	L_{LI} (dB)	$L_{LF(1/3)}$ (dB)	Freq. (Hz)	Dist. (m)	L_{Lpk} (dB)	L_{LI} (dB)	$L_{LF(1/3)}$ (dB)	Freq. (Hz)	Dist. (m)
Mean	83	75	68	94	26.6	107	102	91	35.5	26.4
Mode	80	71	62	40	20	98	97	85	31.5	20
SD	7	7	6	65	22.1	12	13	14	9.7	22.3
Min	75	65	57	16	1	86	79	71	20	1
Max	103	95	85	250	75	137	131	130	100	75

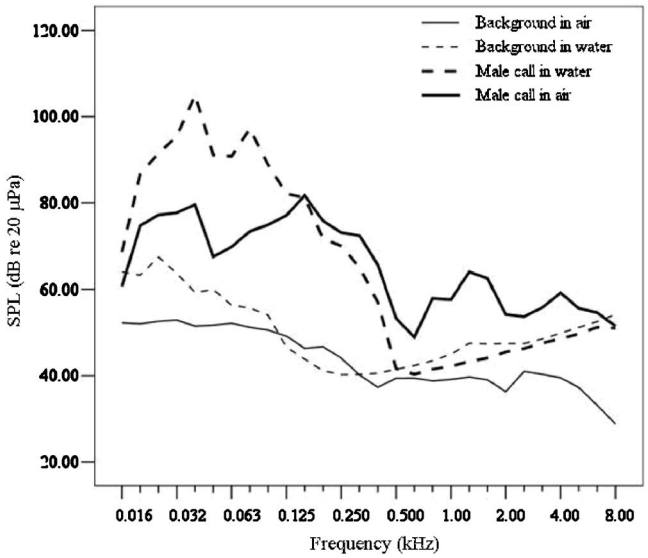


FIG. 3. A 1/3 octave spectral analysis of the same male alligator vocalization in air (thick continuous line) and water (thick broken line) shown against the background in air (thin continuous line) and water (thin broken line). All measurements are maximum values of SPL with fast linear weighting (L_{LF}) and ~20 μPa. Both air- and water-borne components of the vocalization lie well above the background with maximum signal-to-noise ratios of about 30 dB at 160 Hz for air and about 50 dB at 40 Hz for water.

ground increasingly influences the intensity measurement, particularly for the broadband parameters. In the case of water it is also possible that beyond a certain distance there may be an additional process affecting attenuation, e.g., a physical limit on propagation of low frequencies. This would explain why the quadratic regression yielded positive and negative values of the quadratic parameter b_2 for air and water, respectively.

D. Estimated active space in air and water

Having estimated the source and transmission characteristics in air and water we may attempt to estimate the active space as the limit set by either background noise or the thresholds. For air-conducted transmission the data obtained by Higgs *et al.* (2002) using the ABR method indicate that in the low-frequency region, i.e., up to 800 Hz the active space is essentially threshold limited as the threshold values are more than 1 SD above the mean high-water mark for the background. Only in the most sensitive range, between 800 Hz and 2 kHz, do the threshold values drop below the background. Above 2 kHz the active space is again threshold limited. However, as noted above and as shown in Fig. 5, the ABR estimates differ radically from estimates obtained by Wever (1971) using cochlear microphonics. This is particularly so for the low to midfrequency region where the estimates differ by as much as 50 dB. The Wever (1971) estimates therefore indicate that the alligator active space is primarily noise limited.

The 1/3 octave analysis showed that for air the vocalizations are dominated by low frequencies up to 250 Hz and analyses were carried out independently for four octave frequency bands 16–25, 31–50, 63–100, and 125–200 Hz. As the ABR threshold values are not available for frequencies below 100 Hz, estimates of active space based on 1/3 octave values in air were limited to only those values above 100 Hz, for which the threshold lies between 55 and 60 dB. The noise levels for each of the bands were available and the cochlear potential values were available for the top three bands. For the lowest frequency band threshold estimates were extrapolated from the neighboring band. Active space estimates were made using an ideal slope of −1 (spherical spreading). These estimates [Table IV(a)] indicate that for air transmis-

TABLE III. Parameter estimates from regression analyses.

	Air				Water			
	*b*0	*b*1	*b*2	*R*2	*b*0	*b*1	*b*2	*R*2
L_{Lpk}	101.3	−0.67		0.86	128.8	−0.93		0.77
	103.9	−1.12	0.011	0.88	127.1	−0.64	−0.008	0.78
L_{LI}	89.7	−0.60		0.80	124.7	−0.94		0.75
	94.0	−1.18	0.014	0.84	121.9	−0.46	−0.13	0.77
L_{EQ}	78.9	−0.49		0.69	116.0	−1.04		0.78
	83.6	−1.10	0.015	0.78	113.7	−0.64	−0.01	0.79
$L_{LF(1/3)}$	83.9	−0.67		0.81	122.0	−1.04		0.77
	86.8	−1.06	0.009	0.82	118.8	−0.48	−0.015	0.79
$L_{LF(1/3)}$ 16–25 Hz	81.8	−0.54		0.91	120.1	−1.01		0.86
	85.8	−1.14	0.015	0.96	119.6	−0.53	−0.014	0.88
$L_{LF(1/3)}$ 31–50 Hz	79.7	−0.55		0.74	122.4	−1.05		0.78
	80.7	−0.66	0.002	0.74	118.4	−0.55	−0.012	0.80
$L_{LF(1/3)}$ 63–100 Hz	84.1	−0.68		0.86	108.4	−0.99		0.67
	86.6	−0.99	0.007	0.87	106.8	−0.68	−0.009	0.68
$L_{LF(1/3)}$ 125–200 Hz	86.0	−0.74		0.85	104.6	−1.15		0.96
	88.3	−1.07	0.009	0.86	104.8	−1.90	0.015	0.97

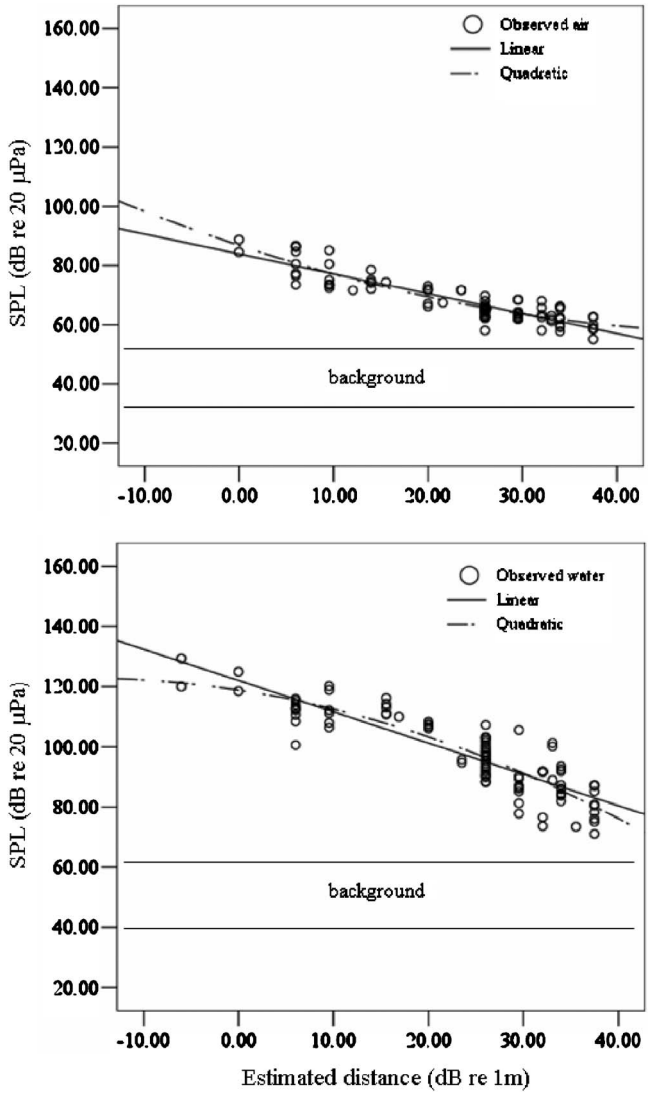


FIG. 4. Linear vs quadratic regression of SPL (L_{LF}) as a function of estimated distance for (top) air and (bottom) water, obtained from the 1/3 octave modal analysis. The distance values are plotted in dB ~1 m, i.e., 20 $\times \log_{10}$ (distance).

sion with spherical spreading the active space if limited by the measured noise (or threshold in the 16–25 Hz band) would range to 38, 44, 98, and 159 m for each of the bands from the lowest upwards. Thus for air-borne transmission with the noise measured at the Reptile Park the active space was quite limited, especially for the lowest bands. In principle, however, in a lower noise environment the active space could range to several kilometers for the highest frequency band.

For water-borne transmission [Table IV(b)] the ABR threshold values lie well above background over the entire hearing range (Fig. 6), even in the most sensitive region of 0.4–1 kHz, so that the active space, if defined by the ABR response, would be entirely threshold limited for water. However, in this case there were no dominant modes in a frequency range for which there were threshold values. Indeed, as is clearly illustrated in Fig. 6, for the example male vocalization the entire 1/3 octave spectrum above 100 Hz lies below the ABR thresholds in water. Nevertheless, an

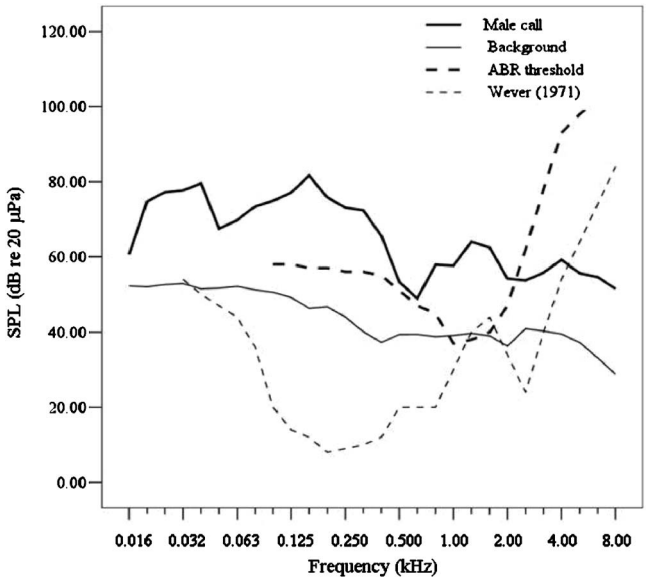


FIG. 5. 1/3 octave analysis of the male alligator vocalization in air (thick line) compared with the background (thin line), the ABR threshold curve (thick broken line) from Higgs *et al.* (2002) and the cochlear potentials threshold curve (thin broken line) from Wever (1971).

estimate was made using the threshold values at 100 Hz, which is about 120 dB ~ 1 μPa or 94 dB ~ 20 μPa. Figure 6 also illustrates the Wever (1971) threshold values elevated by 37 dB. For a stimulus of 30 Hz in air Wever (1971) measured a threshold of −20 dB ~ 1 dyn/cm^2 which corresponds to 54 dB ~ 20 μPa. Thus a reasonable guess at the threshold in water for 30 Hz is about 90 dB ~ 20 μPa. For high frequencies the two threshold curves match up quite well but below 800 Hz the two estimates diverge by 40 dB or more. However, at no point did the alternative threshold values drop below background, thus confirming that for transmission in water the active space is limited by hearing sensitivity.

Analyses for the water-borne component were also carried out independently for the same four octave frequency bands, i.e., 16–25, 31–50, 63–100, and 125–200 Hz, and the results shown in Table IV(b). These indicate that if limited by the adjusted Wever values the active space in water would range to 51 m, 104 m, 1.5 km, and 213 m for each of the bands from the lowest upwards. Thus, like the air-borne component, the active space is quite limited in the lowest bands but potentially quite extensive for the 63–100 Hz band. The noise based estimates show the opposite trend, i.e., they tend to decrease in range with increasing frequency, thus reflecting the signal-to-noise ratio for each of the bands. The ABR based estimates confirm that the active space would be extremely limited, no more that 14 m, if these were an accurate representation of behavioral thresholds.

Table V shows estimations of active space for near versus far-field detection for the low-frequency 1/3 octave bands, for which there were sufficient data, i.e., 20, 31, 40, and 63 Hz. The far-field estimates were based on the adjusted Wever (1971) values and independent estimates of the near field were made for both the fish and frog saccular sensitivity data using the procedure as defined in the methods.

TABLE IV. Estimated active space for hearing in air.

	Frequency band (Hz)	SPL at 1 m	Slope	Threshold or noise limit	Estimated active space
(a)	16–25	84.2–87.6	−1.0	52–53 (Noise)	35–60 m
				56–62 (Wever)	12–38 m
	31–50	77.5–83.9	−1.0	51–52 (Noise)	19–44 m
				47–54 (Wever)	15–69 m
	63–100	83.4–89.8	−1.0	55–60 (ABR)	14–55 m
				50–52 (Noise)	37–98 m
				20–44 (Wever)	93 m–3.1 km
	125–200	86.5–90.1	−1.0	52–58 (ABR)	26–80 m
				46–49 (Noise)	74–159 m
				8–14 (Wever)	4–13 km
(b)	16–25	116.7–123.6	−0.98–−1.10	93–99 (Wever+37 dB)	6–51 m
				63–67 (Noise)	158 m–2.5 km
	31–50	120.9–124.0	−0.99–−1.11	84–91 (Wever+37 dB)	22–104 m
				60–64 (Noise)	363 m–1.7 km
	63–100	105.3–111.5	−0.86–−1.12	92–96 (ABR)	3–14 m
				57–81 (Wever+37 dB)	12 m–1.5 km
				54–57 (Noise)	143–2.2 km
	125–200	101.8–107.4	−1.34–−1.56	92–96 (ABR)	2–4 m
				45–51 (Wever+37 dB)	42–213 m
				41–47 (Noise)	57–300 m

The far-field estimates were similar to those made for Table IV(b) and show a trend to increase range with frequency. In contrast both near-field estimates show a best range at 31 Hz extending to 80 m, which is greater than the 72 m estimated for the far field. Thus if the alligator saccule is endowed with the same sensitivity as that of the frog it could well be a more effective receiver at low frequencies than the basilar papilla.

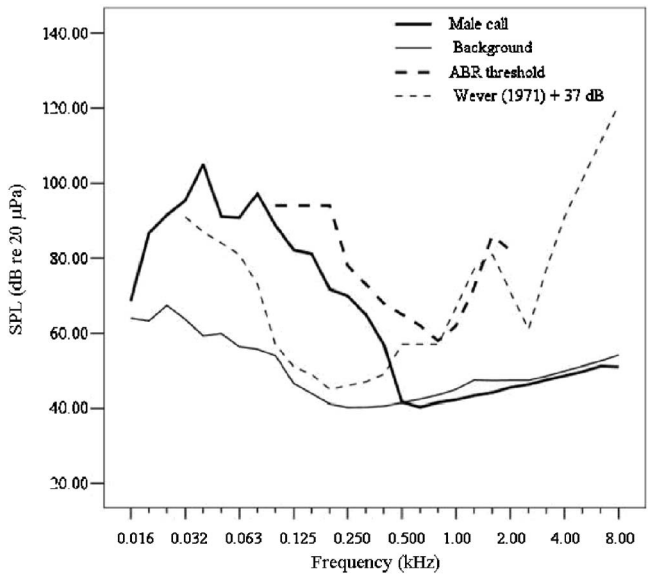


FIG. 6. 1/3 octave analysis of the male alligator vocalization in water (thick line) compared with the background (thin line), the ABR threshold curve (thick broken line) from Higgs *et al.* (2002) and the cochlear potentials threshold cure (thin broken line) from Wever (1971) adjusted by 37 dB. The adjustment value was obtained from the difference between the ABR thresholds in air and water at 100 Hz.

IV. DISCUSSSION

This study of alligator vocal displays has obtained source intensity estimates of 85–90 and 121–124 dB SPL (L_{LF} measured for each 1/3 octave band) in air and water, respectively. In conjunction with existing data and for the measured noise level these have yielded active space estimates of 169 m and 1.5 km for hearing in air and water, respectively, the 125–200 and 63–100 Hz octave bands, respectively. These are comparable with the dimensions of the alligator pond and suggest that individuals would be able to detect conspecific vocal displays from most points within their immediate environment in air, and possibly more extensively in water. They would also be completely aware of most of the environmental airborne sound sources which contribute to the background. There are, however, several reasons for caution in accepting these results.

The most important reason for caution is that as shown above (Fig. 5) very different estimates of hearing capability, as much as 40–50 dB, can be obtained by the different methods. Further, the Higgs *et al.* (2002) data were obtained from juvenile alligators and they speculate that the hearing sensitivity may scale with age and size. Wever's (1971) thresholds were also obtained from young animals but the method of cochlear potentials produced estimates that are comparable

TABLE V. Estimated active space for near versus far-field detection in water.

1/3 octave frequency band (Hz)	Dependent variable	Intensity at 1 m	Slope	Threshold (source)	Active space (m)
20	SPL re 20 μPa	115.7–121.3	−0.72–−0.94	96 (Wever+37 dB)	11–59
	PVL re 1 μm s^{-1}	39.8–45.4	−1.72–−1.94	−10–10 (fish)	4–25
	PAL re 1 μm s^{-2}	81.7–87.3	−1.72–−1.94	20–46 (frog)	5–51
31.5	SPL re 20 μPa	118.8–122.4	−0.84–−1.01	90 (Wever+37 dB)	27–72
	PVL re 1 μm s^{-1}	38.9–42.4	−1.84–−2.01	−20–−5 (fish)	14–55
	PAL re 1 μm s^{-2}	84.8–88.4	−1.84–−2.01	20–46 (frog)	10–80
40	SPL re 20 μPa	122.0–127.6	−1.02–−1.26	87 (Wever+37 dB)	24–98
	PVL re 1 μm s^{-1}	40.0–45.6	−2.02–−2.26	−25–−10 (fish)	13–56
	PAL re 1 μm s^{-2}	88.0–93.6	−2.02–−2.26	20–46 (frog)	8–67
63	SPL re 20 μPa	104.1–113.7	−0.76–−1.19	87 (Wever+37 dB)	9–136
	PVL re 1 μm s^{-1}	18.1–27.7	−1.76–−2.19	−30–−15 (fish)	5–43
	PAL re 1 μm s^{-2}	70.1–79.7	−1.76–−2.19	20–46 (frog)	4–48

with those obtained by single unit recordings in other species. For example, Wever's (1971) sensitivity measurements for *caiman crocodilus* are quite close to absolute fiber thresholds obtained by Kinke and Pause (1980). Both indicate that in the most sensitive range thresholds in air can be as low as about 5 dB SPL ~20 μPa and *caiman crocodilus* has broadly similar hearing to *alligator mississippiensis*. We can be confident that the ABR method does significantly overestimate behavioral thresholds for lower frequencies below 1 kHz. In contrast cochlear potentials are known to give a good representation of behavioral thresholds, certainly in the case of mammals including humans (Kraus and McGee, 1992). Nevertheless, there remains the uncertainty as to whether data obtained from young alligators is valid for a fully grown adult.

In the case of air the highest estimate based on spherical spreading limited by the measured noise was 169 m. It may be possible that due to reflection from the water surface the physical situation approximates two monopole sources of opposite polarity, which would effectively double the power of the source (Rossing and Fletcher, 1995). However this would not affect the spreading law away from the source and more importantly our estimates are based on intensity measurements at variable distances, which would include any effects of reflection. Although the 1/3 octave measurements are less prone to the effect of background noise it is likely, as noted above, that the empirical slopes for air were affected by the background. Hence we should conclude that the upper limit of 169 m as defined by the background at the Australian Reptile Park is reasonable. It is possible that in an environment with a lower level of noise that the active space in air could be significantly larger. For example, with a background of 20–40 dB SPL, as might be found in a quiet rural environment, the active space in air could extend to several kilometers.

In the case of the water-borne component, unlike the case for air, most of the power was concentrated in very low frequencies below 100 Hz and the signal-to-noise ratio was highest in the lowest frequency bands, as illustrated by the estimate of the range of active space if defined by the background. And yet the estimates based on thresholds show the same trend as with air, i.e., that the most effective bands for transmission and reception were not the lowest, for which the range was limited to about 100 m. Nevertheless, for the Wever (1971) adjusted threshold values, the active space at about the 1.5 km range in the 63–100 Hz band appears to be quite extensive. We should, though, treat this estimate with the same caution as outlined above. For water-borne transmission we have the additional uncertainty that we do not have any direct measurements of thresholds below 100 Hz, although the adjustment by 37 dB is likely a good guess. The measurements that we do have based on the ABR at 100 Hz and above would indicate an extremely limited active space in water. Also there must remain some doubt over the effectiveness of low-frequency transmission in shallow water (Rossing and Fletcher, 1995). Our estimates were based on extrapolation from measurements up to only 100 m and the quadratic regression curves suggest an additional process which could be transmission loss. A further complication arises from consideration of transmission from water to the underlying solid substrate in the form of seismic energy (O'Connell-Rodwell *et al.*, 2001), particularly if an alligator is in contact with the bottom when vocalizing or receiving.

The clear trend of both air and water estimates is that there appears to be a disparity between the low-frequency characteristics of the signal and the receiver sensitivity based on detection by the basilar papilla. This apparent signal-receiver mismatch is particularly prominent for the near-infrasound SAV component of the male call in water. This suggests that an alternative sensory mechanism may potentially be more ideally suited for reception of the low-frequency component of the alligator call and the issue of near-field detection becomes all the more important. The analysis of the near-field active space given in Table V was based on two independent estimates: one for particle velocity threshold values in nonspecialist fish (Fay and Simmons, 1999) and the other for particle acceleration thresholds in frogs (Narins and Lewis, 1984). Both estimates produce es-

sentially the same picture which is that an otolithic hearing mechanism would indeed be well matched to the low-frequency signal profile giving a best range in the 31 Hz 1/3 octave band of up to 80 m. This particular estimate was made assuming that the alligator saccule could match the seismic sensitivity of the frog saccule in responding to peak accelerations of less than 0.001 cm/s^2 (Narins and Lewis, 1984). It is generally accepted that the amphibian ear makes use of several sensory end organs, including the basilar and amphibian papillae, but where the saccule has retained a specialization for low-frequency substrate-borne sound (Smotherman and Narins, 2004). It seems plausible that stem reptiles could have derived a similar condition from the basal tetrapods and that the saccule may have conserved a similar amphibian-like role in some groups of amniotes, including crocodilians.

While there remains some uncertainty surrounding the estimates, the results do allow us to infer that the active space must be a complex one. In the most distal region of the active space it is defined by the level of background noise in air. For the background at the Australian Reptile Park this was up to about 170 in the 125–200 Hz band. In water the signal-to-noise ratio is much greater and the distal region of the active space is defined by hearing thresholds. Assuming the adjusted Wever (1971) thresholds and no transmission loss this could extend to 1 km or so in the 63–100 Hz band. In the proximal region of the active space, possibly up to 80 m, the sensitivity of the inner ear to both near-field and far-field components of water-borne sound is the defining factor. Such a hierarchy of range within the active space could both reflect two distinct functions of acoustic sensing in crocodilian behavior—that of scene analysis and reproduction, respectively. The basilar papilla is ideally suited to fulfill the first function, that is to be able to detect, localize, and recognize a conspecific (or indeed potential prey).

With regard to the second function, alligator bellowing is known to be associated with powerful autonomic responses and olfactory signals via musk glands (Vliet, 1989). While papilla input to the autonomic nervous system could mediate such responses, the projections are not well described in amniotes. In amphibians an acoustic autonomic pathway via the secondary isthmal nucleus (SIN), including projections to hypothalamus and amygdala, is specifically engaged in reproductive vocal behavior (Neary, 1988). The SIN receives inputs from the dorsal medulary nucleus (DMN) (a homologue of the amniote cochlear nucleus) and the DMN in turn receives inputs from both saccule and basilar papilla (Horowitz *et al.*, 2007). Phylogenetic comparisons, however, suggest that the SIN is homologous to the parabrachial nucleus (PBN) (ten Donkelaar, 1988a) which is known to be a visceral/gustatory relay station in amniotes (ten Donkelaar, 1988b; Voogd *et al.*, 1988), receiving afferent fibers from the nucleus of the solitary tract, the brain stem nucleus for visceral, and gustatory sensory input. The PBN also receives strong vestibular input and is believed to mediate conventional vestibular-autonomic interactions (Balaban *et al.*, 2002), e.g., in regulating blood pressure for different postures. It has been thought, though, that the sacculus in amphibians has a dual auditory/vestibular function (Horowitz *et al.*, 2007) and this duality could extend to autonomic responses. Thus a saccular-parabrachial projection could mediate a reward function associated with inertial vestibular sensations of self-motion (Todd, 2001; Todd and Merker, 2004) and may plausibly have a reproductive function in response acoustic activation.

The alligators are ancient representatives of the archosaurs, which include both dinosaurs and birds, a group whose ears are similar in structure (Manley, 2000), along with other similarities, such as egg laying, parenting, and vocal behavior. If the otolithic hearing mechanism is proven to be present in this group, as well as in fish and amphibians, then it is likely that it is widespread in all vertebrates, including mammals and humans. Given its possible role in mediating reproductive vocal behavior then it may well have played a central, but hitherto overlooked, role in evolution. Clearly, in order to resolve this issue it is a matter of some urgency for further data to be obtained on the comparative sensitivity of the basilar papilla and sacculus to low frequency and near infrasound in air and water. It is also a matter of some importance to determine the degree to which the low-frequency component may be transmitted in the underlying substrate as seismic energy.

ACKNOWLEDGMENTS

I am grateful to the Australian Reptile Park for permission to record the alligators and to keeper Craig Adams for assistance in working with the alligators. I am also grateful to Dr. Doug Cato of the Australian Defence Science and Technology Organisation (DSTO) for providing technical assistance in the use and calibration of the hydrophone in testing the equipment. The work was carried out using equipment purchased from a Royal Society grant.

Balaban, C. D., McGee, D. M., Zhou, J., and Scudder, C. A. (**2002**). "Responses of primate caudal parabrachial nucleus and Kölliker-Fuse nucleus neurons to whole body rotation," J. Neurophysiol. **88**(6), 3175–3193.

Beach, F. A. (**1944**). "Response of captive alligators to auditory stimulation," Am. Nat. **78**, 481–505.

Brenowitz, E. (**1982**). "The active space of red-winged blackbird song," J. Comp. Physiol. **147**, 511–522.

Brown, C. H. (**1989**). "The active space of blue monkey and grey cheeked magabey vocalisations," Anim. Behav. **37**, 1023–1034.

Christensen-Dalsgaard, J., and Jorgensen, M. B. (**1988**). "The response characteristics of vibration-sensitive saccular fibers in the grassfrog, *Rana temporaria*," J. Comp. Physiol. **162**, 633–638.

Cuthoys, I. S., Kim, J., McPhedran, S. K., and Camp, A. J. (**2006**). "Bone conducted vibration selectively activates irregular primary otolithic vestibular neurons in the guinea pig," Exp. Brain Res. **175**, 256–267.

Fay, R. R., and Simmons, A. M. (**1999**). "The sense of hearing in fishes and amphibians," in *Comparative Hearing: Fish and Amphibians*, edited by R. Fay and A. Popper (Springer, New York), pp. 269–318.

Fritzsch, B. (**1999**). "Hearing in two worlds," in *Comparative Hearing: Fish and Amphibians*, edited by R. Fay and A. Popper (Springer, New York), pp. 15–42.

Higgs, D. M., Brittan-Powell, E. F., Soares, D., Souza, M. J., Carr, C. E., Dooling, R. J., and Popper, A. N. (**2002**). "Amphibious auditory responses of the American alligator (*Alligator mississippiensis*)," J. Comp. Physiol. **188**, 217–223.

Horowitz, S. S., Tanu, L. H., and Simmons, A. M. (**2007**). "Multiple mechanosensory modalities influence development of auditory function," J. Neurosci. **27**(4), 782–790.

Janik, V. (**2000**). "Source levels and estimated active space of bottlenose dolphin (*Tursiops truncatus*) whistles in the Moray Firth, Scotland," J. Comp. Physiol. **186**, 673–680.

Kinke, R., and Pause, M. (**1980**). "Discharge properties of primary auditory fibres in *Caiman crocodilus*," Exp. Brain Res. **38**, 137–150.
Koyama, H., Lewis, E. R., Leverenz, E. L., and Baird, R. (**1982**). "Acute seismic sensitivity in the bullfrog ear," Brain Res. **250**, 168–172.
Kraus, N., and McGee, T. (**1992**). "Electrophysiology of the human auditory system," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. Popper and R. Fay (Springer, New York), pp. 335–404.
Lewis, E. R., Leverenz, E. L., and Bialek, W. S. (**1985**). *The Vertebrate Inner Ear* (CRC Press, Boca Raton, FL).
Lewis, E. R., and Narins, P. M. (**1985**). "Do frogs communicate with seismic signals?," Science **227**, 187–189.
Lewis, E. R., and Narins, P. M. (**1999**). "The acoustic periphery of amphibians: Anatomy and physiology," in *Comparative Hearing: Fish and Amphibians*, edited by R. Fay and A. Popper (Springer, New York), pp. 43–100.
Manley, G. A. (**2000**). "Cochlear mechanisms from a phylogenetic viewpoint," Proc. Natl. Acad. Sci. U.S.A. **97**(22), 11736–11743.
Miller, J. O. (**2006**). "Diversity in sound pressure levels and estimated active space of resident killer whale vocalisations," J. Comp. Physiol. **192**, 449–459.
Narins, P. M., and Lewis, E. R. (**1984**). "The vetebrate ear as an exquisite seismic sensor," J. Acoust. Soc. Am. **76**(5), 1384–1387.
Neary, T. J. (**1988**). "Forebrain auditory pathways in Ranid frogs," in *The Evolution of the Amphibian Auditory System*, edited by M. J. Fritzsch, W. Ryan, W. Wilcynski, T. E. Hertherington, and W. Walkowiak (Wiley, New York), pp. 233–352.
O'Connell-Rodwell, C. E., Hart, L. A., and Arnason, B. T. (**2001**). "Exploring the potential use of seismic waves as a communication channel by elephants and other large mammals," Am. Zool. **41**, 1157–1170.
Popper, A. N., and Fay, R. R. (**1999**). "The auditory periphery in fishes," in *Comparative Hearing: Fish and Amphibians*, edited by R. Fay and A. Popper (Springer, New York), pp. 43–100.
Rossing, T. D., and Fletcher, N. H. (**1995**). *Principles of Vibration and Sound* (Springer, New York).
Smotherman, M., and Narins, P. M. (**2004**). "Evolution of the amphibian ear," in *Evolution of the Vertebrate Auditory System*, edited by G. Manley, A. Popper, and R. Fay (Springer, New York), pp. 164–199.
Southall, B. L., Schusterman, R. J., and Kastak, D. (**2003**). "Acoustic communication ranges for northern elephant seals (*Mirounga angustirostris*)," Aquatic Mammals **29**, 202–213.
Strain, G. M., Tucker, T. A., Graham, M. C., and O'Malley, N. A. (**1987**). "Brain-stem auditory evoked potentials in the caiman (*Caiman crocodilius*)," Electroencephalogr. Clin. Neurophysiol. **75**, 97–104.
ten Donkelaar, H. J. (**1988a**). "Anurans," in *The Central Nervous System of Vertebrates*, edited by R. Niewenhuys, H. J. Donkelaar, and C. Nicholson (Springer, Berlin), Vol. **2**, pp. 1151–1314.
ten Donkelaar, H. J. (**1988b**). "Reptiles," in *The Central Nervous System of Vertebrates*, edited by R. Niewenhuys, H. J. Donkelaar, and C. Nicholson (Springer, Berlin), Vol. **2**, pp. 1315–1524.
Todd, N.P. McA. (**2001**). "Evidence for a behavioural significance of saccular acoustic sensitivity in humans," J. Acoust. Soc. Am. **110**(1), 380–480.
Todd, N.P. McA., and Merker, B. (**2004**). "Siamang gibbons exceed the saccular threshold," J. Acoust. Soc. Am. **115**(6), 3077–3080.
Todd, N.P. McA., Rosengren, S. M., and Colebatch, J. G. (**2003**). "A short latency vestibular evoked potential (VsEP) produced by bone-conducted acoustic stimulation," J. Acoust. Soc. Am. **114**(6), 3264–3272.
Vliet, K. A. (**1989**). "Social displays of the American alligator (*Alligator mississippiensis*)," Am. Zool. **29**, 1019–1031.
Voogd, J., Niewenhuys, R., van Dongen, P. A. M., and ten Donkelaar, H. J. (**1988**). "Mammals," in *The Central Nervous System of Vertebrates*, edited by R. Niewenhuys, H. J. Donkelaar, and C. Nicholson (Springer, Berlin), Vol. **3**, pp. 1637–2098.
Welgampola, M. S., Rosengren, S. M., and Colebatch, J. G. (**2003**). "Vestibular activation by bone-conducted sound," J. Neurol., Neurosurg. Psychiatry **74**, 771–778.
Wever, E. G. (**1971**). "Hearing in the Crocodilia," Proc. Natl. Acad. Sci. U.S.A. **68**, 1498–1500.

Onset, growth, and recovery of in-air temporary threshold shift in a California sea lion (*Zalophus californianus*)

David Kastak[a] and Colleen Reichmuth
Institute of Marine Sciences, University of California, Santa Cruz, Long Marine Laboratory, 100 Shaffer Road, Santa Cruz, California 95060

Marla M. Holt
National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 2725 Montlake Boulevard, East Seattle, Washington 98112

Jason Mulsow
Department of Ocean Sciences, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California 95064

Brandon L. Southall
Institute of Marine Sciences, University of California, Santa Cruz, Long Marine Laboratory, 100 Shaffer Road, Santa Cruz, California 95060 and
NOAA Ocean Acoustics Program, National Marine Fisheries Services, Office of Science and Technology, 1315 East-West Highway, SSMC3 12359, Silver Spring, Maryland 20910

Ronald J. Schusterman
Institute of Marine Sciences, University of California, Santa Cruz, Long Marine Laboratory, 100 Shaffer Road, Santa Cruz, California 95060

(Received 23 July 2007; accepted 10 August 2007)

A California sea lion (*Zalophus californianus*) was tested in a behavioral procedure to assess noise-induced temporary threshold shift (TTS) in air. Octave band fatiguing noise was varied in both duration (1.5–50 min) and level (94–133 dB re 20 μPa) to generate a variety of equal sound exposure level conditions. Hearing thresholds were measured at the center frequency of the noise (2500 Hz) before, immediately after, and 24 h following exposure. Threshold shifts generated from 192 exposures ranged up to 30 dB. Estimates of TTS onset [159 dB re (20 μPa)^2 s] and growth (2.5 dB of TTS per dB of noise increase) were determined using an exponential function. Recovery for threshold shifts greater than 20 dB followed an 8.8 dB per log(min) linear function. Repeated testing indicated possible permanent threshold shift at the test frequency, but a later audiogram revealed no shift at this frequency or higher. Sea lions appear to be equally susceptible to noise in air and in water, provided that the noise exposure levels are referenced to absolute sound detection thresholds in both media. These data provide a framework within which to consider effects arising from more intense and/or sustained exposures.
 [DOI: 10.1121/1.2783111]

PACS number(s): 43.80.Nd, 43.80.Lb [WWA] Pages: 2916–2924

I. INTRODUCTION

Noise-induced temporary threshold shift (TTS) is the reversible elevation in hearing threshold caused by a fatiguing auditory stimulus. As a step toward understanding the potential impacts of underwater anthropogenic noise on marine life, research efforts have focused on conducting laboratory studies with the intent of examining TTS in marine mammals. Marine mammals that have been tested for susceptibility to TTS are the beluga (*Delphinapterus leucas*), bottlenose dolphin (*Tursiops truncatus*), California sea lion (*Zalophus californianus*), harbor seal (*Phoca vitulina*), and northern elephant seal (*Mirounga angustirostris*) [(Kastak and Schusterman, 1996; Kastak *et al.*, 1999; Finneran *et al.*, 2000, 2002, 2003; Nachtigall *et al.*, 2003, 2004; Finneran *et al.*, 2005; Kastak *et al.*, 2005)]. With the exception of a preliminary observation by Kastak and Schusterman (1996), none of these studies has investigated the effects of aerial noise exposure on hearing sensitivity.

Over the past several years, concerns have focused on underwater anthropogenic noise because of increases in overall level as well as correlations between cetacean (whales, dolphins, and porpoises) strandings and military operations using sonar (see, e.g., Frantzis, 1998; Fernandez *et al.*, 2005). Within marine mammals, the cetaceans are fully aquatic and therefore not considered to be susceptible to damaging effects of airborne noise; consequently, most of the literature regarding hearing loss in marine mammals has focused on underwater noise exposure. The pinnipeds (seals, sea lions, and walruses), however, represent a special case in any discussion of noise-induced TTS, because this group is amphibious and is therefore subject to the effects of noise in

[a]Electronic mail: kastak@ucsc.edu

air as well as under water. Although underwater noise could disrupt foraging and predator avoidance in pinnipeds, airborne noise has the potential to disrupt social functions that involve acoustic communication, such as mother-pup bonding and breeding behavior. Furthermore, depending on the rate of recovery, hearing loss induced in air may have major impacts on auditory sensitivity under water and vice versa. Therefore, the present study was designed to examine the effects of airborne noise exposure on a California sea lion, an animal that makes frequent transitions from land to water.

The subject of this experiment had previously participated in two studies involving underwater noise exposure designed to induce small to moderate levels of recoverable hearing loss (Kastak *et al.*, 1999, 2005). These earlier experiments showed that small amounts (<10 dB) of TTS could be induced by exposure to octave bands of noise at or above 80 dB relative to threshold at the center frequency of the noise band.

Despite the results obtained in the underwater studies, there are several factors that make the noise exposure component of TTS experiments more desirable to conduct in air than in water. Under water the subject must surface in order to breathe and to obtain food reinforcement, causing variability in the received sound field. In air, only a small movement of the head is required to receive reinforcement, provided it is delivered to a precise location close to the subject's head. Further consistency in noise exposure and in threshold determination can be ensured by testing in a sound-attenuating hemi-anechoic chamber, which has the additional benefit of reducing interference from outside noises that would mask the test signal or distract the subject. Finally, in-air sessions can be conducted more rapidly than those taking place under water, permitting the rapid acquisition of the large data set needed to test for statistical significance and to obtain robust estimates of TTS onset, growth, and recovery. For these reasons, the present experiment, designed to induce TTS of 10 dB or more, was conducted in air. This level was chosen because at low noise levels (smaller threshold shifts), growth of TTS is small and therefore might be underestimated unless higher threshold shift levels (corresponding to faster growth rates) are attained.

In order to further examine the effects of noise exposure on pinniped hearing, the experimental goals were to (1) show that TTS in air is correlated with some aspect of sound magnitude; (2) determine whether TTS in air and underwater are similar when the exposure levels are related to the subject's thresholds in air and under water; and (3) assess the time course of recovery from threshold shifts induced by noise levels of up to 104 dB above the subject's threshold.

In the discussion of the exposure parameters and data obtained in this experiment, the following abbreviations are used (ANSI, 1994): SPL=sound pressure level (dB re 1 μPa in water and dB re 20 μPa in air); SL=sensation level, or the difference between noise SPL and the subject's auditory threshold (dB); SEL=sound exposure level [dB re 1 μPa^2 s in water and dB re (20 μPa)^2 s in air]. Sound exposure level was used to quantify the fatiguing stimulus so that both sound magnitude and temporal variables could be taken into account simultaneously.

II. MATERIALS AND METHODS

A. Subject

The experimental subject was a female California sea lion (*Zalophus californianus*) named "Rio," whose age ranged from 17 to 20 years during the course of the experiment. She was housed at Long Marine Laboratory in Santa Cruz, CA in pools filled with free-flowing seawater surrounded by haulout space. She was maintained on a fish diet comprising herring and capelin, and received approximately 50%–75% of her daily ration during experimental sessions.

Rio had extensive prior experience in behavioral psychoacoustical testing, including underwater and aerial audiometry (Kastak and Schusterman, 1998; Southall *et al.*, 2005), auditory masking (Southall *et al.*, 2003, 2005), in-air sound localization (Holt *et al.*, 2005), and underwater noise-induced TTS (Kastak *et al.*, 1999, 2005). Prior training in these tasks facilitated acquisition of the behaviors involved in the current study. The protocols used in all stages of training and testing were approved by the UCSC Chancellor's Animal Research Committee. Research was conducted under NMFS Permit Nos. 259-1481-00 and 1072-1771-00.

B. Apparatus

Testing was conducted in a hemi-anechoic chamber (Eckel Industries), located next to the subject's living space. The chamber was a 2.5 m tall rectangular structure that was divided into a 3×5.6 m experimental space where the subject was tested, and a 3×1.4 m control room where an experimenter, an assistant, and the controlling equipment were located. The experimental space and control room were double-walled to eliminate environmental noise contamination. The walls and ceiling of the experimental space were lined with fiberglass-filled stainless steel wedges, and the concrete floor was covered with 2.6-cm-thick closed-cell neoprene. These features provided a relatively uniform acoustic space. Ventilation was provided by quieted fans that were not acoustically coupled to the chamber.

Inside the experimental space, a threshold station, response target, two speakers, and trial light were used for audiometric testing. The threshold station was a 33 cm tall PVC stand atop of which a PVC chin cup was mounted. The 10×10 cm response paddle was mounted 50 cm to the left of the threshold station at the level of the chin cup. The test tone projector, located on-axis to the chin cup at a distance of 70 cm, was a JBL 2123H midrange speaker. The trial-indicator light was placed just below this speaker, and was used to denote the interval of each experimental trial. The second speaker was placed on the floor of the chamber next to a PVC tube that was used to deliver fish reinforcement to the subject. The second speaker emitted an acoustic tone (bridge stimulus) to signal the subject that fish would be delivered.

To the right of the threshold station, a noise exposure station, a noise-projecting speaker, and a station-indicator light were configured for use during noise exposures. The noise exposure station was identical to the threshold station but placed 60 cm to its right and at an angle of approximately 45° to the right. This station faced the noise exposure

speaker that was centered 35 cm in front of it. The noise exposure speaker was either a Fender Princeton Chorus Guitar Amplifier or a Community EM280 compression driver coupled to a P100 horn projector. The light mounted just above the noise exposure speaker was used to cue the subject to remain positioned at the noise generating speaker until reinforcement was provided. During exposure, reinforcement was delivered from the same location as in threshold testing.

C. Acoustic stimuli and measurement

In order to provide continuity from prior underwater testing, pure tone thresholds were determined using a 2.5 kHz stimulus and noise exposures comprised an octave band centered at 2.5 kHz. The center frequency of the noise band was chosen in favor of a frequency $\frac{1}{2}$ octave above the center frequency because previous data from the same subject showed lack of a $\frac{1}{2}$ octave effect (Kastak *et al.*, 2005). Audiometric testing was semiautomated, using LABVIEW (National Instruments) virtual instruments. Pure-tone signals were converted from digital to analog at an update rate of 32 kHz using a National Instruments PXI-6070 multifunction DAQ device housed in a PXI 8176 controller. The pure-tone signals were 500 ms in duration with 40 ms linear rise/fall times and were projected from the JBL 2123H speaker. Octave-band noise with a center frequency of 2.5 kHz was used for noise exposure. This fatiguing noise was generated on the PXI board, digitally filtered, and bandpass filtered using a Krohn-Hite 3530 filter in order to obtain a flat frequency spectrum. It was amplified using a Hafler P9000 power amplifier, and projected from the Fender amplifier during the first phase of the experiment and the Community projector during the second phase.

Signals and noise were calibrated in the acoustic chamber using a Josephson Engineering C550H microphone and either a PC-based signal analysis package (Spectra Plus, Pioneer Hill) or a combination of virtual instruments. Mapping of the sound field in 10 cm^2 grids surrounding the stations ensured that spatial variability of the acoustic stimuli was less than ±2 dB in the vicinity of the subject's head.

D. General procedure

TTS was measured as a function of noise SL, SPL, SEL, and duration. The experimental design called for holding the frequency constant over experimental sessions while varying the level and duration of the fatiguing noise. Measurement of TTS was accomplished by assessing the subject's hearing sensitivity to the test tones before, immediately after, and at least 24 h following exposure to the octave-band noise. Control sessions, which comprised threshold testing associated with mock noise exposure, were also conducted. Multiple replicates of each exposure and control condition were run to ensure that statistically reliable results would be obtained. There were two phases of the experiment, and each phase followed an exposure matrix which defined the levels, durations, and number of replicates to be used. The sequence of testing was pseudorandomized within each phase of testing.

The subject's performance in all stages of testing was voluntary, with behavioral control established by operant conditioning and food reinforcement. During testing the subject was free to terminate participation in the experiment at any time by moving to the door of the acoustic chamber. When this occurred, the experiment was stopped and she was allowed to return to her pool enclosure. This happened rarely and primarily during initial training. There was no relation between the level of the noise and self-termination of sessions.

1. Preexposure threshold determination

A modified staircase method and a go/no-go procedure were used to estimate absolute hearing thresholds. Based on previous audiometric data, a test signal with a level of approximately 25 dB SL served as the stimulus for the first signal trial. A trial began with the subject positioned at the threshold station. The light situated under the tone projection speaker was turned on by the experimenter to initiate a trial. Following a brief, randomized delay of 1–4 s, a pure-tone signal was projected. The subject responded to the presence of the signal by touching the left-mounted response paddle with her nose (HIT). When this occurred, the bridge stimulus was projected and she received a fish delivered through the PVC tube by an assistant in the control room. The ratio of signal trials to no-signal trials (catch trials) was 1:1. Withholding response on a catch trial (CORRECT REJECTION) was reinforced in the same way as a HIT (bridge followed by food reinforcement). There were two types of incorrect response: failing to touch the paddle on a signal trial (MISS) and touching the paddle on a catch trial (FALSE ALARM). The sea lion received no special feedback following either type of incorrect response, except that the trial light was turned off. In these situations the subject normally restationed and waited for the next trial to begin.

On the first signal trial of a session, the maximum tone level (25 dB SL) was presented. Following each HIT, the level was dropped by 4 dB. This process continued until a MISS occurred. Following each MISS the level of the signal was raised by 2 dB. The 2 dB level change was used on all subsequent signal presentations (lowered following each HIT and raised following each MISS). Following nine reversals (transitions from HIT to MISS or vice versa), the signal level was raised to the starting level for several cooldown trials designed to maintain stimulus control over the response behavior. Thresholds were determined following the method of Dixon and Mood (1948) after the conclusion of the session. Data from sessions in which the false alarm proportion exceeded 0.25 were discarded. After the hearing threshold was determined, the subject was given a 20–30 min break prior to beginning the noise exposure.

2. Noise exposure

Noise exposure began when the subject was cued to enter the acoustic chamber and position at the exposure station. The exposure noise was turned on prior to the subject entering the chamber. An octave band of Gaussian white noise was projected from the noise-projecting speakers with a level and duration chosen from a predetermined exposure matrix. Approximately once every 30 s during noise exposure the

TABLE I. Noise exposure matrices for Phase 1 and Phase 2. The noise levels used are expressed in dB SL and dB SPL. The number of sessions conducted for the 16 level/duration combinations are provided, yielding a total of 112 exposure sequences for Phase 1 and 108 exposure sequences for Phase 2. The sound exposure levels resulting from each level/duration combination are shown in the cells.

Phase 1						
Noise exposure			Sound exposure level			
dB SL	dB SPL	*n*	1.5 min	12 min	25 min	50 min
Control	···	4	···	···	···	···
65	94	8	113	122	125	128
80	109	8	128	137	140	143
95	124	8	143	152	155	158
Phase 2						
Noise exposure			Sound exposure level			
dB SL	dB SPL	*n*	6.25 min	12.5 min	25 min	50 min
Control	···	3	···	···	···	···
98	127	8	152	155	158	161
101	130	8	155	158	161	164
104	133	8	158	161	164	167

light was turned off and a fish was delivered to the subject. At these times, the subject's position within the noise field changed only briefly when she moved from the station in order to pick up her fish. The time spent out of the calibrated noise field was small (on the order of several seconds) compared to the overall time of exposure; therefore, only energetically negligible differences in actual versus estimated exposure levels occurred. When the full duration of exposure was achieved, the light and noise at the exposure station were turned off. The sea lion was rewarded with a piece of fish and cued to position at the threshold station where postexposure testing immediately began.

3. Postexposure threshold testing

Following cessation of the noise exposure interval, the sea lion's hearing threshold to the 2.5 kHz test tone was assessed again. Actual determination of the threshold took place between 10 and 15 min following cessation of noise exposure. TTS was measured as the difference in decibels between postexposure and preexposure thresholds. The subject's hearing was tested again the following morning in order to assess whether sensitivity had fully recovered. In all cases where the initial postexposure threshold was more than 20 dB higher than the corresponding preexposure threshold, the subject was retested later the same day. In cases where the retest threshold was still elevated by 3 or more dB from the preexposure threshold, the subject was retested on subsequent days. Testing continued without intervening noise exposures until the threshold was within ±3 dB of the baseline threshold.

E. Exposure conditions

The exposure matrices used in Phase 1 and Phase 2 of the experiment are shown in Table I, which provides the SPL and the SEL for each condition tested. Each exposure matrix was a 4×4 design, with four exposure levels (three noise levels plus the control condition) and four exposure durations. Eight replicates of each of the noise exposure cells and three to four replicates of each of the control exposure cells were completed in a pseudorandomized order within each matrix.

The noise exposure levels in each matrix were established relative to the subject's baseline auditory threshold at 2.5 kHz. The subject's mean threshold at 2.5 kHz was 29 dB SPL prior to the start of the experiment, and all exposures levels were selected relative to this threshold. Therefore, an exposure level of 95 dB SL corresponded to an absolute exposure level of 29+95, or 124 dB SPL.

Phase 1 of the experiment took place between July 2002 and May 2003. In this phase, the exposure levels were 65, 80, and 95 dB SL. The exposure durations corresponding to each of these levels were 1.5, 12, 25, and 50 min.

Phase 2 of the experiment took place between April 2005 and October 2005. The exposure levels were 98, 101, and 104 dB SL and the exposure durations were 6.25, 12.5, 25 and 50 min. These exposure levels were incremented in 3 dB steps and the durations were incremented by doubling. This matrix was designed so that sound levels and durations increased in a systematic manner, both to induce larger shifts, and to allow for further comparison of equal SEL conditions comprising different combinations of exposure SPL and duration.

F. Analysis

Threshold shifts were calculated by subtracting preexposure thresholds from postexposure thresholds. Mean threshold shifts paired by sequence were compared across preexposure, postexposure, and 24 h recovery conditions using repeated-measures ANOVA, followed by a Student-Neuman-Keuls procedure for individual comparisons in the event that the ANOVA results were significant at the 0.05 level. Mean thresholds obtained using various combinations of duration and SPL resulting in equal sound exposure were also compared, using a Student's t-test or one-way ANOVA. Given

TABLE II. The mean TTS values obtained for each cell of the noise exposure matrices used in Phases 1 and 2. Statistically significant threshold shifts are indicated as $^{*}(p<0.05)$, $^{**}(p<0.01)$, and $^{***}(p<0.001)$. Note that none of the control (no-noise) conditions generated significant TTS.

Phase 1						
Noise exposure			Mean TTS			
dB SL	dB SPL	*n*	1.5 min	12 min	25 min	50 min
Control	···	4	0.7	1.1	0.7	−1.4
65	94	8	0.0	2.2	4.7^{**}	1.0
80	109	8	−0.5	4.5^{**}	5.3^{***}	7.2^{***}
95	124	8	1.5	3.8^{**}	6.1^{***}	11.0^{***}
Regression slope			0.05	0.06	0.05	0.34^{***}
Phase 2						
Noise exposure			Mean TTS			
dB SL	dB SPL	*n*	6.25 min	12.5 min	25 min	50 min
Control	···	3	0.5	−0.1	1.4	−0.7
98	127	8	1.7^{*}	4.3^{***}	7.3^{***}	10.5^{***}
101	130	8	3.3^{**}	5.7^{***}	7.8^{***}	16.0^{***}
104	133	8	4.6^{***}	7.6^{***}	12.3^{***}	23.5^{***}
Regression slope			0.50^{**}	0.55^{**}	0.84^{*}	2.12^{***}

that the difference in SEL between the 12 and 12.5 min exposures used in Phases I and II is negligible, the data from these two conditions were combined. Threshold shifts across equal sensation levels in air (this study) and under water (Kastak *et al.*, 2005) were also compared using a Student's t-test. A two-way ANOVA was used to test the effects of noise level and duration on the magnitude of TTS, as well as to test for interaction between the two factors.

Where SELs were plotted against threshold shifts, the following equation was used to fit to the data:

$$\mathrm{TTS} = (10m1)\log_{10}(1+10^{(\mathrm{SEL}-m2/10)})$$

(Kastak *et al.*, 2005). This is a modified form of the equation used to fit asymptotic threshold shift used by Maslen (1981). The parameters of the equation refer to TTS onset ($m2$) and growth of TTS with increasing SEL ($m1$). The latter parameter corresponds to the slope of a straight line fitted to the linear portion of the curve while the former corresponds to the x intercept (threshold shift of zero) of the same line. This equation is descriptive, and is not necessarily explanatory of the relationship between TTS and SEL.

III. RESULTS

Threshold shifts obtained for each exposure condition are shown in Table II. Significant differences between preexposure and postexposure thresholds are marked with asterisks. There were no differences between mean preexposure and 24 h recovery thresholds under any combination of level and duration (with the exception of six threshold values that reflected longer-duration recovery, obtained at intervals greater than 24 h). In Phase I, threshold shifts were not observed at durations of 1.5 min regardless of noise SPL. The only significant linear trend toward increasing threshold shifts with increasing stimulus level occurred at an exposure duration of 50 min, and the slope was extremely small (0.3 dB/dB noise SPL). In Phase II, significant threshold shifts were detected at all levels tested. Mean shifts ranged from 1.7 dB at 98 dB SL/6.25 min to 23.4 dB at 104 dB SL/50 min. All relations between threshold shift and SPL were significant, with small (less than 1) but positive slopes under all durations except the 50 min condition, in which the slope was 2.2 dB TTS/dB noise SPL.

There were no threshold shifts on any control sequence, indicating that no factor other than noise was responsible for the temporary loss of hearing sensitivity observed in the present study. No trends in preexposure thresholds were observed within each phase; however, mean baseline thresholds increased slightly (by 1.5 dB) between the phases (approximately one year). This difference was significant at the 0.5 level, implying some hearing loss between the two phases. However, an audiogram obtained from this subject several months after completion of the experiment showed no significant threshold shift at the test frequency relative to the mean of the Phase 1 preexposure thresholds. These results are difficult to interpret, although the audiogram did show some loss in sensitivity at frequencies above 6.4 kHz. Whether this increase is noise induced or age related cannot be determined; however, the phenomenon of high-frequency hearing loss in the absence of controlled noise exposure has been observed in another individual of the same species (Schusterman *et al.*, 2002).

Results of the 25 and 50 min exposures at 95 dB SL, when directly compared to results obtained under water under the same exposure conditions, revealed no significant differences in mean threshold shifts between the two media, as shown in Fig. 1. An additional comparison between underwater exposures at 80 dB SL and 22 min duration and in-air exposures at the same sensation level at 25 min also indicated no significant differences between the two media.

Comparisons of threshold shifts among equal-energy (SEL) conditions comprising various combinations of level and duration showed a trend toward increasing threshold

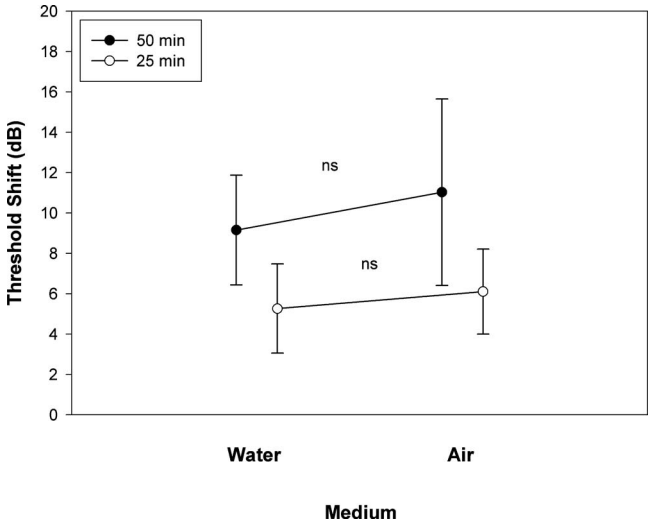


FIG. 1. Comparison of mean threshold shifts induced by equivalent SL exposures (95 dB SL) in air and under water for two duration conditions. There were no significant differences attributable to the differences in media.

shifts with increases in duration (Table III). This effect was most profound at higher sound exposure levels. In equal energy comparisons, longer duration exposures always resulted in greater mean threshold shifts than louder, shorter durations of the same overall sound energy.

Figure 2 shows a response surface fitted to a three-dimensional plot of noise duration and SPL versus threshold

TABLE III. Temporary threshold shifts obtained at combinations of noise SL/duration. Equal sound exposure level conditions are grouped.

Sound exposure level (SEL)	Level/duration combination (dB SL/min) resulting in column 1 SEL	TTS (note increasing TTS with increasing duration)
128	80 dB SL/1.5 min	–0.5
	65 dB SL/50 min	1.0
143	95 dB SL/1.5 min	1.5
	80 dB SL/50 min	7.2
152	98 dB SL/6.25 min	1.7
	95 dB SL/12 min	3.8
155	101 dB SL/6.25 min	3.3
	98 dB SL/12.5 min	4.3
	95 dB SL/25 min	6.1
158	104 dB SL/6.25 min	4.6
	101 dB SL/12.5 min	5.7
	98 dB SL/25 min	7.3
	95 dB SL/50 min	11
161	104 dB SL/12.5 min	4.6
	101 dB SL/25 min	7.8
	98 dB SL/50 min	10.5
164	104 dB SL/25 min	12.3
	101 dB SL/50 min	16

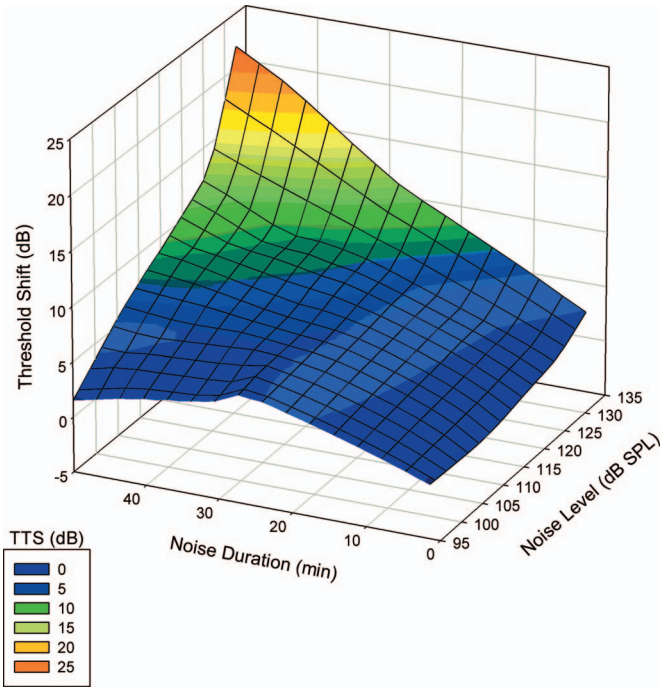


FIG. 2. A surface plot fitted to the mean threshold shifts under all conditions of noise level and duration. There is a significant interaction between noise SPL and duration for both phases of the experiment.

shift, with the data points removed for clarity. The effect of increasing duration can be seen to be largest at higher sound pressure levels (i.e., the duration effect depends on the level of noise), reinforcing the effect shown in Table III. Threshold shifts increased slowly with increasing duration at the lowest noise exposure levels, and increased rapidly with duration at the highest noise exposure levels used. The interaction between sound pressure level and duration was statistically significant for both Phase I ($F_{6,84}=5.28$, $p<0.001$) and for Phase II ($F_{6,84}=3.35$, $p<0.01$).

Figure 3 shows threshold shifts plotted against SEL. There was a significant increasing trend as shown by the experimental model, with greater SELs resulting in greater threshold shifts ($m1=2.5$; $m2=159$; $F_{(1,191)}=224$; $p<0.0001$; $R^2=0.54$). The resultant level of TTS onset was thus determined to be 159 re $(20\ \mu\text{Pa})^2$ s, while the growth of TTS at SELs above this level was 2.5 dB per dB increase in SEL. Caution must be used in the interpretation of this trend, as a relation between SEL and TTS assumes that the equal-energy rule applies, an assumption which is probably not strictly correct, as shown by the equal SEL comparisons.

In six exposure sequences, threshold shifts exceeded 20 dB. In these cases, repeated threshold estimates were obtained at postexposure intervals ranging from 1 to 48 h. Recovery functions for these sequences are shown in Fig. 4. All thresholds eventually returned to acceptable baseline levels (26–32 dB SPL), and all recovery functions were similarly shaped. A regression of threshold shifts versus the logarithm of time showed a highly significant linear relation, with threshold shifts decreasing at a rate of 8.8 dB per log (min).

IV. DISCUSSION

The experiment described here represents the most robust and conclusive data set on TTS in a marine mammal to

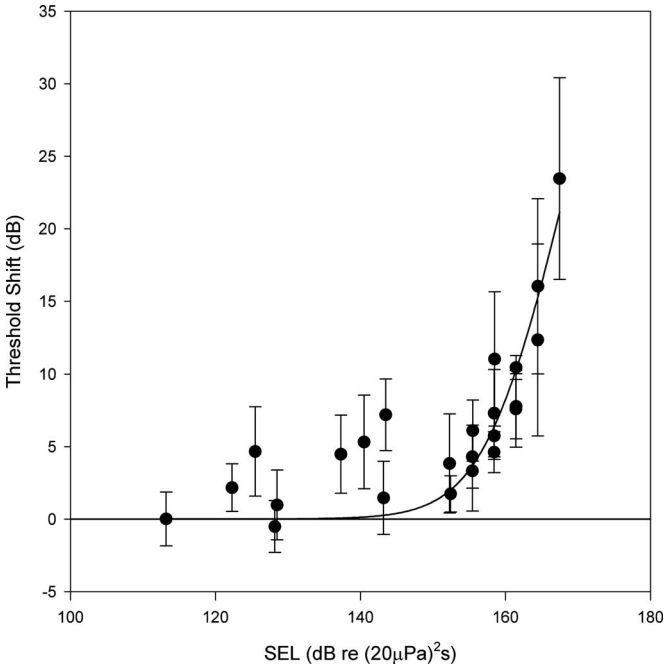


FIG. 3. A plot of TTS in dB vs SEL in dB re (20 μPa)^2 s, showing an exponential increase in magnitude of TTS with increasing exposure levels. The symbols represent mean shifts and the error bars represent standard deviations.

date. Onset, growth, and long-term recovery from TTS were statistically characterized with the benefit of the large data set that could have only been obtained under testing situations similar to those outlined here. Previous studies (Kastak *et al.*, 1999, 2005) have suffered from a limitation in the amount of data that could be collected, primarily because testing occurred in water. Variability of results in the previous studies meant that either a linear model or a nonlinear model with fixed parameters had to be used to fit the data. A linear model presented an unrealistically low TTS onset as well as a shallow slope that most likely underestimated the growth of TTS considerably (Kastak *et al.*, 1999). A curvilinear model based on a small data set with relatively small TTS values required at least one of the model parameters (slope) to be fixed, reducing confidence in the model (Kastak *et al.*, 2005). In the present experiment, both model parameters (slope and onset) were permitted to vary without bounds, resulting in a highly significant fit of the experimental model. This was likely due to the large number of data points as well as the wide range of TTS values included. Overall, the data presented here provide a clearer description of TTS in this species than has previously been the case.

It is important to note, however, that this study has limitations. First, the noise exposure was chosen with very specific characteristics (center frequency and bandwidth), as was the audiometric test stimulus. It is possible that different results would have been obtained had the experimental parameters varied along these dimensions. Second, only one subject was used, a deficiency that is, unfortunately, common in experimental work with marine mammals. However, these factors should not minimize the importance of the results; the subject was repeatedly tested to generate a robust assessment of onset, growth, and recovery of aerial TTS.

Growth of TTS, estimated here as 2.5 dB/dB increase in noise level, is a particularly important piece of data in the assessment of noise-induced hearing loss. It is difficult to quantify, however, and therefore hearing loss at nontested SELs is difficult to predict. At low to moderate SELs, a linear fit best describes the data; however, based on such an analysis, the growth of TTS is unrealistically small (only a fraction of a decibel increase in TTS per decibel increase in noise level). Further, linear models indicate a progressive threshold improvement as noise levels decrease, sometimes intercepting the abscissa at levels that would lead to unreasonably low noise levels corresponding to TTS onset. Rather, a model describing TTS growth should be curvilinear and have a lower asymptote at 0 dB TTS. The data we obtained are better fit by an exponential equation, the parameters of which correspond to TTS onset and growth at moderate to high levels of noise (Maslen, 1981; Kastak *et al.*, 2005). For this subject in air, TTS onset was established at 159 dB SEL [dB re (20 μPa)^2 s]. Growth of TTS was determined to be slightly above 2 dB per dB increase in exposure level. Both of these figures can be reasonably applied to noise exposure in California sea lions in air, with the following caution: Sound exposures that are different enough to be considered qualitatively different (impulse versus continuous, noise versus tone, etc.) are not likely to obey simple rules for predicting noise impacts.

A potential pitfall in using a SEL approach to quantifying TTS was made evident in an analysis of the 18 equal energy exposure conditions presented in this study. In every

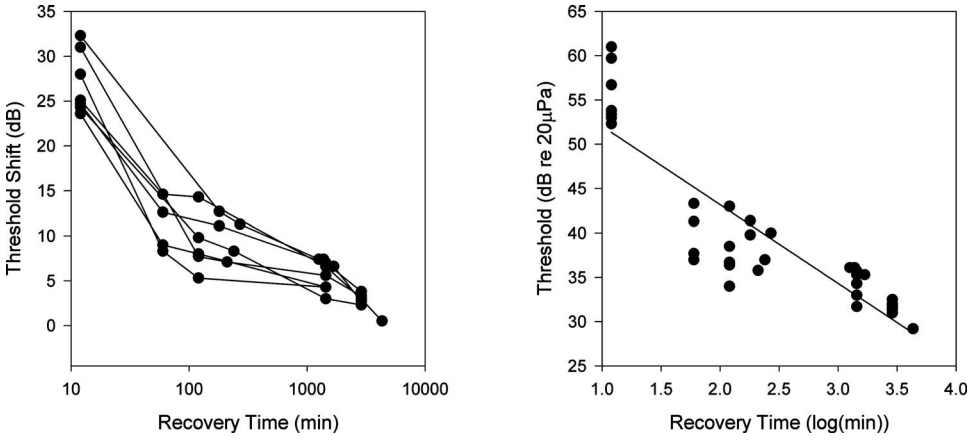


FIG. 4. Recovery from TTS greater than 20 dB plotted against the logarithm of time. All shifts eventually returned to near zero. Each line represents an individual test-recovery sequence. (b) A linear regression applied to the pooled data. The slope of the line is significant, $p < 0.001$, with a slope of −8.8 dB per log(min). The R^2 value is 0.54.

sequence comprising equal energy exposures, the longer, quieter exposure resulted in a greater threshold shift than the shorter louder exposure, which was an unexpected result (Ward, 1962). These results cannot be overlooked when making predictions of the sort that rely on an energy-based relationship. It is possible that peculiar anatomical properties, balances between simultaneous hearing loss and recovery, or differing physiological effects of noise play a significant role in governing the amount of damage a particular noise exposure may cause. Because the results of equal SEL comparisons shown in Table III appear to contradict the equal energy trading rule, a more productive way to examine the TTS data presented here may be to examine the relationship among SPL, duration of exposure, and the magnitude of TTS. The interaction between SPL and duration shown in Fig. 2 confirms that neither factor can be used alone to predict the magnitude of TTS. However, such a graph shown may provide an extremely powerful tool for predicting the potentially detrimental effects of noise on the hearing of a particular species of interest. For the California sea lion in this study, an exposure of 130 dB SPL for a duration of 20 min might be expected to induce a moderate threshold shift on the order of 5 dB, while an exposure of 120 dB lasting 40 min might induce a more severe shift of 10–15 dB. This example demonstrates the disproportionate effect of duration over SPL on the magnitude of TTS induced by noise. It further illustrates the usefulness of such a graph in predicting anthropogenic effects on hearing in free-ranging marine mammals, given that received levels are known or can be estimated. The shapes of these response surfaces will likely vary with sound type and frequency range, as well as species of interest, but the approach remains valid nonetheless.

Noise exposure levels used in this study and in previous research on pinniped TTS (Kastak *et al.*, 1999, 2005) were chosen by referencing them to the subject's threshold at center frequency, expressed as dB SL (sensation level). The use of SL to normalize exposure levels for comparisons across individuals and species is unproven and results have been equivocal. This approach is likely to be valid across the subject's area of best sensitivity—obtaining equal loudness curves for these species would be a good first step in determining noise effects outside areas of best sensitivity. Subjects with higher baseline hearing levels have been shown to require proportionally higher noise exposures to induce threshold shifts in other mammals (Humes and Jesteadt, 1991; Boettcher, 2002), and an approach based on SL was successful when harbor seal and sea lion results were compared (Kastak *et al.*, 2005); however, results obtained from a northern elephant seal were contradictory.

Sensation level may also be a useful tool for comparing the effect of noise SEL between air and water. For example, TTS onset under water for this subject was about 206 dB re 1 μPa^2 s (Kastak *et al.*, 2005), which is equivalent to noise exposure at a level of 206 dB re 1 μPa lasting 1 s. This SPL is 131 dB greater than the subject's absolute underwater threshold (i.e., 131 dB SL). In air, TTS onset for the same stimuli was estimated as 159 dB re (20 μPa)^2 s, corresponding to noise exposure of 159 dB re 20 μPa lasting 1 s, or an onset level of 130 dB SL. Thus, comparisons of the present data with data previously obtained indicate that noise exposures of equal durations can be equated in terms of SL, irrespective of medium. Because of more efficient testing procedures in air, we believe that aerial data can be used as an alternative for predicting the effects of underwater exposures to bands of noise. Tests to determine whether the same relationships also apply for different types of noise or to different aquatic species remain to be conducted.

A final, unexpected benefit of this study was the assessment of long-term recovery from TTS. Typically, recovery from TTS takes place in two qualitatively (if not physiologically) separate phases (Ward *et al.*, 1959). The first is a short-term process during which the level of TTS "bounces" to its highest level at approximately 2 min postexposure then begins to drop (Spieth and Trittipoe, 1958). Evaluation of threshold shifts during this period requires rapid audiometric methodology. The second phase of recovery is longer term, with threshold shifts generally declining in proportion to the logarithm of time (Ward *et al.*, 1959). Unfortunately, the limitations of this study resulted in postexposure threshold estimates that occurred at an average of 12 min following exposure. Whether the 2 min bounce occurs for TTS in pinnipeds is a matter reserved for future studies, but it is likely that many of the threshold shifts obtained in this study are underestimates of the maximum shift occurring at the 2 min postexposure point. It is clear, however, based on trial-to-trial data, that recovery from TTS occurs rapidly in marine mammals (on the order of minutes) for low and moderate levels of noise exposure (Finneran *et al.*, 2005; Kastak *et al.*, 2005; Nachtigall *et al.*, 2004), with TTS levels likely approaching zero within 1 or 2 h.

For threshold shifts greater than about 20 dB, the time course of TTS recovery was much longer, with a maximum of over 48 h elapsing prior to complete recovery. Recovery was proportional to the logarithm of recovery duration. Because recovery from TTS may take a considerable amount of time, it is of great importance for regulatory applications of TTS data to consider that some effects of anthropogenic noise may persist significantly beyond the duration of the noise itself. Based on the rates of recovery reported in this study, recovery from TTS appears to depend on the degree of threshold shift rather than solely on the absolute levels of noise exposure. Therefore, when considering the effects of noise on marine mammals, it is important to note that not only may a given noise exposure result in hearing loss in some species but not others, but that recovery from hearing loss will also differ among species and even individuals. Differences in susceptibility will likely be based on factors such as age, sex, and prior exposure over time scales of months or years.

Although the interpretations we make may not represent susceptibility to noise by California sea lions in general, it is clear that, for this individual, moderate and high levels and durations of broadband noise can cause auditory fatigue greater than the threshold shifts that have previously been reported for marine mammals. Based on the relation between sound exposure levels and TTS, the following may clarify what the experimental data mean in a practical sense. In

water, a passing ship exposing a pinniped to noise for a period of 10–20 min would need to produce received levels of between 170 and 175 dB re 1 μPa in order to begin to induce TTS. Similarly, in air, an aircraft flying over a sea lion rookery and exposing the animals to broadband noise for 30 s to 1 min would need to generate received levels of about 140–145 dB re 20 μPa in order to induce TTS. It cannot be emphasized enough that these are highly specific and simplified examples, ignoring other effects such as startle responses, masking, and differing rates of recovery. Further, noise related effects other than TTS, such as temporary loudness shift, can affect broader frequency ranges and take much longer to recover (Botte and Monikheim, 1994). The caveats regarding sample size, noise type, and exposure intermittence remain applicable; thus, additional research must examine the effects of different noise types on this and other species. However, predictions made by this and other studies do serve to indicate potential damaging effects to the California sea lion's auditory system by relatively intense, sustained, and/or chronic anthropogenic noise exposure.

ACKNOWLEDGMENTS

The authors would like to thank the student assistants and volunteers at Long Marine Laboratory for their dedicated and hard work resulting in the completion of this project. Financial support was provided by Grant No. N00014-04-0284 from the Office of Naval Research; we thank Mardi Hastings and Robert Gisiner at ONR for their continued support. Funding for the hemi-anechoic chamber and control equipment was provided by a Defense University Research Instrumentation Program (DURIP) grant. Two anonymous reviewers provided extremely thorough and valuable comments.

American National Standards Institute (ANSI). (**1994**). "Acoustical terminology," ANSI S1.1-1994, R2004.

Boettcher, F. A. (**2002**). "Susceptibility to acoustic trauma in young and aged gerbils," J. Acoust. Soc. Am. **112**, 2948–2955.

Botte, M. C., and Monikheim, S. (**1994**). "New data on the short-term effects of tone exposure," J. Acoust. Soc. Am. **95**, 2598–2605.

Dixon, W. J., and Mood, A.M. (**1948**). "A method for obtaining and analyzing sensitivity data," J. Am. Stat. Assoc. **43**, 109–126.

Fernandez, A., Edwards, J. F., Rodriguez, F., de los Monteros, A. E., Herraez, P., Castro, P., Jaber, J. R., Martin, V., and Arbelo, M. (**2005**). "'Gas and fat embolic syndrome' involving a mass stranding of beaked whales (Family: Ziphiidae) exposed to anthropogenic sonar signals," Vet. Pathol. **42**, 446–457.

Finneran, J., Dear, R., Carder, D., and Ridgway, S. (**2003**). "Auditory and behavioral responses of California sea lions (*Zalophus californianus*) to single underwater impulses from an arc-gap transducer," J. Acoust. Soc. Am. **114**, 1667–1677.

Finneran, J. J., Carder, D. A., Schlundt, C. E., and Ridgway, S. H. (**2005**). "Temporary threshold shift in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones," J. Acoust. Soc. Am. **118**, 2696–2705.

Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (**2000**). "Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and a beluga whale (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions," J. Acoust. Soc. Am. **108**, 417–431.

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002**). "Temporary shift in masked hearing thresholds in odontocetes after exposure to single underwater impulses from a seismic watergun," J. Acoust. Soc. Am. **111**, 2929–2940.

Frantzis, A. (**1998**). "Does acoustic testing strand whales?," Nature (London) **392**, 29.

Holt, M. M., Schusterman, R. J., Kastak, D., and Southall, B. L. (**2005**). "Localization of aerial puretones in pinnipeds," J. Acoust. Soc. Am. **118**, 3921–3926.

Humes, L. E., and Jesteadt, W. (**1991**). "Modeling the interactions between noise exposure and other variables," J. Acoust. Soc. Am. **90**, 182–188.

Kastak, D., and Schusterman, R. (**1996**). "Temporary threshold shift in a harbor seal (*Phoca vitulina*)," J. Acoust. Soc. Am. **100**, 1905–1908.

Kastak, D., Schusterman, R., Southall, B., and Reichmuth, C. (**1999**). "Underwater temporary threshold shift induced by octave-band noise in three species of pinniped," J. Acoust. Soc. Am. **106**, 1142–1148.

Kastak, D., and Schusterman, R. J. (**1998**). "Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology," J. Acoust. Soc. Am. **103**, 2216–2228.

Kastak, D., Southall, B., Schusterman, R., and Reichmuth-Kastak, C. (**2005**). "Underwater temporary threshold shift in pinnipeds: Effects of noise level and duration," J. Acoust. Soc. Am. **118**, 3154–3163.

Maslen, K. (**1981**). "Towards a better understanding of temporary threshold shift of hearing," Appl. Acoust. **14**, 281–318.

Nachtigall, P., Supin, A., Pawloski, J., and Au, W. (**2004**). "Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials," Marine Mammal Sci. **20**, 673–687.

Nachtigall, P. E., Pawloski, J. L., and Au, W. W. (**2003**). "Temporary threshold shifts and recovery following noise exposure in the Atlantic bottlenosed dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **113**, 3425–3429.

Schusterman, R. J., Southall, B. L., Kastak, D., and Reichmuth Kastak, C. (**2002**). "Age-related hearing loss in sea lions and their scientists," J. Acoust. Soc. Am. **111**, 2342–2343.

Southall, B. L., Schusterman, R. J., and Kastak, D. (**2003**). "Auditory masking in three pinnipeds: Aerial critical ratios and direct critical bandwidth measurements," J. Acoust. Soc. Am. **114**, 1660–1666.

Southall, B. L., Schusterman, R. J., Kastak, D., and Reichmuth Kastak, C. (**2005**). "Reliability of underwater hearing thresholds in pinnipeds," ARLO **6**, 243–249.

Spieth, W., and Trittipoe, W. J. (**1958**). "Intensity and duration of noise exposure and temporary threshold shifts," J. Acoust. Soc. Am. **30**, 710–713.

Ward, W. D. (**1962**). "Damage-risk criteria for line spectra," J. Acoust. Soc. Am. **34**, 1610–1619.

Ward, W. D., Glorig, A., and Sklar, D. L. (**1959**). "Relation between recovery from temporary threshold shift and duration of exposure," J. Acoust. Soc. Am. **31**, 600–602.

Session 1aAA

Architectural Acoustics: Sound Systems in Large Rooms and Stadia

David S. Woolworth, Cochair
Oxford Acoustics Inc., 356 CR 102, Oxford, MS 38655

Timothy W. Leishman, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

Invited Papers

8:45

1aAA1. Acoustical and sound system design considerations in large venues and stadia. Jack Wrightson, Ron Baker, and Kevin Day (WJHW, Inc., 4801 Spring Valley Rd, #113, Dallas, TX 75244, rbaker@wjhw.com)

Large, spectator seating facilities provide unique challenges for the sound and acoustical system designer. The challenges are equally daunting, but different for indoor and outdoor facilities. This presentation provides empirical data of existing facilities in regard to speech intelligibility, uniformity of sound reinforcement coverage, maximum loudness, etc., along with a review of predicted modeling and auralization techniques. Design guidelines for acoustical treatment, loudspeaker configuration, predictive calculation of speech intelligibility, and other aspects will be discussed. Difference or variations in design practice between more conventionally sized spaces and very large environments will be presented.

9:10

1aAA2. The implications of scale on sound systems in large spaces. Bob McCarthy (Alignment and Design Inc., 204 Falling Leaves Ct., Creve Coeur, MO, 63141, bob.mccarthy@charter.net)

The symphonic concert hall is a well documented reference point for matters of scale in the musical side of architectural acoustics. Larger spaces may have comparable shapes, and yet the matters of scale require vastly different approaches in regard to the sound system and architectural acoustics. The causes and consequences of scalar expansion will be discussed as well as the shifting balance between the roles of the sound system and the architectural acoustics. Sound system and acoustical designs must adapt to those aspects which are modified by scale (power, distance, wavelength, etc.) along with those that remain constant, regardless of the size (angular aspects, echo perception, the spacing between and size of audience members, etc.). Expanded room scale requires the sound system to use increasingly effective methods of controlling the direct sound field to tailor its response to the highly asymmetric shapes presented for coverage. The methods of control and their implication on the hall acoustics will be discussed. Scalar independent methods for sound system designs which can be form-fitted into a space will be discussed in the context of a variety of large room examples.

9:35

1aAA3. Stadium acoustics-design challenges and solutions. David Marsh (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247, David.Marsh@pmkconsultants.com)

Large stadiums, whether enclosed or open-air, present many acoustical design challenges including excessive reverberation, echoes from surfaces distant from the sound system loudspeakers, sound absorption by air, refraction, difficulty achieving acceptable speech intelligibility, and synchronization of the sound reinforcement system to the video boards. This paper is an overview of how these challenges have been met in several professional sports facilities. Reverberation times for these large venues are normally underpredicted on the order of -20% by conventional methods. Low frequency reverberation times (down to 63 Hz) tend to be extremely long—about 15 s on average. A novel approach will be described to estimate these values (within a wide range of possibilities) based on *in situ* measurement data. It will be shown that after exhausting all possibilities of acoustical treatments, the most effective way of achieving acceptable speech intelligibility is to use a distributed sound system with no seat being more than about 80 ft from the nearest loudspeaker.

10:00–10:20 Break

10:20

1aAA4. Purpose built sound systems for large rooms. Neil A. Shaw (Menlo Sci. Acoust., Inc., P.O. Box 1610, Topanga, CA 90290-1610, menlo@ieee.org) and John Monitto (Meyer Sound Labs., Inc., Berkeley, CA 94702)

Sound systems for sports facilities have improved over the years with the development of better loudspeaker systems with increased headroom and reduced distortion at high sound levels. Today's loudspeaker systems can produce cleaner sound with less interference between the devices, when installed either temporarily or permanently in these often very reverberant venues. Two basic design philosophies for football, baseball, and soccer stadiums are the point source and the distributed system. Smaller basketball,

hockey and other sports typically have sound systems flanking scoreboards in a point source design with the option of distributed delayed speakers. In some new facilities that are designed for multipurpose use, the use of electronic processing systems enables these rooms to be designed with lower reverberation times, which is typically the preferred acoustic environment for live amplified concerts. For games and other events a variable acoustic system provides additional reverberation as needed, also enhancing natural crowd and court sounds, such as a ''puck'' sound, and increases the intimacy of applause and cheering between the fans and the game itself.

10:45

1aAA5. Signal-to-noise ratios in sound system design. Marshall Long (Marshall Long Acoust., 13636 Riverside Dr., Sherman Oaks, CA 91423, mlacoustics@sbcglobal.net)

An important consideration in sound system design is intelligibility, and the most common measure of intelligibility is some sort of signal-to-noise ratio. Intelligibility can be expressed in terms of a goodness ratio (high signal-to-noise) or a badness ratio (high noise-to-signal), or a modified badness ratio (noise x reverberation time)/signal such as %ALcons. The trick is how to calculate signal, how to calculate noise, and how to design systems with what we calculate. Three types of solutions: Cabinets (horns), distributed loudspeakers, and line arrays are discussed.

11:10

1aAA6. Live sound measurements in stadia. Wolfgang Ahnert, Stefan Feistel, Alexandra Radu Miron, and Enno Finder (Ahnert Feistel Media Group, Arkonastr. 45–49, D-13189 Berlin, Germany, wahnert@ada-acousticdesign.de)

In this presentation, the authors introduce the software-based measuring system EASERA SysTune to be used for measurements by excitation with music or speech signals. It investigates the use of standard signals supplied from a sound system in a stadium in real-time. Using a newly developed program module, live-sound recordings or speech and music signals from a microphone input and from the mixing console can be utilized to obtain impulse response data for further evaluation. New noise suppression methods are presented that allow these impulse responses to be acquired in full-length, even in occupied venues. As case studies, acoustic measurements based on live sound supply are discussed for a soccer stadium. Required measuring conditions and limitations are derived as a result.

TUESDAY MORNING, 27 NOVEMBER 2007 GRAND COUTEAU, 8:00 TO 10:05 A.M.

Session 1aAO

Acoustical Oceanography: Marine Sediment Properties and Inversions

Jon M. Collis, Chair
Woods Hole Oceanographic Inst., Bigelow Bldg, Woods Hole, MA 02543

Chair's Introduction—8:00

Contributed Papers

8:05

1aAO1. The effect of grain shape on the porosity of marine sediments. David R. Barclay and Michael J. Buckingham (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238)

The porosity of marine sediments increases as the mean grain diameter decreases. In contrast, the close-packing structures (including random packing) of uniform-size spheres all show a porosity that is independent of sphere size. Hamilton suggested that, in sediments, the shape of the grains is instrumental in determining the porosity. To investigate this idea, experiments have been performed on a variety of sand samples (marine sediments, beach sands, and desert sands) with mean radii within the range 30 mm–200 mm. The macroscopic properties (porosity and bulk density) are measured in simple tabletop experiments, while the microscopic details of grain shape are quantified using an optical microscope and image-analysis software. For each grain in, typically, a sample of several hundred, the boundary is traced and various geometric parameters are returned, including the perimeter, the cross-sectional area, the centroid, and the best-fit ellipse. Also, a Fourier decomposition yields the radius as a function of angle, and the higher-order coefficients are then combined to provide a description of grain shape (i.e., deviation from sphericity). An attempt is being made to interpret the observed porosities in terms of the measured grain shapes. [Research supported by ONR.]

8:20

1aAO2. Shear waves and the discrepancy between perceived and ideal frequency power laws for sediment attenuation. Jon M. Collis, Allan D. Pierce, and William M. Carey (Boston Univ., Boston, MA 02215, jcollis@bu.edu)

Inverse techniques based on data for long-range propagation in shallow water have recently inferred that the attenuation in certain marine sediments varies at low frequency as the 1.8 power. Idealized models predict the exponent to be exactly 2.0. The inverse inferences usually assume the bottom is a fluid, and this is ordinarily a good approximation, because the shear wave speed in bottom sediments is typically very small. Direct numerical simulation [J. M. Collis et al., Proc. Oceans 2007, Aberdeen (2007)] indicates that shear waves make a sufficient contribution to shallow water attenuation that could account for the small discrepancy in exponents. To better assess whether this is the case, the present paper analyzes the effect of shear waves on modal attenuation. The Pekeris model with a lower elastic half-space is used with the shear wave speed taken to be substantially less than the sound speed in water. The derived dispersion relation has complex roots for the horizontal wave number, and the imaginary part, found by a perturbation analysis, predicts that the shear wave contribution to the modal attenuation is proportional to the cube of the ratio of the shear wave to water sound wave speeds. [Work supported by ONR.]

8:35

1aAO3. Complex-density, equivalent-fluid modeling of acoustic interaction with the seafloor. Michael Vera (Univ. of Southern Mississippi, 118 College Dr., #5046, Hattiesburg, MS 39406, michael.vera@usm.edu)

Acoustic interaction with the seafloor can generate both compressional and elastic shear waves in the solid. Accurate models of shear propagation are often computationally expensive and difficult to apply to long-range propagation. When the sound field in the water is of primary interest, an equivalent-fluid model of the seafloor, with parameters chosen to match the reflection coefficient of the actual elastic solid, can sufficiently characterize the effect of the bottom on energy in the water. The effective density of the seafloor material in this approach can be a complex number. Prior methods for generating equivalent fluids were intended for low shear speeds and low grazing angles. Recent developments in the technique were intended to extend its validity to higher shear speeds and a wider range of angles. These efforts were initially motivated by the need to simulate bottom-interacting arrivals for the broadband Kauai source in the North Pacific Acoustic Laboratory experiment at megameter ranges. The work to be presented involves a more detailed examination of the performance of the method, including comparisons to benchmark models and to shorter range data from this Kauai source collected as part of the Basin Acoustic Seamount Scattering Experiment.

8:50

1aAO4. A direct inversion scheme to obtain sediment sound speed and density in shallow water. Kyle M. Becker (Penn State Univ./Appl. Res. Lab., P.O. Box 30, State College, PA 16804-0030)

A simple inversion scheme is described for estimating the sound speed and density in the seabed for Pekeris-type waveguides. At a fixed frequency, input data to the inversion algorithm are discrete values of the phase of the plane wave reflection coefficient at the water/bottom interface. Above the critical angle, the input data are derived from estimates of the horizontal wave numbers for propagating modes in the wave guide. For two or more propagating modes, a linear system of equations is solved directly to determine sound speed and density in the bottom. The approach is strictly only valid for determining properties of the lower half-space in a Pekeris Wave guide. However, in certain circumstances, the approach can be applied to determine properties of the topmost sediment in layered media. The method is demonstrated on a synthetic data set and then applied to data collected at sea. [Work supported by an ONR Ocean Acoustics Entry-Level Faculty Award.]

9:05

1aAO5. Analysis of the effect of water column sound speed variation on geoacoustic inversion. Yong-Min Jiang and N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, PO Box 3055, Victoria, BC V8W 3P6, Canada)

Bayesian matched field inversion has been applied on multi-tonal data sets acquired on the New Jersey continental shelf in the SW06 experiment in August, 2006. Since the data sets were collected in a range-dependent environment due to water column sound speed variation, sound speed profile (SSP) was decomposed in terms of empirical orthogonal functions and also inverted in the inversion. This presentation examines the sensitivity of the geoacoustic and geometrical parameters to the shape and the gradient of the sound speed in the thermocline by the interparameter correlation from the inversion. The effects of the SSP on the geoacoustic inversion and source localization are studied through the inversions of measured data at different ranges with/without prior information about geometric parameters in the inversion. Further investigations on the variation of SSP to the sediment sound speed, density, and attenuation at different ranges are carried out by simulations. It is found that the geometric parameters are more sensitive to the SSP than are the geoacoustic parameters in this shallow water environment. [Work supported by ONR.]

9:20

1aAO6. Geoacoustic information content of bottom-moored horizontal line array data. Dag Tollefsen (Norwegian Defence Res. Establishment (FFI), Box 115, NO3191 Horten, Norway) and Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria BC, Canada V8W 3P6)

This paper considers geoacoustic information content in matched-field inversion of acoustic data from bottom-moored horizontal line arrays. Bayesian geoacoustic inversion is applied to low-frequency multi-tone data in a synthetic shallow water environment under realistic signal-to-noise ratios and for frequency bands relevant to both controlled-source experiments and ship-noise inversion. The information content is quantified in terms of widths of marginal a posteriori probability densities of model parameters obtained by fast Gibbs sampling. The information content dependence on factors such as array length and sensor configuration, source range and bearing (from endfire to broadside), and frequency content is studied.

9:35

1aAO7. Bayesian geoacoustic inversion of ship-noise data from a bottom-moored horizontal line array. Dag Tollefsen (Norwegian Defence Res. Establishment (FFI), Box 115, NO3191 Horten, Norway) and Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria BC, Canada V8W 3P6)

This paper presents results from matched-field geoacoustic inversion of low-frequency (40–150 Hz) ship-noise data recorded on a horizontal array deployed at the seafloor. Data are taken from an experiment conducted by FFI in shallow waters of the Barents Sea. Estimates of sediment parameters were obtained by inversions of narrowband data from a relatively quiet surface ship at source-array ranges of 1–6 km. The fast Gibbs sampler was used to estimate posterior probability densities of model parameters. Estimates of sound speed and density of quaternary sediment are compared with results from inversion of data from a towed source collected along the same track, and with reference values from other geophysical data collected in the area.

9:50

1aAO8. Improved perturbative inversion schemes for obtaining bottom geoacoustic properties in shallow water. Megan S. Ballard and Kyle M. Becker (Penn State Univ., State College, PA 16802, msd200@psu.edu)

Perturbative schemes using modal wave numbers are used to obtain sound speed in the sediment as a function of depth [S. D. Rajan, et. al., J. Acoust. Soc. Am. **82**(3), 998–1017 (1987)]. The inversion algorithm involves solving an ill-posed problem, with regularization and singular-value decomposition used to stabilize the solution, resulting in a smoothed version of the true sound speed profile. However, owing to geological processes, sediments are often better described by layers having distinct properties and are not well represented by a smooth profile. In this work, two new methods are applied to stabilize the solution of the inverse problem, making it possible to resolve discontinuities in the sound speed profile of the sediment. Piecewise polynomial truncated singular value decomposition (PP-TSVD) is used in place of traditional stabilizing methods and is demonstrated to yield better results when a layered model is a reasonable assumption. The results can be further improved using qualitative regularization if prior knowledge of the location of sound speed discontinuities is available. These methods are shown to yield very accurate estimates of the sound speed profile deep into the sediment using very few perturbations to the forward model. [Work supported by NDSEG and ONR.]

Session 1aUW

Underwater Acoustics: Propagation Modeling

Josette Paquin Fabre, Chair
Acoustics Division, Naval Research Lab., Stennis Space Center, MS 39529

Chair's Introduction—10:25

Contributed Papers

10:30

1aUW1. Integrated signal excess as a metric for environmental acoustic assessment. J. Paquin Fabre (Naval Res. Lab., 1005 Balch Blvd., Stennis Space Ctr., MS 39529)

Signal excess (SE) is often used as a metric for determining acoustic performance through a waveguide or over an area. This quantity can limit the capability to assess the environment because SE is computed given a single acoustic source, receiver and frequency, and is a function of range from the source, thereby limiting the scenarios for which the environment can be assessed as well as the ability to visualize it over an area. Integrated SE (ISE) with phase tracking is proposed as an improved metric for evaluation of acoustic performance. For a given source location (in latitude and longitude), the SE is integrated over all possible source depths, a band of frequencies, and bands of ranges. Additionally, the phase variations across the regions are tracked. This metric is compared to various traditional SE fields and is shown to provide a better representation of the overall acoustic properties of a waveguide and an area. The ISE and the original SE are also examined across an ensemble of sound speeds and the variations of the acoustic environment are characterized. [The author appreciates and acknowledges the funding support from the Naval Research Laboratory Base Program.]

10:45

1aUW2. Modeling surface duct precursors for Littoral Acoustic Demonstration Center 2003 (LADC03) seismic calibration experiment. Arslan M. Tashmukhambetov, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, 2000 Lakeshore Dr., New Orleans, LA 70148, atashmuk@uno.edu), and Natalia A. Sidorovskaia (UL Lafayette, Lafayette, LA 70504-4210)

In summer 2003, LADC conducted a seismic calibration experiment for a 3190-in3 21-element seismic exploration array. After calibration, absolute broadband (up to 25 kHz) pressure-time dependencies for a wide range of offsets and arrival angles were produced. A computational workflow, combining the upgraded version of the Navy standard acoustic propagation model RAM and industry airgun modeling packages, was developed to model the calibrated experimental data. Experimental and modeled data have demonstrated good agreement in absolute pressure amplitudes and large-scale frequency interference pattern for the frequencies up to 1,000 Hz. The results of the application of the developed workflow for investigation of the effect of a seasonally-formed surface duct on seismic array energy distribution is presented. The presence of the surface duct is experimentally confirmed to be responsible for the generation of a unique set of modes forming the precursor arrivals, first systematically observed from a point source on a near-bottom hydrophone by DeFerrari and Monjo. The influence of a range-dependent environment on precursor stability is addressed. An attempt to identify precursors in the LADC03 experimental data is discussed. [Research sponsored by the Industry Research Funding Coalition and the Joint Industry Project.]

11:00

1aUW3. Efficient modeling of range-dependent seismo-acoustics problems. Michael D. Collins, Woo-Yeol Jung (Naval Res. Lab., Washington, DC 20375), Elizabeth T. Kusel (Northeastern Univ., Boston, MA 02115), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

The parabolic equation method is the most effective approach for solving range-dependent problems in ocean acoustics. During the past three decades, many difficulties have been encountered while attempting to generalize this approach to problems involving elastic layers. With the development of an improved single-scattering solution [J. Acoust. Soc. Am. **121**, 808813 (2007)], a large class of range-dependent seismo-acoustics problems can now be solved accurately and efficiently. Previous single-scattering solutions for seismic problems were based on an iteration formula, and limited by convergence problems. Since the improved solution does not require iteration, it provides the efficiency of the energy-conserving solution that has proved to be very useful for acoustics problems. The application of this approach to problems involving fluid layers is currently being investigated. [Work supported by ONR.]

11:15

1aUW4. Mode formulas for shallow water waveguides using a modified asymptotic approximation. Stephen V. Kaczkowski, William L. Siegmann (Dept. of Mathematical Sci., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180-3590, kaczks@rpi.edu), Allan D. Pierce, and William M. Carey (Boston Univ., Boston, MA 02215)

Sound speed profiles in shallow water waveguides are small deviations from isospeed and often decrease monotonically with depth. Approximation formulas for the propagating modes are found using a modified form of the classical WKBJ method. The ocean bottom is taken as homogeneous. The approach is accurate for modes with phase speeds greater than, and even slightly less than, the maximum water sound speed. The validity and accuracy of the approximations over frequency and mode number are illustrated using benchmark numerical calculations. Comparisons with approximations from other methods, including previous WKBJ approaches, are described. The formulas here are typically more compact and convenient. The results demonstrate how changes in parameters, such as frequency, bottom sound speed, and water sound speed profile quantitatively affect the mode functions. Applications for representative waveguide profiles are provided. [Work supported by the Office of Naval Research.]

11:30

1aUW5. Modeling underwater sound propagation from an airgun array using the parabolic equation method. Alexander MacGillivray (JASCO Res. Ltd., 4464 Markham St., Victoria, BC, Canada)

A technique for modeling sound propagation from an airgun array using the parabolic equation (PE) method is presented that takes into full account the far-field, angle-dependent directionality pattern of the array. This is achieved by generating a PE starting field for the array by summing together shaded, phase-shifted replicas of the PE self-starter. The

array starter has been implemented using the RAM parabolic equation code. Validation comparisons are presented of field predictions generated using the array starter against exact normal mode solutions for an array source computed using ORCA. Examples of synthetic waveform airgun array calculations are also presented. The array starter method can be used to accurately predict pressure waveforms from an airgun array in the ocean environment provided that the modeler knows (or can compute) ''notional'' far-field source signatures for individual airguns in the array.

11:45

1aUW6. Model for energy consumption by a broadband shallow-water acoustic communications network. Daniel Rouseff and Warren L. J. Fox (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105, rouseff@apl.washington.edu)

A multistatic active sonar system with several widely distributed sensors must share information between the sensors. Acoustic communication is a viable method to share information and is particularly appropriate when the sensors are mobile. A potential drawback is that a significant fraction of the total energy budget for the system might be consumed just by sending information between the sensors. In the present work, an energy consumption model is developed that is appropriate for incoherent acoustic communication in shallow water. The model builds upon the foundation established by Sozer et al. [IEEE J. Oceanic Eng. 25, 72–83 (2000)], but uses an improved sub-model for acoustic propagation. The sub-model includes environmental factors like the sound speed profile in the water column and the composition of the seabed. By exploiting the waveguide invariant concept, the broadband nature of the communications signals can be included efficiently in the calculation. Numerical results demonstrate how relaying messages between intermediate sensors can save substantial energy compared to direct communications. The calculations also show that energy consumption can vary by more than an order of magnitude depending on the seabed composition. [Work supported by ONR.]

12:00

1aUW7. Relating X-waves and Cerenkov radiation. Shane C. Walker (Univ. of California, San Diego, Scripps Inst. of Oceanog., Marine Phys. Lab., 9500 Gilman Dr., Mail code 0238 La Jolla, CA 92093-0238)

X-waves are shown to be related to Cerenkov-Vavilov (CV) radiation as derived by solving the linearized inhomogeneous scalar wave equation in free space for the case of a source in uniform motion at velocity **v** that is greater than the speed of waves, c, in the medium, $|v|>c$. The results apply to wave phenomena in general, both acoustic and electromagnetic. In this context, the supersonic/superluminal properties of X-waves are unambiguously demonstrated to be phase effects. One implication is that even the ideal diffractionless X-wave is not signal-like nor can it exceed the Shannon data rate bandwidth relation, $\Delta t \Delta\omega = 2\pi$. Further, it is shown that the continuous power required to propagate an X-wave is analogous to the energy dissipation of a CV particle caused by radiation drag. Thus, generating a diffractionless X-wave requires an infinite total energy input over a spatially infinite source aperture. Finally, the connection between CV radiation and X-waves suggests the possibility of generating X-waves using time-reversal.

TUESDAY AFTERNOON, 27 NOVEMBER 2007 GRAND BALLROOM E, 1:00 TO 5:05 P.M.

Session 1pAA

Architectural Acoustics, Musical Acoustics, Signal Processing in Acoustics, and Noise: Even Better Than the Real Thing—Rock, Pop, and All That Jazz

Alexander U. Case, Cochair
Fermata Audio and Acoustics, P.O. Box 1161, Portsmouth, NH 03802-1161

K. Anthony Hoover, Cochair
Cavanaugh Tocci Associates, Inc., 327F Boston Post Rd., Sudbury, MA 01776

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Psychoacoustic evaluation of music reproduction in passenger cars. Hugo Fastl, Martin Ammler (AG Technische Akustik, MMK, TU Muenchen, Arcisstr. 21, 80333 Muenchen, Germany, fastl@mmk.ei.tum.de), and Adam Sulowski (Audi AG, Ingolstadt, Germany)

A method to assess the quality of music reproduction in passenger cars is developed and evaluated. In essence, the subject sits in a standing car and is presented from a CD music of different genres via the car's audio system. The subject has to rate different aspects of the performance on a 10 step scale using a PDA. From the ratings, a net like a spider net, with different attributes like natural, squeeking, dull, reverberant, acceptable, good etc. on its rays is constructed. The goal is a net with a large area, i.e. as many as possible of the attributes should attain the maximum rating. Examples from a pilot study are given for car audio systems of different sophistication, and aspects like gender differences or possible biases towards a product will be discussed.

1:25

1pAA2. Changing the paradigm of practice rooms and music teachers coaching studios. Ron Freiheit (Wenger Corp., 555 Park Dr., Owatonna, MN 55060)

For years the image of practice rooms for many musicians meant playing in places that at times resembled (and sometimes were, in fact) closets. They served the function by creating individual places to practice, but were usually less than desirable due to the internal acoustics provided. The same paradigm has held true for most music teachers coaching studios. Time spent with the teacher was invaluable for improving one's craft, however the acoustics were constrained the rooms physical dimensions, construction materials and room furnishings. With the advent of DSP, these spaces can now provide a variety of acoustic environments for practice and coaching. Integrated into the latest DSP technology is the ability to provide concurrent recording capability, enabling immediate review of the musician's performance. In the coaching studio, it is especially effective when the student can quickly hear a specific point the teacher is making. New technology accomplishes this more quickly than with past recording methods. The net result of these additional tools is to accelerate the potential learning experience for the musician.

1:45

1pAA3. Fictional timbre—Close microphones and warped perspectives. Alexander U. Case (Sound Recording Technol., Univ. of Massachusetts, Lowell, 35 Wilder St., Lowell, MA 01854)

Recording engineers, particularly in pop and rock genres, have a long tradition of placing microphones exceptionally close to the sound sources they wish to record. Not a gesture towards realism, close microphone techniques are born of practical needs for sound isolation in sometimes cramped recording spaces, production pressures to defer signal processing decisions until the mixdown session, and creative desires to variously avoid, enhance, and/or exaggerate sonic features of the instrument in search of a sound that is ''better than the real thing.''

2:05

1pAA4. Psychoacoustics of non-real spaces. Sam Ortallono (MediaTech Inst., 3324 Walnut Bend Ln., Houston, TX 77042)

Psychoacoustics of non-real spaces, modern recording, and mixing techniques provide the opportunity to create acoustic phenomena that do not exist in the real world. This presentation will discuss how recorded music has been manipulated and try to understand what people think they hear and why. Through a combination of recent original recordings and classic records, investigations can be made into how processed sound can disorient, surprise, or please the listener. In the studio, compressors, gates, delays, and artificial reverbs are employed to create unnatural soundscapes, yet the mind still applies the rules that have been learned in everyday life. Such impossibilities as backward sounds and multiple sources will also be considered. Finally, some speculation can be made as to why these records seem to sound even better than the real thing.

2:25

1pAA5. What to do with a five-track woodwind quintet. James W. Beauchamp and Mert Bay (School of Music and Dept. of Elect. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, jwbeauch@uiuc.edu)

Recently we made a five-track recording of a woodwind quintet using a two-step process: First, the group was recorded together in the same room with each instrument close-miked. Second, each instrument was recorded separately with the performer monitoring the others on headphones. There were a few synchronization problems, but basically the method worked fine, and the result is a recording with complete separation between the tracks. Although this was inspired by the necessity for ground truth for a demixing project, several other advantages of this approach to recording have come to mind: (1) Listening to all instruments played back simultaneously from separate speakers is very enjoyable. (2) Music-minus-one recordings are useful for musicians to practice with and are easy to produce. (3) Data reduction on each track enables analysis, modification, and resynthesis projects of various sorts. For example, pitch, tempo, relative loudness, and timbre are easy to manipulate. Examples of music produced by processing the woodwind quintet will be demonstrated.

2:45

1pAA6. Making outdoor acoustic venues "better than the real thing." Pamela Harght (School of Architecture and Urban Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

Traditionally-acoustic music organizations, particularly orchestras, are developing new ways to reach out to a wider audience, extend their season, and promote their music in non-traditional summer homes. Many of these outdoor venues accommodate far more patrons than the established, enclosed concert hall. Furthermore, with these larger venues comes the responsibility to enhance the performing group through amplification to satisfy the expectations of the patrons and to compensate for the vastly different atmosphere regular concert-goers are accustomed to. Electronic reinforcement is now required to accommodate the thousands more outdoors and with digital reverberation and modern recording equipment, these goals and expectations can be met with little compromise in the listening experience. This paper will discuss the role of sound reinforcement in orchestral music, focusing on the popular Tanglewood Music Center, the summer home of The Boston Symphony Orchestra, and other successful venues.

3:05–3:15 Break

3:15

1pAA7. Pulse and swing: Quantitative analysis of hierarchical structure in swing rhythm. Ken Lindsay (tlafx, 180 Ohio St., Ashland, OR 97520, samba4ken@gmail.com) and Pete Nordquist (Southern Oregon Univ., Ashland, OR 97520)

Swing feeling in music is generally attributed to the presence of timing variations in the rhythm as played, contrasted with standard tablature form. Substitution of triplet subdivision for the divide by two format (quarter, eighth, sixteenth notes) of Mozart-Bach (MB) notation is a well known timing variation associated with Swing. We present evidence that an important part of Swing is also found in nonuniform timing of the Pulse, or basic beat, of a Swing tune i.e., time delta between downbeats of successive musical measures does not conform to a regularly spaced time grid as defined by the metronome. To our knowledge, this fundamental aspect of Swing has not been previously reported. The departure from a uniform time grid and presence of different types of time variations played by different instruments both make important contributions to the ''lively'' quality found in Swing style of many genres: Classic Jazz Swing, Pop, Gospel/Soul, Rock and Roll, Samba, etc. Although different rhythms may not be locked together as in MB style, nonetheless they are all highly synchronized. The degree of synchronization corresponds to varying degrees of tight or loose rhythmic style. This also contributes to the ''live'' feeling in the musical performance.

3:35

1pAA8. Reality is not a recording/a recording is not reality. Jim Anderson (New York Univ., 194 Mercer St., Rm. 501, New York, NY 10012)

The former New York Times film critic, Vincent Canby, wrote all of us have different thresholds at which we suspend disbelief, and then gladly follow fictions to conclusions that we find logical. Any recording is a fiction, a falsity, even in its most pure form. It is the responsibility, if not the duty, of the recording engineer, and producer, to create a universe so compelling and transparent that the listener isnt aware of any manipulation. Using basic recording techniques, and standard manipulation of audio, a recording is made, giving the listener an experience that is not merely logical but better than reality. How does this occur? What techniques can be applied? How does an engineer create a convincing loudspeaker illusion that a listener will perceive as a plausible reality? Recordings will be played.

3:55

1pAA9. Audio networks—From the main stage to main street. Deb Britton (K2 Audio, LLC 4900 Pearl East Cir., Ste. 201E, Boulder, CO 80301)

In recent years, digital audio networks have become a more and more prevalent means of distributing audio signals. Audio networks are not unique to large, permanently-installed sound systems, but are also used in touring sound, homes, and even some musical instruments. The audio network trend was not sparked by a desire to be on the cutting edge of technology, but to fill some very real needs. Audio technology has been steadily moving from analog to digital, yet the interconnection of the individual components remains in the analog domain. The resultant numerous A/D and D/A conversions can cause signal quality issues, severely limiting the performance of the system. A network provides a common, digital interconnection, and reduces the infrastructure requirements by routing many signals over fewer cables. For the facility owner, the greatest benefit could be the drastic reduction in cabling. For the engineer or system end user, any input to any output routing capability enables the implementation of different system configurations on the fly, and can allow for future expansion, without the need for hardware or wiring changes. For the musician or home studio enthusiast, a networked instrument provides a direct interface to a computer for recording.

4:15

1pAA10. A survey of small live music venues of New Orleans. David Woolworth (Oxford Acoust., Inc., 356 CR 102, Oxford, MS 38655, dave@oxfordacoustics.com)

This survey of small live music venues of New Orleans will summarize aspects of history, capacity, sound reinforcement specifications, and acoustic properties. There will also be information on the factors affecting sound quality for small venues and some examples of ''on the fly'' problem solving, using available materials and sound processing.

Contributed Papers

4:35

1pAA11. Acoustic investigations of concert halls for rock music. Niels W. Adelman-Larsen (Flex Acoust., Scion-DTU, Diplomvej 377, 2800 Lyngby, Denmark, nwl@flexac.com), Eric R. Thompson, and Anders C. Gade (Tech. Univ. of Denmark, 2800 Lyngby, Denmark)

Objective measurement data and subjective evaluations have been collected from 20 small-/medium-sized halls in Denmark used for amplified rhythmic music concerts (pop, rock, jazz). The purpose of the study was to obtain knowledge about optimum acoustic conditions for this type of hall. The study is motivated by the fact that most concert tickets sold in Denmark relate to concerts within these genres in this kind of venue. The subjective evaluations were carried out by professional musicians and sound engineers who responded on the basis of their experiences working in these (and other) halls. From the relationships between the subjective evaluations and the objective data, recommended values for reverberation time across frequency are suggested. For halls with volume between 50.000 ft^3 and 160.000 ft^3, the optimal range is 0.7 s to 1.0 s without audience (and chairs). T30 should be frequency independent, so that clarity remains high also at the lower frequencies. A lower limit for T30 was identified. The paper also proposes possible hall designs that seek to fulfill the demands of sound engineers and musicians, as found in the investigation.

4:50

1pAA12. Compliance with city noise regulations by small venue restaurants, lounges, and sports bars. Marlund E. Hale (Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065, mehale@aol.com)

There are many small venues that offer recorded and/or live musical entertainment to their regular customers as well as to attract new customers to their establishments. Many municipalities have enacted noise regulations requiring that such commercial entertainment meet exterior community noise limits and sometimes even indoor noise standards. This paper presents field performance test results for several small businesses that failed to meet the regulatory noise limits and were in jeopardy of being fined or closed down until they could comply with the regulatory noise requirements. The businesses all provided musical entertainment, and were located on busy streets in commercial buildings. All of the businesses also had a customer base that expected loud music. Most of the businesses were afraid that reducing the volume of their musical entertainment would cause them to lose customers and money. Several solutions were explored and recommendations made that satisfied both sides of the issue.

TUESDAY AFTERNOON, 27 NOVEMBER 2007 NAPOLEON C3, 1:55 TO 4:30 P.M.

Session 1pAB

Animal Bioacoustics: Acoustics of Bats, Birds, Insects, and Rodents (Lecture/Poster Session)

Mardi C. Hastings, Chair
Pennsylvania State Univ., Applied Research Lab., State College, PA 16804

Chair's Introduction—1:55

Contributed Papers

2:00

1pAB1. Link-based feature extraction for model-based detection of echolocation calls. Mark D. Skowronski and M. Brock Fenton (Biol., Univ. of Western Ontario, London ON N6A 5B7, Canada, mskowro2@uwo.ca)

Frame-based acoustical analysis of echolocation calls from bats has recently been proposed to improve the robustness of automated call detection. Statistical models, popularized and honed by automated human speech processing research, learn the distributions of call features extracted identically from short overlapping frames of time. A fundamental assumption is that each frame contains at most one echolocation call. In field recordings of bats, this assumption is often violated by the presence of echoes and overlapping calls from multiple bats, which degrade automated detection performance. Simple rules, based on expert knowledge of echolocation calls, were used to follow local spectrogram ridges and to group ridge points into links. Frame-based features were extracted from the links, which may overlap in time, allowing links to be used in the existing model-based analysis paradigm. Improved performance using links is demonstrated for several pathological cases, including conspecific interaction and high-duty-cycle foraging.

2:15

1pAB2. Simulating bat echolocation. Fons De Mey, Jonas Reijniers, and Herbert Peremans (Universiteit Antwerpen, Labotheek, Prinsstraat 13, 2000 Antwerpen, Belgium, fons.demey@ua.ac.be)

In accordance with the growing practice of performing acoustical experiments in virtual reality, the acoustic properties of the echolocation system of a number of bat species are characterized in simulation. The head related transfer function (HRTF) and interaural intensity differences (IIDs) from the pinnae and the radiation pattern from the noseleaves are calculated with a boundary-element method. The obtained results are compared to actual acoustic measurements performed on other specimens of the same species. This comparison shows that performing the experiments in virtual reality is a viable alternative to performing measurements on real specimens. It has the additional advantage that the spatial resolution that can be achieved compared to the more commonly used acoustical measurements, is higher. Furthermore, computer manipulation of the virtual morphology model allows characterizing the influence of different head parts on the HRTF and the noseleaf radiation pattern. To illustrate the power of this technique, the effects of different head parts on the HRTF of the phyllostomid bat *Phyllostomus discolor* are investigated. For this species, it is concluded that the pinna has a larger effect on the interaural intensity difference (IID) pattern than the head itself.

2:30

1pAB3. Frequency-swept directivity lobes—An emerging functional principle of biosonar beamforming. Rolf Müller, Hongwang Lu, Zhiwei Zhang (School of Phys. & Microelectronics, Shandong Univ., Hongjia Lou 5, 250100 Jinan, China, mueller@sdu.edu.cn), and John R. Buck (Univ. of Massachusetts Dartmouth, North Dartmouth, MA 02747)

In all bat species, the spatial sensitivity of the biosonar system is mainly determined by the shape of the outer ears. In addition, a noseleaf surrounds the nostrils in many species and shapes the spatial distribution of the emitted sound energy. Predicting the beam-forming effects of these structures numerically from their surface geometry has enabled comparative studies of high-resolution beam pattern estimates across a large set of different bat species. The beam patterns obtained so far differ from common manmade transducers, such as loudspeakers and microphones. Strong asymmetric side lobes, which change direction as a function of frequency, are a distinguishing feature consistently seen in emission and reception beam patterns of the bats. Where investigated, the occurrence of these side lobes appears to be tied to the presence of flap-like structures (including the tragus of the pinna), which may be hypothesized to have evolved for this purpose. The presence of these side lobes suggests that bat biosonar has not been evolved to maximize the ensonification energy and sensitivity for a single small target unconditionally. Instead, the function of these strong side lobes could be explained as maximizing the information obtained in analogy to squinted beams used in manmade systems.

2:45

1pAB4. Sound recordings from the 2007 emergence of Brood XIII cicada. Ralph T. Muehleisen (Civil, Architectural, and Environ. Engr., Illinois Inst. of Technol., 3201 S. Dearborn St., Rm. 228, Chicago, IL 60616, muehleisen@iit.edu)

Level calibrated recordings of the Brood XIII 17-year periodic cicadas were made in June, 2007 in River Forest, Illinois. Brood XIII consists of three separate species: *Magicicada cassini*, *M. septendecim*, and the more rare *M. septendecula*. Power spectrum analysis of a 2 min segment of the recording shows a 5.95 kHz resonance with $Q \approx 6$, indicative of *cassini*, and a 5 dB lower resonance at 1.35 kHz with a $Q \approx 13.5$, indicative of *septendecim*. Since other researchers have estimated that the power of *septendecim* is 20 dB below that of *cassini*, the high level of the *septendecim* song suggests that 30 times more *septendecim* than *cassini* were singing. Computation of the sound pressure level shows a 2–5 dB variation with a period of 7.5 s, indicative of synchronized chorus singing of *cassini*.

3:00–3:15 Break

3:15

1pAB5. Possible vocalizations of the ivory-billed woodpecker (*Campephilus principalis*). Michael D. Collins (Dept. of Mathematical Sci., Rensselaer Polytechnic Inst., Troy, New York 12180), Richard S. Martin (529 Durham Dr., Homewood, AL 35209), and Dalcio K. Dacol (19149 Hempstone Ave., Poolesville, Maryland 20837)

The Ivory-billed Woodpecker (Campephilus principalis) was feared to be extinct for decades, but there have been sightings in several states during the past few years. During a study at the last known nest sites in the 1930s, the characteristic kent calls of this species were recorded and other calls were observed but not recorded. There have been recent observations and recordings of high-pitched calls that are suspected to be ivorybill vocalizations. In the Pearl River, the calls were heard mixed in with kents during an encounter with two ivorybills. At the same site two days later, the calls were heard and recorded after an ivorybill was flushed, and the movements of the source of the calls was correlated with the movements of the ivorybill (video was also obtained during the encounter). Similar calls were recorded along with kent-like calls in the Choctawhatchee River at a site where there were sightings the same day and the previous day. The calls were also heard in the Congaree Swamp after a recording of kents was played back. The calls are similar to one of the calls of the Blue Jay (Cyanocitta cristata), but this conspicuous species was not observed during the encounters.

3:30

1pAB6. Design of LabVIEW-based software for insect acoustic research. Vijay Ramalingam, Tom Fink, John Seiner (The Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677, virjay@olemiss.edu), Douglas Streett (ARS-USDA, Stoneville, MS), and Alan Lax (ARS-USDA, New Orleans, LA)

LabVIEW-based software has been developed for acquiring and analyzing acoustical data from insects under field and laboratory conditions. The following features will be discussed in the presentation: Data acquisition using single or multiple transducers using National Instruments hardware. Timed, threshold, and periodic data acquisition routines. Periodical playback of WAV files or sine waves during data acquisition. Embedded META data information to avoid field notebooks. Real-time saving or deleting acquired data without losing acquisition continuity. Sequential, logical, and automatic file saving routines with time and date stamped file names. Playback of acquired data at different speeds ($0 \leqslant v \leqslant 1$), and at different time frames. Compare multiple waveforms (up to four) using synchronized pan and zoom routines. Create annotations, image file, and small segments of the original waveform with a single mouse click. Create spectrograms, FFT, and power spectrums of the acquired data. [Work supported by USDA ARS.]

3:45

1pAB7. A vowel identification procedure for gerbils. Joan M. Sinnott and Kelly W. Mosteller (Psych. Dept., Univ. of South Alabama, Mobile, AL 36688)

Mongolian gerbils are small rodents with auditory specializations that provide them with excellent hearing in the human speech range. Our lab has previously developed simple go/no-go response procedures to measure gerbil detection and discrimination of human vowels. Here is presented our new identification procedure that uses a two-alternative, forced-choice, left versus right response. The gerbil is tested in a small wire mesh cage that contains a center platform, a light, and two feeder cups left and right of the light. The gerbil begins a trial by mounting the platform when the light flashes. One of two vowel sounds (e.g. /u/ or /i/) then starts repeating at the rate of 1 per sec. The gerbil is reinforced with food if s/he goes to the left feeder cup for /u/, or to the right feeder cup for /i/. Incorrect responses are punished with timeouts. Gerbils easily learn to differentiate the spectrally-dissimilar vowels /u/ versus /i/, and data will be presented on their ability to differentiate more similar vowels such as /a/ versus /ae/. Gerbil identification ability appears related primarily to second formant frequency differences. A video will be shown of gerbils working on the procedure. [Research supported by NIH.]

All posters will be on display and all authors will be at their posters from 4:00 p.m. to 4:30 p.m.

1pAB8. The substrate vibration generating behavior of *Aphaenogaster carolinensis* (Hymenoptera: Formicidae). Timothy O. Menzel (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), Jake R. Marquess (Univ. of Mississippi, University, MS 38677), John Seiner (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), and Douglas Streett (USDA-ARS-BCMMUR, Stoneville, MS)

Aphaenogaster carolinensis Wheeler produces substrate vibrations by striking a substrate with its mandible and dragging its mandible across the surface. We have described this behavior and conducted laboratory trials to determine the conditions under which it occurs. Individual strikes were consistent in pattern and duration, but there was great variation in the duration of, and the number of strikes associated with, a single substrate vibration generating event. In laboratory trials, this behavior occurred most frequently in response to confrontation with non-nest mate conspecifics and did not require the presence of a food source or territory to be initiated.

1pAB9. Buzz digging and buzz plastering in the black-and-yellow mud dauber wasp, *Sceliphron caementarium* (Drury). Tom Fink, Vijay Ramalingam, John Seiner (NCPA, Univ. of Mississippi, One Coliseum Dr., University, MS 38677, tfink@olemiss.edu), Niels Skals (Univ. of Southern Denmark, DK-5230 Odense M, Denmark), and Douglas Streett (USDA-ARS, Stoneville, MS 38776)

The black-and-yellow mud dauber digs mud with its mandibles at pond or puddle edges, flies with a mud ball to a locality that is dark and protected from rain, and then immediately begins to form a mud tube for eventual provisioning with paralyzed spiders as food for its single developing larva (only one egg is depositied per tube). Digging and plastering (making a mud tube) are always accompanied by intense vibrations (intense enough to shake one's hand when holding the wasp with forceps) caused by the wing muscles. Visual and acoustic details of digging and plastering were studied by high speed video (usually 2,000 fps) correlated with acoustical data recorded with laser Doppler vibrometry, microphones, and accelerometers. High definition video was also used to view the general process of digging and plastering, but it is impossible to visualize the

vibrations except with high-speed video. The vibration of the wing muscles causes the head to vibrate at the same frequency, which mechanically aids digging and plastering. Pulling and pushing mud with the mandibles while digging are at different frequencies, and plastering is at an even lower frequency. The wasp can modulate the amplitude of these vibrations while digging or plastering.

TUESDAY AFTERNOON, 27 NOVEMBER 2007 GRAND COUTEAU, 1:35 TO 3:45 P.M.

Session 1pAO

Acoustical Oceanography: Storms and Intense Air-Sea Interactions

Jeffrey A. Nystuen, Chair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Chair's Introduction—1:35

Invited Papers

1:40

1pAO1. The origin of storm wave noise and its application to passive acoustic remote sensing of bubbles. Grant Deane and Helen Czerski (Scripps Inst. of Oceanogr., Code 0238, UCSD, La Jolla, CA 92093-0238)

Breaking waves in storms create large numbers of bubbles at the ocean surface. Direct measurement of the bubbles is difficult, and there are few data on the numbers, sizes, and creation rates of the large, transient bubbles formed in whitecaps. These bubbles could be measured using passive acoustic remote sensing if the excitation mechanism of sound production from newly created bubbles was understood. Using bubbles released from a nozzle as a model system, we have determined that sound production is driven by the collapse of an air neck formed immediately after bubble pinch-off. Laboratory experiments and model calculations of the mechanism will be presented. [Work supported by ONR and NSF.]

2:00

1pAO2. Acoustic surface-interaction phenomena during storms: A semi-empirical approach. Roger Gauss and Joseph Fialkowski (Naval Res. Lab., Code 7144, Washington, DC 20375-5350)

When storms arise at sea, winds transfer energy to the water surface, raising waves that grow higher and longer, and generating subsurface bubbles via breaking waves. This increased complexity of the undersea air-sea boundary-interaction zone makes it important to be able to isolate and characterize the contributions of the rough air-sea interface, bubbles, and fish on an impinging acoustic field to allow realistic assessments of their impact on sonar system performance. Active transmissions can be used to remotely sense this environment and help decipher the interaction process. In this paper, a coupled approach is presented that provides self-consistent oceanographic and acoustic models. Through a combination of oceanographic and acoustic measurements and modeling, semi-empirical oceanographic models of key phenomena (wave age, air-void fraction) are first developed that depend on readily-observable environmental parameters (wind speed and significant wave height). In turn, these models are used to develop broadband semi-empirical acoustic scattering and loss models that environmentally depend on these same parameters. The paper concludes with illustrations of how strong this dependence is during storm conditions, as well as the extent to which other scattering phenomena (fish) can be masked. [Work supported by ONR.]

Contributed Paper

2:20

1pAO3. Multibeam sonar observations of near-surface bubbles during a storm. Thomas C. Weber (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03284)

In October 2006, a reson SeaBat 7125 multibeam sonar (400 kHz) mounted on the drop keel of the R/V Hugh Sharp was used to collect acoustic backscatter from bubbles generated by breaking wind-waves in the Gulf of Maine. The multibeam sonar has a very high spatial resolution, with 256 1.0×0.5 deg beams. The data of interest here were collected as the ship returned to port because of bad weather (sustained 15 m/s wind speed), traveling through the plumes of breaking waves. Data were collected at a ping rate that varied between 5–10 pings/s. These observations show frequent breaking wave bubble plumes extending up to 12 m deep. The data also show that preferential concentrations of bubbles (clusters) were often present. Correlation functions (e.g., the pair correlation) describing the bubble clustering have been extracted from this data, and their affect on acoustic propagation characteristics (sound speed, attenuation) will be presented. [Work supported by ONR Ocean Acoustics.]

Invited Papers

2:35

1pAO4. Acoustic monitoring of severe weather in the northeast Pacific Ocean. Jeffrey A. Nystuen (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105, nystuen@apl.washington.edu) and Svein Vagle (Fisheries and Oceans Canada, Sidney, BC, Canada V8L 4B2)

Wind and rainfall are the principal physical processes responsible for the production of high frequency (1–50 kHz) ambient sound in the ocean. The primary source of the sound is the resonant ringing of individual bubbles created during wave breaking and raindrop splashes. Larger bubbles (?‘ 300 μ diam) quickly return to the surface, while smaller bubbles can be mixed downward many meters. During severe weather, a layer of smaller ambient bubbles forms and effectively absorbs higher frequency ($>$10 kHz) sound. These processes are revealed in a two-year record of ambient sound recorded from a sub-surface mooring at 50N, 145W in the northeast Pacific Ocean as part of the Canadian SOLAS program. The passive acoustic signal of wind, rain, and ambient bubble clouds are compared to the sub-surface mooring data including data from an upward looking 200 kHz active sonar and a 300 kHz ADCP. The acoustic signatures of light, moderate and heavy rainfall are superimposed on the signature of high wind, demonstrating rainfall detection even in the presence of high wind. [Work supported by ONR, Fisheries and Oceans Canada, and the Canadian CFCAS & NSERC.]

2:55

1pAO5. Quantifying the destructive power of hurricanes with naturally occurring undersea sound. Nicholas C. Makris and Joshua D. Wilson (Massachusetts Inst. of Technol., Cambridge, MA 02139, makris@mit.edu)

A passive ocean-acoustic method for quantifying the destructive power of a hurricane with accuracy similar to that of in situ aircraft measurements is presented. The method has been demonstrated by comparing underwater acoustic data with aircraft wind speed data. Both the acoustic and wind speed data were obtained in 1999, when hurricane Gert passed over an autonomous underwater hydrophone in the North Atlantic. The implications of this relatively safe and inexpensive ocean acoustic method for hurricane classification and disaster planning are discussed. A new experimental initiative between Mexico and the United States on ocean acoustic hurricane quantification will be described. [This work is supported by the Office of Naval Research and Seagrant.]

Contributed Papers

3:15

1pAO6. Comparison of wind driven ambient noise in shallow and deep-water ocean environments. David Knobles, Steven E. Cho, Sumedh M. Joshi (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713), Roy Gaul (BlueSea Cooporation, Houston, TX 77014), and Brian J. Sperry (Sci. Application Intl., Inc., McLean, VA)

Measured sound pressure time series in shallow and deep-water ocean environments are used to investigate the character of wind-generated noise over a range of wind speed of 2–40 knots and a frequency band of 10–3,000 Hz. In 2006, acoustic data were collected during tropical storm Ernesto off the New Jersey continental shelf in shallow (70 m) water. Ambient noise from these data is compared to deep-water measurements at several locations in the Northwest and the Northeast Pacific. The results also are compared to previously reported measurements by [Kuperman and Ferla (1985)], [Wenz (1962)], and others. The analysis includes a technique for separating background shipping noise from the wind-induced component of ambient noise. The system electronic noise floor and other extraneous sources of noise are evaluated to assure that the ambient noise measurements are not corrupted. Physical mechanisms examined include a wind-saturation effect that limits the ability to acoustically monitor distant storms and the effect of shallow water propagation on the character of wind noise. The identification of the distant shipping part of the spectrum allows for inferences to be made on the low frequency part of the ambient noise spectrum. [Work supported by ONR.]

3:30

1pAO7. High resolution acoustic characterization and change detection to evaluate hurricane effects on a coastal bay bottom. Angelina M. Freeman, Harry H. Roberts (Coastal Studies Inst., Dept. of Oceanic and Coastal Sci., Louisiana State Univ., Baton Rouge, LA 70803, afreem8@lsu.edu), and Patrick D. Banks (Louisiana Dept. of Wildlife and Fisheries, Baton Rouge, LA 70898)

Using high-resolution acoustics (sidescan sonar, chirp sonar subbottom, and echo sounder profile data) the shallow surface and subsurface of Sister Lake (Terrebonne Basin) of coastal Louisiana’s deltaic plain was imaged pre- and post-hurricanes Katrina and Rita. The pre-hurricanes survey in May 2005 imaged the entire lake bottom. To assess the effects of hurricanes Katrina and Rita, a second survey in October 2005 included acoustic resurveying of corridors of data covering all major sectors of the lake. Regions of sidescan reflectance differences between pre- and post-storm periods, corroborated with regions of onlapping sediment or zones of scour in the chirp sonar subbottom data, show areas both of sedimentation and erosion between surveys. Ground truth box cores were collected to calibrate acoustic data. Preliminary image processing results of the georeferenced pre- and post-storms sidescan datasets indicate 60% of the area experienced sedimentation and 40% was scoured between surveys. Amplitude reflectivity of chirp horizons were mapped, and incorporated with sidescan amplitude and texture attributes in change detection analysis. Standardized digital processing, supervised classification, and change detection techniques are developed to characterize the bay bottom and shallow subsurface and to assess and quantify storm effects. [Work supported by NOAA NA03NMF4520316.]

Session 1pPA

Physical Acoustics and Acoustical Oceanography: Acoustic Applications for Hurricane and Storm Preparedness

Preston S. Wilson, Chair
Univ. of Texas at Austin, Dept. of Mechanical Engineering, 1 University Station, Austin TX 78712-0292

Invited Papers

4:00

1pPA1. Monitoring hurricanes using infrasound. Carrick L. Talmadge, Claus Hetzer, Kenneth E. Gilbert, and Roger Waxler (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Oxford, MS 38677, clt@olemiss.edu)

The University of Mississippi is leading a multi-institutional collaboration to monitor hurricane activity using infrasound, in particular the microbarom signal produced by any strong cyclonic storm. Monitoring technologies currently in use do not permit accurate long-term predictions of hurricane intensity, nor in general even for the rapid warning of sudden intensification of the storm strength. The current status of the hurricane acoustic monitoring system, including advances in sensor technology at the NCPA, will be reviewed. The potential for microbaroms to be used as an indicator of storm strength and as a metric to aid in predicting future changes in storm strength will be discussed.

4:20

1pPA2. Underwater surveys after the storm. Peter J Alleman (C & C Technologies, Inc., 730 E. Kaliste Saloom Rd., Lafayette, LA 70508)

Following hurricanes Katrina and Rita, many types of acoustic systems were used to assess the effects of the storms. Swath multibeam sonars were used to survey erosion in ports, channels, and estuaries. Sidescan and scanning imaging sonars were used to locate debris, obstructions, missing oil platforms and ships, and dislocated pipelines. Acoustic positioning systems and acoustic modems were used to support the sonar imaging systems. A wide array of acoustic systems were used in the overall effort. Samples of data from several types of sonar systems will be presented along with discussions of the challenges of collecting data in a real-world environment.

Contributed Papers

4:40

1pPA3. Infrasound measurements during hurricane *Katrina*. Ronald Wagstaff, Eric Goggans, Heath Rice, and Carrick Talmadge (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, rwagstaf@olemiss.edu)

Infrasound is used to detect both natural and man-made events. Microbarometric measurement stations have been placed worldwide in order to detect infrasonic signals in the atmosphere. Such signals range in frequency from less than 0.01 to 10 Hz. Infrasound is produced by many natural phenomena including violent storms, gravity waves, the Aurora Borealis, and even some animals. Because infrasound signals have such low frequencies, they can travel great distances in the atmosphere, while maintaining good temporal coherence. This makes the atmosphere an excellent medium for infrasound signal propagation and corresponding detection of large storms, such as hurricanes and tornados, which generate strong infrasonic signals. This paper will discuss hurricane Katrina's infrasound spectral results obtained from data measured by piezo-ceramic infrasound sensors in Oxford, MS, about 400 miles to the north of New Orleans. Spectral and temporal coherence results will be presented and discussed. [Work supported by ARDEC.]

4:55

1pPA4. Generation of microbaroms by hurricanes. Claus Hetzer, Carrick L. Talmadge, Kenneth E. Gilbert, and Roger Waxler (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Coliseum Dr., University, MS 38677)

Under quiet conditions, a background peak at approximately 0.2 Hz is observed worldwide on infrasound stations. The peak is understood to be the ''microbarom peak'' associated with ocean waves. The intensity of microbaroms increases dramatically on inland infrasound stations when a hurricane is offshore and within a few thousand km. Hurricanes, thus, are unique phenomena for investigating microbarom generation by ocean waves. We review recent microbarom data and examine existing theories for the generation of microbaroms. In particular, we propose two possible models for the generation of microbaroms by hurricanes. In the first model, the microbaroms are generated near the eye of the hurricane, and in the second model, by counter propagating ocean wave fields created by the hurricane and relatively far from the eye. Bearing data from infrasound stations are presented that support the second model. [Supported by the U.S. Army Space and Missile Defense Command.]

Session 1pUWa

Underwater Acoustics: Geoacoustic Inversion

David P. Knobles, Chair
Environmental Sciences Dept., Applied Research Labs., Univ. of Texas at Austin, Austin, TX 78713

Chair's Introduction—1:00

Contributed Papers

1:05

1pUWa1. Analyses of broadband and narrowband propagation measurements in Shallow Water 2006 (SW06). David P. Knobles, Robert A. Koch, Preston S. Wilson (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713), Timothy F. Duda, and Arthur E. Newhall (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Acoustic measurements were made during the Shallow Water 2006 Experiment near the New Jersey continental shelf in 70–80 m of water. For a portion of the experiment, several L-arrays and a variety of broadband and narrowband sources were deployed for the purpose of investigating the nature of the dispersion of the acoustic propagation in the seabed. A geo-acoustic inversion methodology is used to infer the dispersion from the acoustic measurements in the water column. The waveguide possesses significant inhomogenities in range, azimuth, and time due to bathymetry, sound speed profile fluctuations resulting from internal waves, and variability in seabed properties. Physical oceanography measurements are used in part to obtain the best estimate for a water column description for the analyses of specific acoustic measurements. Various inversion methods along with different propagation models, such as two-way coupled modes and finite element parabolic equation algorithms, are used to obtain an accurate mean description and uncertainty estimates for the properties that control dispersion of the seabed. [Work supported by ONR.]

1:20

1pUWa2. Is higher resolution bathymetry always better for undersea sound propagation modeling? Juan Arvelo (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723)

Bathymetry is one of the most important environmental parameters affecting undersea sound transmission in shallow waters. There has always been a preference towards bathymetry with the highest resolution and accuracy when conducting sound propagation modeling. However, extreme care must be exercised in the selection of the correct bathymetric resolution. It will be demonstrated that the correct resolution depends on the source of the geoacoustic model used. Higher resolution bathymetrics, such as multibeam surveys, introduce steeper slopes that yield larger bottom losses. If a lower resolution bathymetry was implemented to conduct geoacoustic inversion, then the same bathymetric resolution must be used to conduct subsequent sound transmission modeling. In addition, when geoacoustic inversion results are reported, the bathymetric resolution used should also be reported.

1:35

1pUWa3. An autonomous underwater vehicle as sound source for geoacoustic inversion. A. Vincent van Leijen (NLDA, P.O. Box 10000, 1780 CA, Den Helder, The Netherlands)

During the MREA07 trials an autonomous underwater vehicle (AUV) was deployed to run a bathymetrical survey in a shallow water environment off the coast of Italy. Apart from much coastal and recreational activity in the area, self-noise of the REMUS AUV was continuously observed on a sparse vertical receiver array that was deployed from a small boat at anchor. As the vehicle ran its survey with a speed of 5 kts and at a depth of 30 m, the propulsion system radiated harmonic related tones of low frequencies that were received in strong interaction with the sea bottom. Three selected tones were then exploited for the geoacoustic inversion of a local range-independent environmental model. The inverted parameters describe a soft sediment layer on top of a hard sub bottom; the obtained model with specified uncertainty is shown to match seismic and bathymetric data of the area.

1:50

1pUWa4. Inversion of sediment properties from observations of scattering of buried cylinders using subcritical ensonification. William Sanders, Dale Bibee (Naval Res. Lab., Sea Fl. Sci. Branch, Stennis Space Ctr., MS 39529), and David Calvo (US Naval Res. Lab., Washington, DC 20375)

It is well known that objects buried near the water-sediment interface may be ensonified beyond the critical angle. Although the evanescent portion of surface waves decays rapidly with depth, objects near the surface may reradiate significant amounts of energy, especially at lower frequencies. Moreover, surface roughness may serve as a mechanism for enhancing the penetration of energy at higher angles. This study addresses the viability of inferring and then utilizing geo-acoustic properties from such scattering observations. A synthetic data set is generated for testing inversion using genetic algorithms (using the SAGA toolkit developed by Peter Gerstoft) and by use of a new model for predicting scattering from buried cylinders, target response using matched asymptotics. Distributions of posterior errors in the inverted parameters are presented. [This work was supported in part by the Naval Research Laboratory under program element PE62782N.]

2:05

1pUWa5. Accuracy of passive matched-field, thick-sediment inversions in range-dependent environments. Donald Del Balzo and Lisa Pflug (Planning Systems Inc., 40201 Hwy 190 East, Slidell, LA 70461)

Accuracy of shallow-water acoustic predictions is severely limited by uncertainty in sediment property characteristics. Recent research, using inverse methods to invert signals from controlled active sources, has been conducted to estimate seabed properties. However, some applications prefer a more covert approach. Our work is focused exclusively on passive, inversion techniques using signals from surface ships-of-opportunity. This study addresses the accuracy of low-frequency ($\sim$100 Hz) matched-field correlations using 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 observations are made. Matched-field techniques are applied in a simulated shallow-water environment with a single vertical line array and high signal-to-noise ratios. The simulations indicate

significant potential for accurate estimates of thick-sediment characterizations of grain size and layer thickness out to ranges of tens of water depths in shallow water, despite moderate mismatch conditions in the environmental model. [Work sponsored by SPAWAR.]

2:20

1pUWa6. Bottom and sub-bottom roughness effects on coherence and geo-acoustic inversions. Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039)

In this discussion, the effect of bottom and sub-bottom roughness on acoustic propagation in shallow water is examined. Generally in underwater acoustics, the acoustic medium is quantized as a stratified medium, with dominant changes occuring in depth and the range-dependent environmental variability (apart from bathymetry) is neglected. For a few environments presented in this paper, this approximation is quite poor. Scattering from bottom and sub-bottom roughness can lead to loss of signal and is often treated as an anomolous attenuation. Energy scattered from a rough hard bottom or basement is not removed from the water column but instead propagates as incoherent energy reducing the measured coherence and leading. A numerical model that incorporates random bathymetric and sub-basement roughness into the parabolic equation is presented. Comparisons of the effects of roughness will be made for an experiment over Basalt (BASSEX2004) and limestone (CALOPS 2007), both of which are thought to be hard, rough basements, covered with sand. Comparisons of measured and predicted coherence, geo-acoustic inversion results, and eigenspectra will be presented.

TUESDAY AFTERNOON, 27 NOVEMBER 2007 GRAND BALLROOM D, 3:00 TO 4:20 P.M.

Session 1pUWb

Underwater Acoustics: Rough Surface Interface Scattering

Guy V. Norton, Chair
Acoustics Div., Naval Research Laboratory, Stennis Space Center, MS 39529-5004

Chair's Introduction—3:00

Contributed Papers

3:05

1pUWb1. Acoustic scattering and propagation through spatially complex dispersive media via the direct numerical solution of the wave equation. Guy V. Norton (Naval Res. Lab., Stennis Space Ctr., MS 39529, guy.norton@nrlssc.navy.mil)

Analytical approaches for modeling propagation and scattering in inhomogeneous media are difficult, and often require extremely simplifying approximations in order to achieve a solution. To avoid such approximations, the direct numerical solution of the wave equation via the method of finite differences offers the most direct tool, which takes into account diffraction and refraction. It also allows for detailed modeling of spatially complex media. In all cases, the correct inclusion of the dispersive properties of the medium can make the difference in the interpretation of the results. However, the inclusion of dispersion directly in the time domain proved until recently to be an elusive problem. In order to model the transient signal, a convolution operator that takes into account the dispersive characteristics of the medium is introduced to the linear wave equation. In this work, numerical modeling of scattering from a variety of different dispersive media with physical properties associated with real world material is presented. Where possible, the numerical solutions are compared with a known solution. It is shown that in those cases, the use of the convolutional propagation operator leads to the correct solution for the scattered field. [Work supported by ONR/NRL.]

3:20

1pUWb2. A geometric method for calculating the Doppler spreading of surface scattered signals. James Preisig (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Grant Deane (Scripps Inst. of Oceanogr., La Jolla, CA 92093), and Chris Tindle (Univ. of Auckland, Auckland, New Zealand)

Recent work has shown that the characteristics of surface scattered high frequency signals can significantly impact the performance of underwater acoustic communications systems in shallow water environments. A combination of high signal intensities due to the focusing of the signals by the curved surface waves and high Doppler spreads due to the motion of the surface waves make it difficult to accurately track the time varying channel impulse response as required by most high rate communications systems. However, algorithms that can model and accommodate the Doppler spreading have been shown to maintain acceptable performance in these environments. The ability to predict the expected Doppler spreads of the surface scattered signals as a function of source/receiver geometry and surface wave characteristics can improve the performance of some of these algorithms. A method for calculating the Doppler spreading of surface scattered signals as a function of the source/receiver geometry and surface wave slope, curvature, and velocity is presented. Comparisons between theoretical predictions and observed Doppler spreads in collected acoustic data are presented.

3:35

1pUWb3. Low frequency enhancement of water-to-air sound transmission. B. Edward McDonald and David Calvo (Naval Res. Lab., Washington, DC 20375)

Low-frequency acoustic radiation into the air from an underwater point source is investigated using plane wave expansion of the source spectrum and Rayleigh reflection/transmission coefficients in order to clarify a recent claim of ''anomalous transparency'' of the water-air interface at low acoustic frequencies [Phys. Rev. Lett. 97, 164301, 2006]. Carrying the analysis to first order in the air/water density ratio, it is found that as the depth-wave number product kD becomes small, radiation into the water is quenched by the source's acoustic image, while the power radiated into air increases to about one percent of the power that would be radiated into unbounded water. For kD less than about 0.093, a majority of the one percent surviving total energy is transmitted to the air. In our view it seems an exaggeration to refer to this as ''anomalous transparency.'' [Work supported by the Office of Naval Research.]

3:50

1pUWb4. Resolving travel-time variability of the coupled oceanographic-tomographic analysis and prediction system (COTAPS). Irena Lucifredi and Peter Stein (Sci. Solutions Inc., 99 Perimeter Rd., Nashua, NH 03063)

It has been suggested that the tomographic approach that feeds the 8–11 kHz ocean acoustic measurements into an ocean model using data from bottom-mounted hydrophones, suffers from a significant travel-time variability due to the surface waves. In this work, the influence of the ocean surface motion on the acoustic travel times has been studied using both the experimental data collected at the Pacific missile range facility (PMRF) off Kauai, and an appropriate acoustic model. The ray-based acoustic model developed is an adaptation of the range-dependent version of Bellhop with the modified top surface characterizing ocean surface waves. The model-generated time series suggests a direct relationship between the phase of the top surface wave and the time-of arrival, which is further confirmed by the experimental data. Furthermore, an approach is developed that automatically decouples the effect of the top surface waves from the environment-related phenomena, making it possible to assign the travel time variations produced by the physical phenomena independently of the surface geometry, in turn eliminating the surface related error. [Work supported by the ONR.]

4:05

1pUWb5. Effect of rough surface on received signals. Jeremy M. Joseph, Richard Lee Culver, Jeffrey A. Ballard, and Colin W. Jemmott (Grad. Program in Acoust., The Penn State Univ., PO Box 30, State College, PA 16804, jmj269@psu.edu)

The overall goal of our research is to incorporate knowledge of variability and uncertainty in received passive sonar signals into the signal processing. As part of this goal, we need to better understand the sources of this variability. It is well known that wave motion on the ocean surface can introduce significant effects on undersea acoustic signals. This talk focuses on predicting the impact of rough ocean surfaces on the variability of the received signal. An acoustics propagation code, range-dependent acoustic model (RAM) rough surface fine coarse (RRSFC), is utilized to simulate environments with flat and rough ocean surfaces. To explore the effects of random rough surfaces, a Monte Carlo method is applied to yield environment realizations with varying rough surfaces. From these simulations, probability density functions are generated and are compared with amplitude figures from environments with a flat surface in order to distinguish any overall effect on a received signal. [Work supported by ONR Undersea Signal Processing Code 321US.]

Note: Payment of additional registration fee required to attend tutorial.

TUESDAY EVENING, 27 NOVEMBER 2007 GRAND BALLROOM E, 7:00 TO 9:05 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Weather and Acoustics

James P. Chambers, Chair

National Center for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Tutorial lecture on weather and acoustics. Alfred J. Bedard (NOAA Environ. Technol. Lab., 325 Broadway, Boulder, CO 80303)

The relationships between sound and weather can be fascinating, frightening, useful, and at times mystifying. This tutorial lecture explores the range of intersections between weather and acoustics. Weather can affect acoustic environments causing noise increases, noise reductions, and sound focusing. One aspect of this tutorial reviews results from propagation modeling, indicating that under some conditions, the atmosphere can produce vertical wave guides. Conversely, sound can be used to actively interrogate the atmosphere and provide information valuable for weather prediction and warning. Probing capabilities reviewed, with examples, show that wind profiles, temperature profiles, wind shears, gravity waves, and inversions can be defined acoustically. There are also possibilities for monitoring other difficult-to-observe parameters such as humidity profiles. In addition, weather processes can generate sound, detectable at long ranges using lower frequencies. Specifically, observing networks have observed infrasound from a growing number of meteorological events (e.g., severe weather, tornadoes, funnels aloft, atmospheric turbulence, hurricanes, and avalanches). Efforts to develop an infrasonic tornado detection system are described in some detail. Results indicate promise to help improve tornado detection and warning lead times, while reducing false alarms. Clear opportunities exist for infrasonic systems to provide operational weather data.

Session 2aAA

Architectural Acoustics and Structural Acoustics and Vibration: Impact and Footfall Noise I

David Braslau, Chair
David Braslau Associates, Inc., 1313 5th St., SE, Ste. 322, Minneapolis, MN 55414

Chair's Introduction—8:25

Invited Papers

8:30

2aAA1. Transmission of impact sound in wood-framed construction. J. David Quirt and Trevor R. T. Nightingale (Natl. Res. Council, Ottawa, Canada, K1A 0R6)

Ongoing studies in the flanking transmission facilities at NRC-IRC have developed a framework for predicting the overall sound transmission, including flanking, in wood-framed multifamily buildings. Power flow via each flanking path is affected by five factors that determine a path transfer function specific to the type of excitation (airborne or impact) and the construction details. This paper examines how common construction details affect transmission of impact sound between adjacent rooms, which are either side-by-side or one above the other. Since the source room floor is an element common to all impact flanking paths, a series of toppings were evaluated as treatment options. Additional layers of gypsum board and improvement due to resilient mounting were evaluated for the walls in the receiving room. Estimates of the resulting impact SPL were obtained by summing the energy transmitted by the direct path through the floor assembly separating the pair of rooms (if applicable) and all the flanking paths involving the relevant wall-floor junctions. These estimates provide the basis for a design guide to predict sound isolation in typical wood-framed row housing or apartment buildings.

8:50

2aAA2. Novel method of measuring sound from footfall to understand impacts in space below. Steven D. Pettyjohn (The Acoust. & Vib. Group, 5700 Broadway, Sacramento, CA 95820-1852, spettyjohn@acousticsandvibration.com)

Sound generated by people walking overhead made a conference room and law library unusable in the offices of a lawyer. Employing sound transmission class (STC) and impact insulation class (IIC) test procedures to quantify sound impacts in the space proved useless in understanding the complaints caused by the footfall. That is, the STC ratings and IIC ratings were sufficiently high that complaints would not have been expected. The floor/ceiling assembly was light weight with 28 ft long girders and open web trusses 35 ft long and 24 in. on center. The floor was constructed with two layers of plywood on top, and the trusses and the ceilings were either gypsum board or suspended acoustical tiles. Floors were finished with commercial carpet, vinyl composition tile, and carpet and padding. Short duration sound measurements in the frequency range 12.5 Hz to 10,000 Hz were made to quantify the difference in background sound and footfall generated sound. The very large differences in sound pressure levels at individual one-third-octave band frequencies and the A-weighted sound level clearly shows why complaints would be expected. The influence of floor and ceiling finish is seen in these measurements. The results provide clues to solutions to the problem with the assembly.

9:10

2aAA3. Development of a two-parameter system for evaluating impact noise insulation. John J. LoVerde and D. Wayland Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Impact insulation is currently described with a single number metric impact insulation class (IIC), which correlates poorly with subjective reaction and does not often properly rank order acoustical performance of assembly components. Modifications to IIC have been proposed, but the methods are limited to changing the source or the metric. Experience indicates that impact noise can be divided into two broad classes: Low frequency thudding (e.g., footfalls) and mid/high frequency noise (e.g., heel clicks, dragging furniture, etc.). Low-frequency impact sound pressure level (ISPL) is generally influenced by different variables than the mid/high-frequency ISPL; the authors propose that the two are largely independent. Therefore, two domains are required to adequately characterize the impact noise insulation of an assembly [LoVerde and Dong, 33rd Inter-Noise Proceedings Paper 296, 167 (2004)]. A variety of candidate two-domain systems have been reviewed, where one domain describes the low-frequency performance and the second domain the mid/high-frequency performance [LoVerde and Dong, 14th International Congress Sound and Vibration Paper 476 (2007)]. The two domains are calculated using third-octave ISPL data obtained from recent field impact noise tests. The results suggest that a two-domain system offers considerable improvement over current single-number metrics for evaluating and designing assemblies.

9:30

2aAA4. Repeatability and reproducibility of field impact insulation class testing. John J. LoVerde and D. Wayland Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Field impact sound insulation in the United States is typically quantified in terms of field impact insulation class, which is defined in ASTM E1007. According to the method, the precision of this test method has not been established, and to our knowledge there have been no studies attempting to evaluate the precision of this type of measurement. Precision is typically defined in terms of repeatability and reproducibility, which will have different definitions for field and laboratory applications. Repeatability in the field may be thought of as the variation in acoustical performance when the test is repeated for the same specimen in the same location. Reproducibility in the field may be thought of as the variation in acoustical performance of the same design, constructed in different locations but in the same manner and with the same personnel and materials. The repeatability and reproducibility of a number of field tests of similar assemblies are calculated, and the precision of field impact testing is evaluated.

9:50

2aAA5. Evaluation of field impact sound pressure levels as a function of tapping machine location. John J. LoVerde and D. Wayland Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Field impact sound insulation in the United States is typically quantified in terms of field impact insulation class, which is defined in ASTM E1007. This standard requires averaging the receiving room impact sound pressure levels (ISPL) generated by the tapping machine in four defined positions. In a previous paper [LoVerde and Dong, J. Acoust. Soc. Am. **121**, 3113 (2007)] the authors analyzed several hundred field tests to determine effects of reducing the number of tapping machine positions on the FIIC rating. In this paper the analysis is extended to look at the detailed effects on the third-octave band ISPL, which may impact design decisions more than the overall FIIC rating. The effect of reducing the number of positions as a function of assembly construction is also examined.

10:10–10:25 Break

10:25

2aAA6. Impact noise and walking noise in multifamily housing. Jerry Lilly (JGL Acoust., Inc., 5266 NW Village Park Dr., Issaquah, WA 98027)

Field measurements of impact sound transmission using the ISO standard tapping machine are compared with field measurements of human walkers in both wood and concrete buildings. In wood frame construction, low-frequency noise dominates the spectrum with most floor coverings. In concrete buildings, high-frequency noise is the controlling factor in non-carpeted areas. The relative effectiveness of different noise mitigation measures will also be discussed.

10:45

2aAA7. Building vibration excited by step dancing—Case study. Ronald Eligator (Acoust. Dimensions, 145 Huguenot St., New Rochelle, NY 10801, religator@acousticdimensions.com)

A new three-story building (steel frame, metal deck, concrete floor slabs) for the performing arts includes a dance studio on the upper level. An isolated concrete floor was installed within the dance studio, which is effective at attenuating impact sources (as modeled by a standard tapping machine), and normal foot fall (walking, light dancing, or similar). Hard-shoe, percussive dance (e.g., step dancing) by solo dancers and especially a chorus of 20 to 30 dancers, generates significant foot fall noise and vibration within the building, that is audible in the spaces one and two floors below the dance studio, and causes perceptible vibration of the building structure. Structural analysis relating natural frequency of building structure with natural frequency of isolated floor and forcing frequency of dancers was performed. Noise and vibration testing, with tapping machine and dancers, was also performed. Stiffening of slab with additional steel column to reduce effective span was effective in reducing magnitude of intrusive noise and vibration, as demonstrated by testing with dancer as source.

Contributed Papers

11:05

2aAA8. Field impact isolation testing failures due to lowest frequency performance deficiencies. Marlund E. Hale (Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065, mehale@aol.com)

There is a continuing demand in the construction of multifamily residences and in the renovation of apartments and condominiums to install hard-surfaced flooring in upper level dwelling units. Some municipalities and/or homeowners' associations have enacted regulations requiring that floor/ceiling assemblies meet specific minimum laboratory tested and/or field impact insulation performance standards. This paper presents field performance test results for several new and renovated condominiums in different multifamily residential buildings that failed both regulatory and homeowners' association impact performance requirements. Further investigations and research into tests by others indicated that an inordinate number of new and renovated condominiums are failing the FIIC performance requirements only in the lowest test band frequency. Potential reasons for such failures are discussed and remedies are recommended.

11:20

2aAA9. Acoustic characterization of in-room footfall noise. Samantha B. Rawlings, Joshua A. Magee, and Robert D. Celmer (Acoust. Program + Lab., Mech. Eng. Dept., Univ. of Hartford, 200 Bloomfield Ave., W. Hartford, CT 06117, celmer@hartford.edu)

Existing literature concerning footfalls is primarily focused on its transmission between spaces, such as a floor/ceiling's Impact Isolation Class. This study measured in-room sound power spectra produced by footfalls on eleven different floor surfaces using human subjects and a standard tapping machine. Within a reverberation room, 7 male and 7 female subjects walked on each floor while wearing 3 different types of footwear: Leather-soled with hard heels, rubber-soled with rubber heels,

and sneakers. Sound power spectra were measured in 1/3 octaves using the procedures of ISO 3741. A tapping machine was also used on each floor profile using both standard drop weights and with cored samples of the same shoe soles attached to the bottom of each weight. The data for each floor profile produced averages by shoe type along with corresponding 95% confidence intervals. A total of 143 average sound power spectra were generated, based on eleven floors with thirteen averages per floor (3 male, 3 female shoes, 6 shoed tapping machines, and 1 bare tapping machine). Spectral results for each shoe/floor were studied and compared. Correlations between sound power spectra produced with human walkers versus tapping machine were investigated. [Work supported by The Paul S. Veneklasen Research Foundation.]

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Tissue Response to Acoustics and Vibrations

Tyrone M. Porter, Chair
Boston Univ., Dept. of Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Invited Papers

8:35

2aBB1. Spatiotemporal response of mammalian cells to sonoporation measured using calcium imaging. R. E. Kumon and C. X. Deng (Dept. of Biomed. Eng., Univ. of Michigan, 2200 Bonisteel Blvd., Ann Arbor, MI 48109-2099, rkumon@umich.edu)

Sonoporation, ultrasound-induced plasma membrane poration, permits the passage of extracellular agents into the cytoplasm that may otherwise be impermeable, thereby making it useful for drug and gene delivery. Ca^{2+} is an important second messenger (signaling) molecule found to play a role in cell recovery after sonoporation. To better understand the dynamical responses of cells to sonoporation, calcium imaging of Chinese hamster ovarian cells was performed using digital video fluorescence microscopy and the Ca^{2+}-indicator dye fura-2AM. Spatiotemporal evolution of intracellular $[Ca^{2+}]_i$ was determined from real-time fluorescence measurements of cells in the presence of Optison^{TM} microbubbles and various extracellular $[Ca^{2+}]_o$ levels before, during, and after exposure to 1 MHz ultrasound tone bursts (0.2 s, 0.45 MPa). In cells where ultrasound caused a direct microbubble-cell interaction, breaking and eventual dissolution of the microbubble were observed with concomitant permeabilization of the cells to Ca^{2+}. These cells exhibited a large influx of Ca^{2+}, and sometimes the $[Ca^{2+}]_i$ did not return to its equilibrium level. Subsequently, other surrounding cells often exhibited complex $[Ca^{2+}]_i$ dynamics, including intracellular oscillations and intercellular waves. Efforts to model the time evolution of $[Ca^{2+}]_i$ in individual cells with analytical and numerical approaches will also be discussed. [Work supported by ACS and NIH.]

8:55

2aBB2. Ultrasound-assisted skin diagnostics. Sumit Paliwal, Makoto Ogura, and Samir Mitragotri (Dept. of Chemical Eng., Univ. of California, Santa Barbara, CA 93106)

The biomolecular composition of human skin, represented by a multitude of lipids, proteins, nucleic acids, and other miscellaneous molecules, is a sensitive indicator of skin's local health, as well as the body's systemic health. However, this information is not easily accessible, and consequently not used in the current diagnostic methods. To address the lack of quantitative methods for non-invasive skin diagnostics, herein, we describe the use of ultrasound to rapidly extract biological markers from skin. Specifically, we extracted several types of proteins (keratins, beta-actin, cytokines, and others), lipids (cholesterol, ceramides, fatty acids, and others), nucleic acids, small molecules (natural moisturizing factors and exogenous molecules), and microbes enabling us to create skin's unique biomolecular signature. The extracted biomolecules showed quantitative correlations with their composition in the epidermal skin. Extraction of these entities from skin using our method is convenient and reproducible, thus making it an advantageous alternative to biopsy and tape-stripping. By quantifying changes in skin's biomolecular profile, we report several specific in vivo applications of our method for assessing chemical-induced skin irritation, atopic dermatitis, and psoriasis in animal models. As a corollary to this technique, measurement of skin's drug bioavailability in rats is also demonstrated.

9:15

2aBB3. Acoustic hemostasis: Underlying mechanisms. Shahram Vaezy, Lawrence Crum, Steve Carter, Grant O'Keefe (Univ. of Washington, Box 355061, Univ. of Washington, Seattle, WA 981195), Vesna Zderic (George Washington Univ.), Roy Martin (Biz Res. Inc.), and Riyad Karmy-Jones (Southwest Medical Ctr.)

Cessation of hemorrhage using extrinsic, interventional methods is possible with delivery of energy to bleeding tissues, i.e., cauterization. High intensity focused ultrasound (HIFU) is one such method, with significant advantages of delivering high levels of energy to well-defined regions of deep-seated tissues, even during profuse bleeding. The physical mechanisms involved in this process include thermal and mechanical effects of HIFU, leading to various biological effects. Our results using HIFU devices of 1–5 MHz, and focal, derated intensities of 1,000–10,000 W/cm^2, in solid organs such as liver, spleen, and kidneys and major and minor blood vessels, show that temperature of targeted tissues reaches 70–100°C within seconds, with formation of microbubbles, approximately

5−−200 μ in size, and concentration of 100 bubbles/mm^3. It appears that boiling of interstitial fluids and blood and acoustic cavitation are both involved. The biological effects include coagulative necrosis, mechanical disruption of tissue structure potentially leading to release of tissue factors enhancing the coagulation, coagulum and thrombus formation at a wound site, tissue fusion via collagen and elastin remodeling, and fibrin plug formation, with minimal damage of the surrounding tissues. These mechanisms appear to provide an effective and safe method of hemorrhage control. [Work supported by NIH, DoD, NSBRI.]

9:35

2aBB4. Temporary blood-brain barrier disruption via focused ultrasound for targeted drug delivery in the brain. Nathan McDannold (Brigham & Women's Hospital, Dept. of Radiol., Harvard Med. School, Boston, MA 02115)

The ability to deliver drugs to the central nervous system is severely limited by the presence of the blood-brain barrier (BBB). This barrier normally serves to protect the brain and restricts the passage of most molecules from being delivered to the brain tissue from the blood stream. We have developed a method by which the BBB is temporarily disrupted in localized regions using low-power ultrasound pulses combined with a circulating ultrasound contrast agent. By systematically focusing at multiple overlapping locations, one could prescribe the region in which the BBB is disrupted to conform to a desired target volume, reducing drug penetration to the rest of the brain. Ultrasound parameters suited for this procedure are compatible with a device that can focus through the intact skull. In this presentation, we will describe recent research on this technique, discussing methodology, safety, possible mechanisms, and potential means to guide and monitor the procedure.

9:55−10:10 Break

Contributed Papers

10:10

2aBB5. Influence of inhomogeneity and geometry on lung response to low-frequency underwater sound. Mark S. Wochner (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029), Yongjie Zhang (Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213), Yurii A. Ilinskii, Mark F. Hamilton, and Evgenia A. Zabolotskaya (Univ. of Texas, Austin, TX 78713-8029)

An effective medium, finite-element model has been developed for human lung response to low-frequency underwater sound [Wochner et al., J. Acoust. Soc. Am. **121**, 3082 (2007)]. Macroscopic properties are determined by averaging over microscopic properties associated with lung tissue elasticity and alveolar geometry. The resulting elastic constants depend on lung volume, and thus vary with the phase of the breathing cycle. This presentation discusses effects due to spatial variations in lung and geometric features specific to the anatomy of lung. An important variation in elastic properties occurs where soft tissue of the parenchyma attaches to much stiffer bronchial tissue, resulting in stresses that may cause injury at high levels of acoustic excitation. A factor affecting inhomogeneity within the parenchyma is the orientation of the swimmer. When vertical, the suspension of the lung under its own weight may cause the stiffness to be higher at the top of the lung than at the bottom, resulting in a change in the lung resonance frequency. Finally, patient-specific geometries based on volumetric data of the human thorax are considered. Results will be presented showing stress and strain fields, lung resonance, and mode shapes. [Work supported by ONR and ARL:UT IR&D.]

10:25

2aBB6. Auscultation using ultrasound-Doppler techniques. William A. Ahroon, Adrianus J. M. Houtsma, and Ian P. Curry (US Army Aeromedical Res. Lab., Fort Rucker, AL 36362-0577)

An earlier presentation described a hybrid instrument using electromechanical and ultrasound-Doppler techniques to permit auscultation in extremely noisy environments, such as Army medical evacuation vehicles [Houtsma et al., J. Acoust. Soc. Am. **120**, 3361 (2006)]. That paper presented the theories behind the hybrid instrument and provided simulations of auscultation in noisy environments. An update on the progress since the first report is presented. In one study, recordings of cardiac and pulmonary sounds from healthy volunteers aboard an in-flight Black Hawk helicopter are presented. As anticipated, auscultation with the electro-mechanical (acoustic) mode was impossible inside the aircraft. Using the ultrasound-Doppler mode, all heart sounds and most lung sounds were clearly audible. The best-audible lung sounds came from the mid-axillary left and right locations. A second study was performed to determine if the hybrid instrument could be used to detect the difference in sound in a correctly versus incorrectly ventilated swine. After some experimentation with precise auscultation sites, the listener was able to detect both heart and lung sounds reliably in both electromechanical and ultrasound-Doppler modes. Recommendations to improve the performance of the instrument in the next iteration of the device are presented.

10:40

2aBB7. Finite-element modeling of microsphere surface modes and high-frequency ultrasound scattering from a single cell. Omar Falou (Dept. of Elec. and Comput. Eng., Ryerson Univ., 350 Victoria St., Toronto, ON, Canada, M5B 2K3, ofalou@ryerson.ca), J. Carl Kumaradas, and Michael C. Kolios (Ryerson Univ., Toronto, ON, Canada, M5B 2K3)

A finite-element model of wave propagation using COMSOL Multiphysics (COMSOL Inc., Burlington, MA) was developed to solve the problem of high-frequency ultrasound (10–70 MHz) scattering from spheres. This model is being used to predict ultrasound backscatter from cells for ultrasound tissue characterization. In this work, backscatter from an elastic (polystyrene) microsphere was used to validate the computational model against experimental results. Agreement between experimental results and finite-element solutions were found in the frequency domain solutions. Moreover, conversion of the theoretical solutions to time-domain backscatter data showed good agreement with experimental time domain measurements. To develop a better understanding of the interactions of the incident pressure field with the scattering object, oscillations of the elastic microsphere at various resonance frequencies were also investigated. A systematic relationship between the resonance frequency and its corresponding surface mode was found. An ultrasound scattering model by a single cell is also presented. The model treats the cell as an elastic sphere (nucleus) surrounded by a fluid shell (cytoplasm). Comparison of the theoretical backscatter predicted by the model and experimental measurements for acute Myeloid Leukemia (AML) cell is also shown. Finally, the implications of these results on ultrasound tissue characterization techniques are discussed.

10:55

2aBB8. Characterization of trabecular bone structure through acoustic scattering. Max Denis, Charles Thompson, and Kavitha Chandra (Univ. of Massachusetts Lowell, One University Ave., Lowell, MA 01854)

In this work, observations of the relationship between the internal structure of a three-dimensional trabecular bone and scattering acoustic waves from the bone tissue are investigated. A three-dimensional scattering field base on the perturbation expansion of the fluid and solid constituents of the bone tissue is presented. Due to the anisotropic nature of the trabecular bone, eigenvalues and eigenvectors of the bone tissue's moment of inertia tensor were used to determine its principal axis. The eigenvector associated with the largest eigenvalue was considered as the unit vector along the principal axis. This allowed for a stationary analysis of the medium. Using spatial autocorrelation, estimates are made of the inner length scales of the medium. Observations were made of the simulated exterior field's relation to the bone tissue's inner length scales. Of particular interest are features of anisotropy and frequency dependence of the scattering field.

11:10

2aBB9. Fast integral equation algorithm for simulation of acoustic energy transfer to the inner ear through multiple pathways. Elizabeth Bleszynski, Marek Bleszynski, and Thomas Jaroszewicz (Monopole Res., 739 Calle Sequoia, Thousand Oaks, CA 91360)

Recent progress is described in our development of a volumetric integral equation-based algorithm applicable to accurate large scale simulations of propagation of elasto-acoustic waves through inhomogeneous media. The primary goal of our work is to construct an efficient and high fidelity numerical simulation tool for investigating such effects as, e.g., acoustic energy transfer to the inner ear via nonairborne pathways. The considered algorithm achieves high efficiency through the use of an FFT-based nonlossy matrix compression and makes possible rigorous simulations involving realistic geometries characterized by highly subwavelength details, large density contrasts, and described in terms of several million unknowns. The recently added algorithm elements are the node-based Galerkin discretization and the solver extension to include pressure and shear elastic wave propagation. Representative examples of numerical simulations involving several million of unknowns and illustrating the method accuracy and effectiveness in simulating acoustic/elastic field distribution in human head are presented. [Work is supported by AFOSR.]

11:25–12:00
Panel Discussion

WEDNESDAY MORNING, 28 NOVEMBER 2007 NAPOLEON B1, 8:00 TO 11:40 A.M.

Session 2aEA

Engineering Acoustics and ASA Committee on Standards: Infrasonic Instrumentation I

Qamar A. Shams, Cochair
NASA Langley Research Center, 4 Langley Blvd., Hampton, VA 23681

Allan J. Zuckerwar, Cochair
4909 Camberley Cir., Williamsburg, VA 23188-8801

Chair's Introduction—8:00

Invited Papers

8:10

2aEA1. Developments in infrasound sensors and wind noise reduction. Henry E Bass, Carrick Talmadge (Natl. Ctr. for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University MS 38677), Michael Hedlin, Kris Walker, Mark Zumberge (Univ. of California, San Diego, La Jolla, CA 92093), and Milton Garces (Univ. of Hawaii, Manoa, Kailua-Kona, HI 96740)

As applications of infrasound grow so does the need for reasonably priced microphones and attendant wind noise reduction approaches. Many applications require sensor systems with frequency response between 0.01 and 10 Hz and a dynamic range between a few micropascals and ten Pascals and operate under severe environmental conditions. Sensor systems are often in remote areas with only occasional maintenance visits. Current installations usually employ classic condenser microphones or sliding wire microbarographs along with porous hose wind filters. Next generation sensors use optical fiber sensors long enough to average out wind noise or multiple sensors which allow for electronically canceling wind noise. Some of the advantages and characteristics of these sensors will be described. Calibration of sensors over the range of conditions remains a challenge. Progress on calibration facilities will be included in the discussion. [Work supported by the US Army Space and Missile Defense Command.]

8:35

2aEA2. Extraction of infrasonic waveforms from conventional condenser microphones. Thomas B. Gabrielson and Timothy M. Marston (Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

There are several commercial measurement-grade condenser microphones that are suitable for measurements down to about 0.1 Hz; however, these one-half-inch microphones reach those low frequencies with an in-line attenuator that compromises the overall dynamic range. In a previous paper [T. M. Marston and T. B. Gabrielson, J. Acoust. Soc. Am. **119**, 3378 (2006)], a process was described for digital reconstruction of the lowest frequencies that avoids the attenuator and the sacrifice in dynamic range. This work has been extended to evaluate a reconstruction process for a one-quarter-inch condenser microphone, which has significantly higher

low-frequency roll-off than the one-half-inch infrasonic microphones. One reason for using the one-quarter-inch microphone is its higher peak pressure limit. By modeling the preamplifier input-impedance bootstrapping and the pressure-equalization leak, extension to a few tenths of a hertz can be performed. However, the pressure-equalization leak characteristic frequency varies from microphone to microphone and with temperature. Consequently, there is a single free parameter that requires a measurement of the actual very-low-frequency response or measurement of a physically constrained reference waveform (like the *N*-wave from a clean sonic boom). [Work supported by the FAA/NASA/Transport-Canada Center of Excellence for Aircraft Noise and Aviation Emissions Mitigation.]

9:00

2aEA3. A device for reducing wind and breaking flow into small-scale eddies to decrease infrasonic noise. Alfred J. Bedard, Jr., Randal T. Nishiyama, and Bruce W. Bartram (NOAA/Earth Systems Res. Lab., Physical Sci., Div., Mail Code R/PSD99, 325 Broadway, Boulder, CO 80305)

The purpose of the fence developed is to raise the atmospheric boundary layer, and at the same time, break up the wind into small eddies, which interact destructively with one another and rapidly break down into smaller scales. This produces a calm region inside an enclosure or behind a wall with only random weak variations of the wind. Severe weather environments often have strong, complex regional flows. In the past, infrasonic systems have used spatial filters to average out the pressures from boundary layer turbulence. However, such spatial filters, while quite effective at lower wind speeds, do not reduce the wind noise sufficiently at higher wind speeds (e.g., above 10 m/s). This wind noise represented a critical limitation for infrasonic tornado detection systems. To solve this problem, we affixed the tops of barriers with corrugations. These cause the creation of adjacent vortex sheets of opposite signs, which interact strongly with one another and rapidly break down into smaller scales, and will not penetrate into regions behind the barrier. We use the eddy fence in conjunction with porous hose spatial filters. Although our motivation was to extend the effectiveness of infrasonic detectors in high wind conditions, there are numerous other applications.

9:25

2aEA4. Low frequency electret condenser microphone. Christopher C. Lawrenson (PCB Piezotronics, 3425 Walden Ave., Depew, NY 14043, clawrenson@pcb.com), Allan J. Zuckerwar, and Qamar A. Shams (NASA Langley Res. Ctr., Hampton, VA 23681)

A condenser microphone has been fabricated for measuring low-frequency sound pressure. The goal of this design is to keep the background noise as low as possible. The microphone features a high membrane compliance with a large backchamber volume, a prepolarized backplane, and a high impedance preamplifier located inside the backchamber. Methods for characterizing the performance of the microphone will be presented including background noise levels, which will be compared to commercially available microphones.

9:50

2aEA5. The Rotorcraft Noise Model (RNM). David Conner (U.S. Army AFDD-JRPO, AMRDEC, RDECOM, MS461, NASA Langley Res. Ctr., Hampton, VA 23681, david.a.conner@nasa.gov), Casey Burley (NASA Langley Res. Ctr., Hampton, VA 23681), Charles Smith, (Lockheed Martin; NASA Langley Res. Ctr., Hampton, VA 23681), and Dennis Pope (Analytical SVCS and Mater.; NASA Langley Res. Ctr., Hampton, VA 23681)

The Rotorcraft Noise Model (RNM) is a simulation program that predicts how sound will propagate through the atmosphere and accumulate at receiver locations located on flat ground or varying terrain, for single and multiple vehicle flight operations. Developed by the NASA Langley Research Center and Wyle Laboratories, RNM was initially intended to provide a tool to aid in the development of low noise terminal area operations for rotorcraft and tiltrotors, but has since been widely adopted for environmental impact studies. At the core of RNM are the vehicle source noise characteristics, input as sound hemispheres. As the vehicle flies along its prescribed flight trajectory, the source noise propagation is simulated and accumulated at the receiver locations (single points of interest or multiple grid points) in a systematic time-based manner. These sound signals at the receiver locations may then be analyzed to obtain numerous ground metrics or noise animation videos. A PC executable version of RNM is publicly available. This presentation will provide details on the theoretical framework of RNMs propagation model and the required inputs, discussion on capabilities for civil and military applications as well as details on output results. Code validation results and example outputs will be presented.

10:15–10:30 Break

10:30

2aEA6. Helmholtz resonator as a mobile source of infrasound. Hector L. Soto (NASA Langley Res. Ctr., M.S. 238, Hampton, VA 23188) and Allan J. Zuckerwar (Analytical Services and Mater., 107 Res. Dr. Hampton, VA 23666)

A Helmholtz resonator is designed to obtain a single-frequency, high sound pressure level (SPL) source of infrasound with minimum distortion. This apparatus is used to provide a simulated point source for applications ranging from sound propagation studies to locating an infrasonic sound source using triangulation. The resonator is designed to yield a frequency of 20 Hz and a high quality factor. The presentation will show background theory, design of the resonator, results of a built prototype, and challenges in obtaining a clean pure tone signal using this system.

10:55

2aEA7. Optical fiber infrasound sensor arrays: Signal detection and characterization capabilities in the presence of wind noise. Kristoffer Walker, Matthew Dzieciuch, Scott DeWolf, Mark Zumberge, Michael Hedlin, and Jon Berger (Inst. of Geophys. and Planetary Phys., Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., MC 0225, La Jolla, CA 92093-0225)

Optical fiber infrasound sensors (OFIS) are compliant tubes wrapped with two optical fibers that integrate pressure variation along the length of the tube with laser interferometry. The coherent signals traveling at acoustic speed are averaged along the length of the OFIS at the speed of light, while spatially incoherent noise traveling at wind speed is attenuated. We present new wind noise measurements, which augment initial work that suggested an OFIS has a lower noise floor above 1 Hz than a traditional microbarometer with a rosette pipe or hose filter in light wind. We also discuss OFIS configurations and techniques to determine the signal phase velocity direction. The instrument response is a function of the orientation of the OFIS relative to that of the wavefront. That property can be exploited to help determine the phase velocity of infrasound signals. This type of instrument-response-dependent beamforming is theoretically an improvement to standard beamforming for the same number of sensors when the signals are immersive or continuous in nature. When the signals are more impulsive in nature, the success of these techniques mostly depends on how successful the OFIS is in reducing wind noise and the signal bandwidth.

11:10

2aEA8. Simulation-based response analysis of porous and non-porous infrasonic arrays. Timothy M. Marston (Grad. Program in Acoust., The Penn State Univ., P.O. Box 30, State College, PA, 16802, tmm357@psu.edu) and Tetsushi Biwa (Tohoku Univ, Aoba-ku, Sendai 980-8579, Japan)

A numerical simulation for porous and non-porous infrasonic arrays has been developed to aid in the analysis of the response of these systems. For design flexibility, the simulation makes use of differential T-elements, which may be used as building blocks to assemble pipes and junctions of various shapes and sizes. The simulation adds to the capabilities of a previously developed model [Alcoverro and Le Pichon, J. Acoust. Soc. Am. **117**, 1717 (2005)], by including variable tube porosity, a calculation of wall-losses based on a thermo-viscous function, a calculation of the change in response as the azimuth and elevation of the incoming wave are varied, and a calculation of the wave shape as it propagates through the array in the time domain. [Research funded by the Applied Research Laboratory Educational and Foundational Fund.]

11:25

2aEA9. Advances in distributed arrays for detection of infrasonic events. Kevin Dillion, Wheeler Howard, and F. Douglas Shields (Mil-tec Res. Technol., Oxford Enterprise Ctr., Oxford, MS 38655)

Infrasound arrays normally consist of four to eight microbarometers spaced kilometers apart. Each of these microbarometers is connected to a pipe or porous hose array to reduce wind noise. This presentation describes a 100 sensor all weather distributed array and its use in continuously logging infrasound signals over an extended period of time. The array has been deployed next to a conventional porous hose array connected to a Chaparral 2.5 microbarometer. The effectiveness of the two arrays in canceling wind noise and detecting infrasound signals is compared. The hardware and software used by the distributed array to log the data and to process it so as to cancel wind noise, identify and separate the infrasound signals, and locate their sources will be discussed.

WEDNESDAY MORNING, 28 NOVEMBER 2007 GRAND COUTEAU, 8:30 TO 11:50 A.M.

Session 2aMU

Musical Acoustics and Speech Communication: Musical Pitch Tracking and Sound Source Separation Leading to Automatic Music Transcription I

James W. Beauchamp, Chair

Univ. of Illinois Urbana-Champaign, School of Music, Dept. of Electrical and Computer Engineering, Urbana, IL 61801

Chair's Introduction—8:30

Invited Papers

8:35

2aMU1. Computing pitch of speech and music using a sawtooth waveform inspired pitch estimator. Arturo Camacho and John G. Harris (CNEL, Univ. of Florida, P.O. Box 116130, Bldg. No. 33, Ctr. Dr., Rm. NEB 486, Gainesville, FL 32611, arturo@cnel.ufl.edu)

A powerful pitch estimation algorithm called SWIPE has been developed for processing speech and music. SWIPE is shown to outperform existing algorithms on several publicly available speech and musical instrument databases, and a disordered speech database, reducing the gross error rate by 40%, relative to the best competing algorithm. In short, SWIPE estimates the pitch as the fundamental frequency of a sawtooth waveform, whose spectrum best matches the spectrum of the input signal. The short-time

Fourier transform of the sawtooth waveform provides an extension to older frequency-based, sieve-type estimation algorithms by providing smooth peaks with decaying amplitudes to correlate with the fundamental frequency (if present) and its harmonics. An improvement on the algorithm is achieved by using only the first and prime harmonics, which significantly reduces subharmonic errors commonly found in other pitch estimation algorithms.

8:55

2aMU2. Error correction and accuracy scaling for a joint subtraction multiple fundamental frequency estimation algorithm. Geoffrey Herman and James Beauchamp (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champain, 1406 W. Green St., Urbana, IL 61801, glherman@uiuc.edu)

This talk presents an improved multiple fundamental frequency (F0) estimation algorithm, which is loosely based on a previously developed joint subtraction algorithm [A. Klapuri, Proc. 7th Int. Conf. on Music Information Retrieval, 216–221 (2006)]. The improved algorithm uses some auditory scene analysis techniques to supplement a maximum likelihood based F0 estimator to predict the best combination of F0s. The accuracy of this base algorithm is then improved by use of error detection/correction. Error rates of the algorithm were collected using random mixes of sounds gathered from the University of Iowa Theremin database of musical instrument samples, as well as mixes of continuous audio. The performance of both the base algorithm and the improved algorithm will be presented together. It was found that the improved algorithm decreases error rates by as much as 50%, compared to the base algorithm with minimal loss in computational efficiency. The trade-offs between computational efficiency and accuracy, tested via a tunable parameter built into the algorithm, will be discussed. Trade-offs between computational costs and accuracy as a function of the number of F0s will also be discussed.

9:15

2aMU3. A method for multiple F0 detection of continuous audio with dissimilar timbres. Mert Bay and James W. Beauchamp (School of Music and Dept. of Elect. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, mertbay@uiuc.edu)

This paper discusses a method for multiple fundamental frequency (F0) tracking performed on a polyphonic mixture of musical instruments with dissimilar timbres recorded on a single monaural track. For each time frame, given a number of possible F0 candidates, F0 trajectories are extracted using an HMM-like method where observations are the input spectra and each F0 candidate is represented by one of the hidden states. The F0 trajectories are optimized according to a harmonic overlap score for each F0 candidate and a continuity score that is based on a distance between the F0 candidate's harmonic amplitudes and frequencies calculated for all possible candidate pairs at successive adjacent frames. This step leads to the formation of note tracks. In a second step, instrument identification is performed on harmonic spectra associated with these note tracks to group similar timbres together. The method is demonstrated with an acoustic audio recording of a woodwind quintet. Applications include audio-to-MIDI conversion, instrument/voice separation, and automatic music transcription. F0 tracks can also be used as high level features for music information retrieval systems.

9:35

2aMU4. Towards estimation of the number of concurrent F0s. Chunghsin Yeh and Axel Roebel (IRCAM/CNRS-STMS, 1, place Igor Stravinsky, 75004 Paris, France)

Two central problems common to many multiple F0 estimation algorithms will be presented: The estimation and handling of the noise level and the modeling and representation of overlapped partials. Contrary to the widely used assumption that the noise is white Gaussian, it is proposed to use an adaptive and frequency-dependent noise level to be able to improve the description of sinusoidal partials and the expected residual. By using the frequency-dependent noise level, an F0 estimation algorithm can be guided to extract sources having designated harmonic-to-noise ratios. One of the major problems in multiple F0 estimation is partial collision. The assumption that is usually made to estimate the expected amplitudes of overlapped partials is the additivity of magnitude spectra or the additivity of power spectra. A mathematical investigation into the properties of overlapped amplitudes allows us to present an improved probabilistic description of the expected amplitudes. The implications of these two refinements for the estimation of the number of sources that are present in a recording will be discussed, and preliminary experimental results will be presented.

9:55

2aMU5. A new strategy on pitch estimation in polyphonic musical audio. Karin Dressler (Fraunhofer IDMT, Langewiesener Str. 22, 98693 Ilmenau, Germany, dresslkn@idmt.fraunhofer.de)

In this paper, a new perceptually motivated approach for pitch estimation in polyphonic musical audio is presented. The proposed algorithm is based on the pair-wise analysis of spectral peaks. Starting from a spectral representation of the audio signal, audible spectral peaks are selected with the help of a simplified psychoacoustic model. The idea of the technique lies in the identification of partials with a successive harmonic number (1 and 2, 2 and 3, etc.). Since two successive harmonics of a harmonic sound have a well-defined frequency ratio, a possible fundamental can be derived from the instantaneous frequencies of the two peaks. Furthermore, under certain conditions the harmonic integer ratios 5/3, 7/5, and 9/7 are also considered. Consecutively, the identified harmonic pairs are evaluated according to a perceptually motivated rating scheme, and the resulting pitch strengths are added to a pitch spectrogram. In addition to the efficient estimation of the fundamental frequency, the presented approach results in a quantitative prediction of the perceived pitch strength, allowing the identification of the predominant voices (e.g., melody and bass line) in music recordings. The method is robust in regard to the missing fundamental frequency phenomenon and works well with quasi-harmonic sounds.

10:15–10:30 Break

10:30

2aMU6. Multipitch detection for piano music: Benchmarking a few approaches. Bertrand David, Roland Badeau, Nancy Bertin, Valentin Emiya, and Gaël Richard (Ecole Nationale Supérieure des Télécommunications, 37-39 rue Dareau, 75014 Paris, France)

When trying to find a solution to the critical and sometimes ill-posed problem of multipitch estimation, it is common to have to choose between several approaches: Using a processing that resembles that of the human auditory perception, or using a decomposition adapted to the spectral content of the targeted sound category, taking into account an a priori knowledge of the spectral properties of musical notes or trying to learn some of its charateristics or even more, to run an algorithm that blindly tends to separate the sound into multiple elementary entities. This work involves some recently published techniques, such as the non-negative matrix factorization with sparsity constraints, a likelihood approach, based on a smooth spectral envelope model for both the background noise and for the partials, and a deterministic method combining spectral and temporal criteria. Their performance is comparatively assessed on a common multipitch database restricted to piano music, drawn both from home-made recordings and soft-synthesized sounds. The results are discussed with respect to the complexity, the versatility, the sensitivity to fine tuning, and the precision reached by each approach. [This work is granted by the French Groupement des Ecoles de Télécommunication (TAMTAM project).]

10:50

2aMU7. An approach for melody extraction from polyphonic audio: Using perceptual principles and melodic smoothness. Rui P. Paiva (CISUC Ctr. for Informatics and Systems of the Univ. of Coimbra, Dept. Informatics Eng, Univ. Coimbra, Polo II, P3030 Coimbra, Portugal)

In this research work, the problem of melody extraction from polyphonic audio is addressed. A multi-stage approach is followed, inspired on principles from perceptual theory and musical practice. Physiological models and perceptual cues of sound organization are incorporated into the method, mimicing the behavior of the human auditory system to some extent. Moreover, musicological principles are applied, in order to support the identification of the musical notes that convey the main melodic line. The system comprises three main modules, where a number of rule-based procedures are proposed: (i) Pitch detection, where an auditory model-based pitch detector is employed for selecting multiple pitches in each analysis frame; (ii) determination of musical notes (with precise temporal boundaries and pitches); and (iii) identification of melodic notes, based on two core assumptions that we designate as the salience principle and the melodic smoothness principle. Experimental results were conducted, showing that the method performs satisfactorily under the specified assumptions, namely when the notes comprising the melody are, in general, more intense than the accompanying instruments. However, additional difficulties are encountered in song excerpts where the intensity of the melody in comparison to the surrounding accompaniment is not so favorable.

11:10

2aMU8. Melody-based retrieval in audio collections. Matija Marolt (Faculty of Comput. and Information Sci., Univ. of Ljubljana, Trzaska 25, 1000 Ljubljana, Slovenia)

Mid-level representations are increasingly used to bridge the gap between high-level (semantic) and low-level audio representations. A mid-level representation that integrates melodic and rhythmic aspects of a music signal is introduced. The representation is formed by first performing multi-pitch detection on consecutive audio frames, and then searching for dominant melodic lines within the detected pitches. Beat-tracking is also performed to yield a beat-synchronous representation, independent of tempo variations. The representation is used as the basis for a melody-based audio retrieval system. An approximate nearest neighbor search algorithm is employed to compare sections of the mid-level representation of the query to indexed sections of mid-level representations of songs in a collection. Symbolic queries may also be used. Results are ranked according to the number of matched sections. A locality sensitive hashing function for cosine similarity is used for indexing, and the point location in equal balls algorithm for searching. Retrieval was tested on the cover song identification task, where the goal is to retrieve different interpretations of a song in a collection. Results on a collection of 2,400 songs are presented. [Work supported by ARRS.]

1130

2aMU9. Query-by-singing/humming systems based on speech recognition technologies. Motoyuki Suzuki, Akinori Ito, and Shozo Makino (Grad. School of Eng., Tohoku Univ., 6-6-05, Aramaki-Aza-Aoba, Aoba-ku, Sendai 980-8579, Japan, moto@m.ieice.org)

While many query-by-singing/humming systems have been developed, several problems have prevented their practical use. For example, pitch extraction methods are commonly used in QbSH systems, but they often yield inaccurate information. To rectify this problem, a QbH system (i.e., no lyrics required), which uses multiple pitch candidates, has been developed. For each input note, multiple pitch candidates are extracted, and all combinations of pitch candidates are considered in the retrieval step. This system has shown higher performance than QbH systems, which use single pitch candidates. QbS systems (i.e., lyrics required) have also been developed. In this case, one of the biggest problems is to segment an input singing voice into musical notes. Several methods have been proposed. However, it is hard to segment the singing voice without linguistic information. In the proposed system, a lyrics recognizer outputs several hypotheses for a recognition result. Each hypothesis gives time alignment information between the singing voice and the recognized text. Once this information is available, it becomes easy to segment the input singing voice into musical notes, and a hypothesis can be verified from a melodic point of view. This system, which has 156 songs in the database, yields 93% retrieval accuracy for a query consisting of five notes.

Session 2aNS

Noise: Lawn, Yard, and Portable Noise and Noise Generation, Modeling, and Measurement

Richard D. Godfrey, Cochair
Owens-Corning, Science and Technology, 2790 Columbus Rd., Route 16, Granville, OH 43023-1200

Karl B. Washburn, Cochair
John Deere and Company, Moline Technology Innovation Center, One John Deere Place, Moline, IL 61265

Chair's Introduction—9:00

9:05

2aNS1. Aero-acoustic noise generation of a lawnmower deck. Chris Peak (John Deere SouthEast Eng. Ctr., 14401 Carowinds Blvd., Charlotte, NC 28273, PeakChristopherA@JohnDeere.com)

Sound power and dynamic pressure measurements were used to identify where noise is generated by a single spindle lawnmower deck. The sound power measurements were performed with different blades and deck profiles to quantify the overall levels. An electric motor was used to turn the blade, and the measurements were made in a semi-anechoic chamber over an absorptive surface. The test procedures outlined in ISO 11094 were used as a guideline. The microphones used for the dynamic pressure measurements were located on a two-inch by two-inch grid, one inch below the cutting plane of the mower. The dynamic pressure data were used to generate one-third octave colormaps and movies of the pressure distribution under the mower deck. The colormaps and movies identified local hot spots where noise is generated, and they were able to distinguish between blade and deck configurations with a difference of two dB in the sound power level. [The work was completed to fulfill the requirements for a Masters Degree from The Pennsylvania State University Graduate Program in Acoustics. Dr. Jerry Lauchle was my advisor for the project.]

9:25

2aNS2. Computational aeroacoustic modeling of rotary mower noise. Karl Washburn (John Deere Moline Technol. Innovation Ctr., One John Deere Pl., Moline, IL, 61265), Christopher Peak (John Deere SouthEast Eng. Ctr., Charlotte, NC 28273), and Robert Sandboge (Pioneer Solutions, Inc., Northville, MI 48167)

The primary source of noise from rotary mowing machines is the aeroacoustic sound generated by the cutting blades. The requirements for maintaining high quality of the cut contradict those for reducing noise from the blade. The key to balancing these competing requirements is understanding the aeroacoustics of blades within the cutting environment. Until recently, there has been no reliable way to predict the noise generated by a configuration of blades and housing decks. To address this, John Deere and its partners have undertaken a program in Computational Aero-Acoustics (CAA). By leveraging existing computational fluid dynamics and parallel processing capabilities, the participation of several software vendors, and the cooperation of industrial and academic partners, this project has provided a capability to accurately predict the sound generated by low Mach number flows in industrial systems. This presentation reviews the results of applying a nascent CAA capability. The noise generated by a single-blade walk-behind mower was computed and compares successfully with laboratory tests of radiated sound from the mower. Details of the computational process and the results will be presented.

Contributed Papers

9:45

2aNS3. Comparison of the conical beam and image source methods for predicting sound fields in a branching tunnel. Sai Tung So (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong) and Kai Ming Li (Purdue Univ., West Lafayette, IN 47907-2031, mmkmli@purdue.edu)

There are many different models for the numerical analysis of a tunnel with a branch, i.e., a T-intersection. Typical numerical models include the finite-element method, boundary-element method, conical beam, ray-tracing approach, and the image-source method. Among these different models, the image-source method is one of the simplest and requires the least computational time. In this study, the prediction of sound transmission along a branching underground tunnel is computed by using the image-source method. There were published data for the measured sound fields along a branching underground tunnel with an arched ceiling. Numerical predictions of sound transmission by means of the conical-beam method were also presented in this earlier study. By applying the information extracted from the published data, the image-source method was applied to predict the transmission of sound for an identical arrangement, except that the tunnel was assumed to have a rectangular cross section. The reverberation time and rapid speech transmission index of this branching tunnel were also computed by means of the image-source method. It was demonstrated that the predicted results according to the image source method compared reasonably well with the published experimental results.

10:00

2aNS4. Prediction of sound fields in a long enclosure due to a dipole source. Chenly Y. C. Lai (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong) and Kai Ming Li (Purdue Univ., West Lafayette, IN 47907-2031, mmkmli@purdue.edu)

The speech intelligibility in long enclosures is primarily affected by two factors: The reverberation time and the signal to noise ratio. To predict the speech intelligibility, previous studies were mainly focused on predicting either the sound field or the reverberation time in long enclosures. In these predictions, the sound sources were usually assumed to be a monopole source. However, it is well known that many noise sources have distinct directional characteristics. For example, noise radiated from rail-

way vehicles and from ventilating fans is best modeled as a dipole source. This paper presents a numerical model to predict the sound field due to a dipole source in a long enclosure. The numerical model is validated by field measurements. The experimental dipole source was built by using a pair of identical loudspeakers placed at a close distance to each other that generate sound at equal amplitude but opposite in phase. Measurements are conducted at two test sites. Tolerable agreement between numerical predictions and experimental data is obtained.

10:15–10:30 Break

10:30

2aNS5. In-duct acoustic source model considering flow characteristics at orifice and valve inside fluid machinery. Tae-Kyoon Kim and Jeong-Guon Ih (Ctr. for NOVIC, Dept. of Mech. Eng., KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea, thaad@kaist.ac.kr)

Intake and exhaust systems of fluid machinery are generally nonlinear and time-varying due to the flow activities at orifice, periodically moving valve, etc. Because of difficulties in considering these effects in the analysis of the sound propagation, a linear time-invariant source model has been used. Negative source resistance and spectral errors at high frequencies of a linear time-invariant model are due to the neglect of such reality. To investigate the effect of nonlinear and time-varying nature of the in-duct source, flow characteristics through orifice and periodic change of valve area were modeled. A simplified fluid machinery model composed of reservoir, valve, and duct was tested. To obtain a reference acoustic characteristic at the source plane, the method of characteristics (MOC) was employed. Pressure spectrum at the source plane using the present model was compared with those from a linear time-invariant source model and MOC. At the fundamental frequency and low harmonics of valve frequency, the pressure spectrum has little difference in both models, which results in a similar overall sound level. However, at high harmonics, the suggested model showed large discrepancy with the linear time-invariant model, but was similar to the MOC result. This revealed a good possibility of the suggested model for the practical application to the fluid machinery with further extended study. [Work supported by BK21.]

10:45

2aNS6. Simpified theory of wind turbine noise. Evangelia G. Kotsari and Allan D. Pierce (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

The noise is expected to be accounted for by the forces exerted on the air by the blades and pillar; the idealized design is presumed sufficiently optimal that the power is maximum and the generated sound is minimal for that power generation. Betz's law applies, and the steady torque exerted on the air by the turbine has a relation to the net drag force (in the direction of incoming flow) exerted on the turbine blade system. The forces associated with the torque create an acoustic pressure field that is even in the coordinate along the turbine axis. The forces associated with the drag create a field that is odd in this coordinate. These two contributions can constructively reinforce or interfere, so the various pure tone contributions at multiples of the blade passage frequency can have nodes in the polar plot of radiated far-field intensity. The locations of these are explained with the use of Betz's law. At moderate to higher frequencies, interest is typically in finite-frequency bands, which contain a large number of discrete tones. To predict the noise within such bands, sums that appear in the basic Gutin theory are replaced by integrals, and appropriate smoothing techniques are applied.

11:00

2aNS7. Sound field modeling in urban squares using a diffusion equation model. Yun Jing and Ning Xiang (Grad. Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

A diffusion equation model which has been recently applied to room-acoustic prediction is used to model the sound field in urban squares surrounded by diffusively reflecting building facades. A previous diffusion equation model [V. Valeau, et. al., J. Acoust. Soc. Am., 119, 2006, pp. 1504–1513; Y. Jing and N. Xiang, J. Acoust. Soc. Am. 121, 2007, pp. 3284–3287; A. Billon, et. al., Applied Acoustics, 2007, in print] is more suitable for room surfaces with low and medium absorption coefficient, whereas in this case the space is featured with an open top characterized by the absorption coefficient 1.0. Another diffusion equation model [T. Polles, et. al., J. Acoust. Soc. Am., 116, 2004, pp. 2969–2983] can well model the sound field in a space where one specific dimension is larger than the others, like street. However, it has not yet been extended for urban squares. A modified diffusion equation model is proposed for modeling the sound field in urban squares. Both simulations and scale-model experiments are employed to investigate its effectiveness in terms of the reverberation time and sound pressure level distribution.

11:15

2aNS8. Environmental noise measurements using a coded signal. David C. Waddington, Jamie A. Angus, and Yui W. Lam (Acoust. Res. Ctr., School of Computing, Sci. and Eng., Univ. of Salford, Salford M5 4WT, UK)

Application of a coded acoustic signal to the investigation of environmental sound propagation is presented. The problem identified is that of obtaining adequate signal-to-noise ratio in environmental measurements, since the sound power level of the source is often required to be limited so as not to cause a noise disturbance. An established method of increasing signal-to-noise ratio is by means of coded signals such as maximum length sequences. However, this method cannot be applied outdoors because of the time-variant nature of the atmosphere. Consequently, a coded acoustic signal design is proposed for the study of environmental sound propagation, involving the modulation of multiple acoustic carriers by orthogonal correlation sequences. An experiment is described in which the correlation properties of the coded signal are demonstrated under outdoor conditions typical of environmental noise measurements. The results illustrate that a coded signal approach can provide researchers, engineers, and environmental officers with an improved method for investigating the propagation of environmental noise. This raises the possibility that investigations could be performed in the urban environment without disturbance to residents, and propagation could be determined in the presence of high background noise, such as from roads. [Work supported by EPSRC UK.]

11:30

2aNS9. Measurement and evaluation of blade passage frequency fluctuations. Cole V. Duke, Scott D. Sommerfeldt, Kent L. Gee, and Connor R. Duke (Brigham Young Univ., N281-A ESC, 75 W. 960 N. No. 7 Provo, UT 84602)

In the active control of tonal noise from cooling fans, one factor that can limit the achievable attenuation is fluctuation of the blade passage frequency in time. Large fluctuations in a short time can hinder the algorithm from converging to the optimal solution. Some fans have steadier speeds than others, which can be due to unsteady driving mechanisms or the physical structure of the fan. Environmental effects, such as back pressure and unsteady blade loading, can also cause the fan speed to fluctuate. The shifting in the blade passage frequency will be measured using a zero-crossing technique to track the frequency of each cycle. Blade passage frequency fluctuations will be presented for various driving mechanisms and environmental conditions. Techniques to minimize frequency shifting will also be discussed.

11:45

2aNS10. Development of a low cost microphone array system for vehicle noise source identification. Yuri Adson Ribeiro Silva, Willoium Fonseca, and Samir N. Y. Gerges (Lab. of Acoust. and Vib., Mech. Eng. Dept., Federal Univ. of Santa Catarina, Univ. Campus—Trindade, Florianopolis, Santa Catarina, Brazil 88040-900, samir@emc.ufsc.br)

The evolution of data acquisition systems and the resulting cost reduction allow a higher channel count system to be used in the development of noise source identification systems. Unfortunately, the cost of measurement microphones remains impracticable for most applications containing dozens of channels. To overtake this problem, it is possible to use low cost sensors with a duly magnitude and phase response characterization. This article will present the low cost microphone array systems design and evaluation with 32 microphones. The microphones were constructed with electrets cartridges and mounted with an electronic preamplifier circuit. Different shapes for the sensors' placement were computationally simulated for the adequate choice of the array geometry. The signal processing and data acquisition are implemented in a commercial platform. This allows the development of algorithms based on the beam-forming technique to obtain noise maps overlapped to the photographic image of the analyzed system. The goal of this work is to present the design, construction, and qualification of the 32-channel noise source identification system for applications in the automotive industry.

Session 2aPA

Physical Acoustics: General Physical Acoustics

Joel Mobley, Chair
National Center for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677

Contributed Papers

8:30

2aPA1. A fast method of deriving the Kirchhoff formula for moving surfaces. F. Farassat and J. Posey (NASA Langley Res. Ctr., Hampton, VA 23681 feri.farassat@nasa.gov)

The Kirchhoff formula for a moving surface is very useful in many wave propagation problems, particularly in the prediction of noise from rotating machinery. Several publications in the last two decades have presented derivations of the Kirchhoff formula for moving surfaces in both time and frequency domains. The method presented here, originally developed by Farassat and Myers in time domain, is both simple and direct. It is based on generalized function theory and the useful concept of imbedding the problem in the unbounded three-dimensional space. An inhomogeneous wave equation is derived with source terms that involve Dirac delta functions with their supports on the moving data surface. This wave equation is then solved using the simple free space Green's function of the wave equation resulting in the Kirchhoff formula. The algebraic manipulations are minimal and simple. The derivation does not require the Green's theorem in four dimensions and there is no ambiguity in the interpretation of any terms in the final formulas. Furthermore, this method also gives the simplest derivation of the classical Kirchhoff formula which has a fairly lengthy derivation in physics and applied mathematics books. The Farassat-Myers method can be used easily in frequency domain.

8:45

2aPA2. Air turbulence-induced vibration of SCUBA breathing apparatus. Dimitri Donskoy, Len Imas, Timothy Yen, Nikolay Sedunov, and Michael Tsionskiy (Maritime Security Lab., Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030)

The first study of the underlying mechanism of noise generation in SCUBA was reported at the previous 153rd ASA meeting (D. Donskoy, Acoustic Emission Mechanism from Scuba Diving Equipment). It was shown that the primary originating source of SCUBA vibration and subsequent acoustic emission is a combination of non-stationary shocks and separated turbulent flow that occurs within the first stage air pressure regulator during the inhale phase of breathing. The separation mechanism is due to a combination of two factors: Shock-induced separation and geometrical discontinuities in the regulator's internal surface topology. The separated flow manifests itself through the presence of three-dimensional vortices and turbulent pressure fluctuations that excite structural resonances of the regulator and connected air tank, resulting in a complicated resonance dynamic behavior of the system. For the air flow dynamic analysis, we utilized a scale-adaptive detached eddy simulation model, which is based on the solution of unsteady, three-dimensional compressible Navier-Stokes equation solved on a hybrid unstructured computational mesh. The CFD simulations were applied to different topologies representing major types of regulators (piston and diaphragm types) to reveal the complex mechanics of the flow and their dependence on the regulator design. [Work supported by ONR.]

9:00

2aPA3. The effects of frequency and sound pressure level on acoustic agglomeration of flyash. Bart Lipkens (Western New England College, 1215 Wilbraham Rd. Box S-5024, Springfield, MA 01119) and Shaozeng Dong (Virginia Commonwealth Univ., Richmond, VA 23284)

Previously, a new concept of effective agglomeration length, which measures the maximum particle separation distance for effective collisions, was proposed (*J. Aerosol Sc.*, Vol. 37, 540–553, 2006) for acoustic agglomeration of polydisperse aerosols. Two mechanisms are taken into account, orthokinetic collision and acoustic wake in a horizontal acoustic wave. Predictions of the model indicate that orthokinetic collision dominates at low frequencies for intermediate particle size ratios while the acoustic wake effect is more significant at higher frequencies for all particle sizes. Experimental results of acoustic agglomeration of flyash in two resonators, one a constant cross-sectional resonator and the second a variable cross-sectional resonator, are reported. The effects of frequency and sound pressure level (SPL) on acoustic agglomeration are measured. The experimental results are compared with predictions from the effective agglomeration length model. It is confirmed that for orthokinetic agglomeration an optimum frequency exists. However, the optimum frequency is found to be a function of SPL. Resonator design has an effect on the maximum achievable SPL and the power consumption of the device.

9:15

2aPA4. Time-domain observations of fast and slow ultrasonic group velocities in microsphere suspensions. Joel Mobley and Robert Evans Heithaus (Dept. of Phys. and Astron., Univ. of Mississippi, University, MS 38677)

In dispersive media, propagating wave packets can exhibit abnormal group velocities that differ substantially from the phase velocity of the dominant mode comprising the pulse. Recently, both arbitrarily large and negative acoustic group velocities were shown to exist in an aqueous suspension of polymer microspheres, using a broadband measurement technique. In this talk, we report on the direct time-domain observation of abnormal group velocities in microsphere suspensions. These dispersion-related propagation effects include slow group velocities, where the envelope speed is reduced by as much as 25%, relative to the phase velocity, and fast group velocities, which can be arbitrarily large or negative. The role of signal bandwidth in creating pulses in which such group velocities are manifested is discussed. The practical limits of slowing the group velocity in microsphere suspensions is also explored.

9:30

2aPA5. Nonlinear phenomena in acoustics: Traveling waves, bifurcations, and singular surfaces. Pedro M. Jordan (Naval Res. Lab., Code 7181, Stennis Space Ctr., MS 39529, pjordan@nrlssc.navy.mil)

Traveling wave solutions (TWS) are explored in the context of nonlinear acoustics. Exact solutions are given, including one involving the recently introduced Lambert W-function, along with asymptotic and stability results. Poroacoustic propagation under Darcy's and Forchheimer's laws is examined, as well as acoustic phenomena in thermoviscous fluids. Additionally, a connection between discontinuity formation in the TWS and the associated singular surface, which is known as an acceleration wave, is pointed out. Last, if time permits, applications to nonlinear kinematic wave phenomena (e.g., second-sound and traffic flow) are briefly noted. [Work supported by ONR/NRL funding (PE 061153N).]

9:45

2aPA6. Computer simulations of a maximum length sequence modulated photoacoustic spectrometer. Arash Soleimani-Karimabad and Ralph T. Muehleisen (Civil, Architectural, and Environ. Engr., Illinois Inst. of Tech., 3201 S. Dearborn St., Rm. 228, Chicago, IL 60616, muehleisen@iit.edu)

There is a great need for inexpensive, rugged, portable, and versatile chemical detectors for use in both security and environmental measurement, but current technologies that meet that criterion are usually limited in sensitivity and/or are dedicated to detecting a limited number of chemicals. Photoacoustic spectrometers are among the most sensitive instruments for trace analysis in gases, liquids, and on surfaces, and many are able to detect a wide range of different substances. Unfortunately, photoacoustic systems are usually expensive, are usually not portable, and usually cannot be used for both surface and airborne measurements. A new photoacoustic measurement system has been proposed that utilizes maximum length sequence (MLS) modulation of small, rugged, inexpensive laser diodes. The use of MLS modulation eliminates the need for lock-in amplifiers, optical modulators, or resonant measurement cells allowing for portable use on both gas and surfaces. In this paper, results of computer simulations to examine the sensitivity limits and noise immunity of the proposed design are presented.

10:00–10:15 Break

10:15

2aPA7. Validation of laser Doppler anemometer measurements of power dissipation in an arbitrary termination. Ki Won Jung and Anthony A. Atchley (Grad. Program in Acoust., Penn State Univ., University Park, PA 16802, kuj102@psu.edu)

The two-microphone technique is commonly used to measure power dissipation in arbitrary acoustic terminations. This method is well established and provides reliable results when care is taken to avoid problematic microphone configurations (e.g., a pressure node at a microphone location or the microphone spacing being too small in comparison to the wavelength). A laser Doppler anemometer (LDA) technique of measuring power dissipation is discussed in this paper. The LDA method has several disadvantages over microphone methods, including cost of the equipment, the need to use periodic test signals, the need for tracer particles, and the need for a transparent duct. However, rather than being limited to fixed microphone locations and working under the plane wave assumption, the LDA method allows measurement of multiple components of the velocity field at numerous locations with arbitrary spacing in three dimensions. In this paper, the LDA method is validated by measuring the power dissipated in a variable RC load [Fusco and Ward, J. Acoust. Soc. Am. **91**, 2229–2235 (1992)]. The results are also compared to those obtained with the two-microphone technique. [Work supported by the Penn State Graduate Program in Acoustics.]

10:30

2aPA8. The wave behavior in the simple 1 dimensional lattice. Kanghyun Chu and Yang-Hann Kim (Ctr. for Noise and Vib. Control, KAIST, 373-1 Guseong, Yuseong, Daejeon, Republic of Korea, yanghannkim@kaist.ac.kr)

The behavior of an acoustic wave in a periodic lattice was mathematically shown for both infinite and finite cases. For simplicity, the lattice is assumed to be composed of two kinds of medium in ABABABAB order for the infinite case, and BBBABABABABBB order for the finite case. For the infinite lattice, by using the Blochs theorem, the dispersion relation of the wave in the periodic lattice was driven. This gives an acoustic band gap (the frequency region that the wave cannot propagate) and negative group velocity. In addition, for the finite lattice, by using symmetry and the superposition principle, reflection coefficient (between the outer medium and the lattice surface) and transmission coefficient (through the finite lattice) were driven. It was then shown that the result of the finite case agrees to the result from the infinite lattice case, if the number of periods goes to infinity. Finally, possible applications by using what we have learned would be discussed.

10:45

2aPA9. Anisotropic dynamical mass density by two-dimensional arrays of solid cylinders in air. Daniel Torrent and José Sánchez-Dehesa (Polytechnic Univ. of Valencia, Valencia, Spain)

The long wavelength limit (homogenization) of two-dimensional (2-D) periodic arrangements of elastic cylinders in a fluid or a gas (i.e., sonic crystals) is analyzed through the linear part of their band structure. It is found that sonic crystals behave (in the homogenization limit) as effective metamaterials with anisotropic dynamical mass density and fluidlike speed of sound. The special case of rigid cylinders embedded in several 2-D lattices is studied here. It is shown that these systems are suitable to build 2-D anisotropic fluidlike metamaterials. Finally, the reflection and transmission through an anisotropic-elastic medium will be discussed. [Work supported by MEC of Spain and GVA of Valencia.]

11:00

2aPA10. A gradient index sonic lens based on acoustic metamaterials. José Sánchez-Dehesa and Daniel Torrent (Polytechnic Univ. of Valencia,Valencia, Spain)

We report a method to design and characterize broadband gradient index lenses. The lenses are based on two-dimensional sonic crystals, which are periodic arrangements of elastic cylinders embedded in air. It will be shown that gradient index sonic lenses built with rigid cylinders are more powerful than conventionally curved lenses. We also demonstrate that the designing possibilities increased by considering the mixture of two types of cylinders in the sonic crystal. So, a gradient index lens with zero reflectance at the surface is proposed by employing a mixture of aerogel and rigid cylinders. Finally, a multiple scattering theory has been employed to compare the performance of conventionally curved lenses and gradient index lenses. [Work supported by MEC of Spain and GVA of Valencia.]

11:15

2aPA11. Nonlinear ultrasound beams and Eckart streaming using interdigital transducers. Tomoo Kamakura, Miyuki Maezawa, and Rui Kamata (Univ. of Electro-Commun., 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan, kamakura@ee.uec.ac.jp)

The Inter-digital transducer (IDT) of X-Y $LiNbO_3$ 128 deg substrate has a rectangular aperture 10 mm$\times$2 mm, and has a center frequency of 20 MHz. The transducer radiates ultrasound beams in two different directions at the Rayleigh angle, that is determined by the propagation speeds of the surface wave on the substrate, and of the longitudinal wave in water, the angle being $\pm$22 deg to the plane normal to the aperture and containing the IDT center electrode. The fundamental and second harmonic sound pressure levels are measured along and across the beam, using a hydrophone whose active element is 0.4 mm in diam and whose frequency response is calibrated up to 40 MHz. The KZK equation is utilized to predict theoretically sound pressure amplitudes [T. Kamakura et al., Acustica **80**, 332–333 (1994)]. Especially, the fundamental pressure curves are compared with the Rayleigh integral formula. It has been found that the KZK equation is almost useful. The Eckart streaming in the ultrasound beam is also observed using PIV.

Session 2aPP

Psychological and Physiological Acoustics: Potpourri

Glenis R. Long, Chair
CUNY Graduate Center, Speech and Hearing Science Program, 365 Fifth Ave., New York, NY 10016

Chair's Introduction—8:30

Contributed Papers

8:35

2aPP1. Probing the temporal resolution and bandwidth of human hearing. Milind N. Kunchur (Univ. of South Carolina, 712 Main St., PSC 404, Columbia, SC 29208)

Two types of experiments were conducted to assess the human discriminability of temporal convolution—one employed lowpass filtering and the other employed delays due to spatial misalignment. By using special ultrahigh-fidelity equipment, both experiments demonstrated discernment at a $\sim$5 μs timescale, which is much shorter than found previously. While the signal manipulations affect both the spectrum and temporal definition of the signal, the spectral changes in these experiments fall below the known just noticeable differences. The discrimination may therefore involve mechanisms additional to the auditory system's ability to distinguish spectral amplitude differences. Furthemore the present work shows that typical instrumentation used in psychoacoustic research has insufficient temporal speed and fidelity for properly assessing all aspects of human hearing and also that digital sampling rates used in consumer audio are insufficient for fully preserving transparency. [This work was partially supported by the University of South Carolina Office of Research and Health Sciences Research Funding Program.]

8:50

2aPP2. Estimates of loudness enhancement and induced loudness reduction at short masker-target intervals. Daniel Oberfeld (Dept. of Psych., Johannes Gutenberg-Universitaet, 55099 Mainz, Germany, oberfeld@uni-mainz.de)

Recent data by Oberfeld [J. Acoust. Soc. Am. 121, 2137–2148 (2007)] supported the hypothesis [Arieh and Marks, J. Acoust. Soc. Am. 114, 1550–1556 (2003)] that a forward-masker causes two processes influencing the loudness of a temporally proximal target tone. One process results in a shift in target loudness towards masker loudness (e.g., loudness enhancement) and decays within 400 ms. Induced loudness reduction (ILR) caused by the second process remains constant for several seconds. According to this model, an estimate of the effect of the faster-decaying process is the change in loudness level at a short masker-target interval (both processes are in effect), minus the negative change in loudness level at a long interval (only ILR is effective). At a 100-ms interval, loudness matches between a 2500-Hz target and a 500-Hz comparison showed that 2500-Hz maskers more than 10 dB higher in level than a 35-dB SPL target caused an increase in loudness. For a 60-dB SPL target, a loudness increase was found at a masker-target level difference of 30 dB, but not at smaller differences. At the 1000-ms interval, ILR was consistently observed. The estimated amount of loudness enhancement is compared to data from experiments using a same-frequency matching task.

9:05

2aPP3. Unmasking the components of a complex tone with stretched harmonics. William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 1226 BPS Bldg., East Lansing, MI 48824)

In the unmasking paradigm [W.M. Hartmann and M.J. Goupell, J. Acoust. Soc. Am. **120**, 2142–2157 (2006)], removing harmonic n of a periodic complex tone causes harmonic $n+1$ to be suddenly audible, though sometimes $n-1$ is heard instead. The unmasking experiment was repeated with harmonically stretched tones, where the frequency of the nth ''harmonic'' was proportional to the 1.32 power of n (octave ratio of 2.5 to 1). After a six-interval sequence with component n alternately on and off ($1<n<21$), listeners matched the pitch of the unmasked component by adjusting the frequency of a sine tone. Pitch matching results were comparable to results for harmonic tones, showing regions of: (1) Accurate matches; (2) reproducible pitch shifts; and (3) inconsistent matches as the removed component frequency increased. The agreement in results proves that unmasking does not depend on harmonicity. It also indicates that pitch matches were not biased by the listener's internal template for the harmonic series in any important way. Because of the stretch in component frequencies, listeners were able to make reliable matches at higher frequencies, as high as 3,500 Hz, demonstrating the importance of frequency resolution in the unmasking effect. [Work supported by the NIDCD Grant No. DC00181.]

9:20

2aPP4. Context effects in the lateralization of sine tones? William M. Hartmann and Erin M. Nolan (Dept. of Phys. and Astron., Michigan State Univ., East Lansing, MI 48824)

Low-frequency sine tones can have large interaural time differences, but high-frequency tones cannot. If listeners estimate the lateral positions of tones of different frequency, using a fixed lateralization response scale, the estimates may depend on the experimental blocking of the tones. A

context where all frequencies appear equally often (mixed) serves as a standard. It was hypothesized that a context with tones blocked on frequency would lead to abnormally large lateral position estimates for high-frequency tones because of a tendency to use the entire response scale in any stimulus-range context. Four listeners estimated the lateral positions of sine tones with frequencies 200, 500, 750, and 1,000 Hz having all possible interaural phase differences (IPD) spaced by 30 deg. Fixed-frequency blocks were done first, then mixed. Very little trace of the hypothesized effect was observed, either in the overall range of responses or in the slope of the response-IPD function for small IPD values. The same listeners matched the laterality of the same tones using the interaural level difference (ILD) of a narrow noise band. Relative slopes of the ILD-IPD functions agreed with slopes of the lateral position estimates. [Work supported by the NIDCD.]

9:35

2aPP5. Salient phase density model for the lateralization of binaural pitches. Peter Xinya Zhang (Dept. of Audio Arts & Acoust., Columbia Coll. Chicago, 33 E. Congress Pkwy, Chicago, IL 60605; Dept. of Phys. & Astron., Michigan State) and William M. Hartmann (Michigan State Univ., East Lansing, MI 48824)

Binaural pitches are sensations of pitch generated from a broadband noise having a narrowband boundary region where the interaural phase difference varies rapidly as a function of frequency. These pitch sensations are perceived to be lateralized in the head. A binaural model, the *Salient Phase Density model*, has been derived to predict the lateralizations. The model begins by separating the boundary from the background. Because the boundary is narrow, the boundary components can be represented statistically by a density. Components of the boundary with phases similar to the background are discounted in the density as *not salient*. The density is then used to compute the cross-correlation function of the boundary region. The predicted lateralization is based on the interaural time difference of a template that best matches the cross-correlation of the boundary. The model was tested against psychoacoustical experiments using four types of Huggins pitch boundary-linear phase, stepped phase, uncorrelated phase, and π-shifted linear phase. The model was further tested against other binaural pitches: Binaural edge pitch and binaural coherence edge pitch. The experiments supported the model. [Work supported by the NIDCD, Grant No. DC00181.]

9:50

2aPP6. Anomalyous interaural coherence minima near 45 degrees. Neil L. Aaronson and Eric Wolf (1300 BPS, Dept. of Phys. & Astron., Michigan State Univ., East Lansing, MI, 48824, aaronson@pa.msu.edu)

High-resolution (200 kHz sample rate) measurements of interaural coherence for a KEMAR manikin were made in an anechoic environment. As the angle of incident sound was varied in the horizontal plane, there was a minimum in coherence around $\pm 45°$. These coherence minima occured at low frequencies, up to about 3000 Hz, and decreased in depth for increasing frequency, f, at a rate approximately proportional to $1/f$. In a 1/3-octave band centered at 500 Hz, the interaural coherence was approximately 0.7 for angles of $\pm 45°$. These results were confirmed by measurements in KEMAR using probe microphones. Measurements made using probe microphones with human listeners at angles of $\pm 45°$ and $0°$ showed a similar pattern in coherences, but the minima were far more shallow. In the 500 Hz band, a coherence of, on average, 0.94 was measured in listeners facing $\pm 45°$, compared to an average coherence of 0.99 in that band at $0°$. Headphone experiments using measured values of coherence showed that the deeper measured coherence minima lead to a deficit in sound localization. [Work supported by the NIDCD grant DC00181.]

10:05–10:20 Break

10:20

2aPP7. Lateralization of linear and logarithmic frequency-modulated (FM) sweeps at high frequencies. Kourosh Saberi and I-Hui Hsieh (Dept. of Cognit. Sci., Univ. of California, Irvine, CA 92697, saberi@uci.edu)

The ability to lateralize linear and logarithmic frequency-modulated (FM) sweeps at high frequencies was investigated using a 2IFC adaptive block design. The FM stimulus swept from 3 kHz to 8 kHz with the onset frequency roved by 15% on each presentation. The cue to be detected was an interaural delay in the waveform's fine-structure (i.e., delay in instantaneous frequency) that varied on each trial according to a 2-down 1-up adaptive rule. The waveforms to the left and right channels had simultaneous onset/offset envelopes. The stimulus duration (0.5, 1, 5, 10, 25, or 50 ms) and thus the sweep rate, as well as the sweep direction (up or down), were additional parameters of the study. Thresholds were higher than expected from studies of sinusoidal FM and ranged from approximately 100 μs for the 0.5 ms sweep to near chance for durations exceeding 10 ms. Sweep direction and type (log versus linear) did not substantially affect thresholds. Implications of these findings for models of binaural interaction at high frequencies and within-channel mechanisms of FM-to-AM transduction will be discussed. [Work supported by NSF.]

10:35

2aPP8. Neuroimaging correlates of spatial cues in speech and noise stimuli. Kourosh Saberi, Kevin R. Smith, I-Hui Hsieh, and Gregory Hickok (Dept. of Cognit. Sci. and the Ctr. for Cognit. Neurosci., Univ. of California, Irvine, CA 92697, saberi@uci.edu)

Three fMRI experiments are described that investigated neural activity patterns in the human auditory cortex in response to motion and multi-source cues in speech and noise stimuli. In the first study, ten subjects were scanned while listening to either moving or stationary AM-modulated noise. We found that motion stimuli activated STG/planum temporale (bilaterally) and right parietal lobe, previously implicated in motion processing. Stationary stimuli, whose perceived lateral position varied across presentations, activated these regions equally well. In the second study, we used an event-related fMRI-adaptation paradigm in which pairs of stimuli, consisting of permutations of sequentially presented motion/stationary sounds (m/m, m/s, s/m, s/s), were used during scanning of ten subjects. Consistent with study 1, we observed no motion-selective activation in the planum or parietal regions. In the third study, we examined cortical activity patterns in response to multiple speech streams filtered through individualized HRTFs originating from either a single or multiple locations. Preliminary results show a greater level of cortical activation in ROIs in response to a single speaker, contrasted to multiple speakers. [Work supported by NSF and NIH.]

10:50

2aPP9. Sound scattering from two concentric fluid spheres. Jared McNew and William D. O'Brien, Jr. (Bioacoustics Res. Lab, Dept. of Elec. and Comput. Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801)

Recently, the finite-element method has been used to analyze the paths through which sound enters the human head. In this work, the solution to the problem of scattering by two concentric fluid spheres is derived in order to verify this model. The effect of differences in sound speed and density is taken into consideration while dissipation is neglected. The concentric sphere solution is then used as a simplified model of the human head where the outer sphere has the properties of bone and the inner sphere has the properties of brain. A Fourier cosine series is used to approximate a pulse propagating into the concentric sphere model. A correlation-based wavefront reconstruction is then performed, followed by ray tracing to visualize the results. [Work supported by AFOSR FA9550-06-0128.]

11:05

2aPP10. New procedure for measuring stimulus frequency otoacoustic emissions. Glenis R. Long (Speech and Hearing Program, Grad. Ctr., CUNY, 365 Fifth Ave., New York, NY 10016) and Carrick L. Talmadge (Univ. of Mississippi, University, MS 38677)

Stimulus frequency otoacoustic emissions (SFOAE) are generated by single tones. This is both an advantage and a disadvantage. The advantage is that they are more likely to be primarily generated from a more limited region of the cochlea than other types of OAE. The disadvantage is that when tone pulses are used, they occur at the same frequency and time as the stimulus, making it very difficult to separate the OAE from the signal, unless one uses suppression. The SFOAE is generated in the cochlea and has to travel along the basilar membrane before being measured in the ear canal. Cochlear travel times mean that the SFOAE is always delayed relative to the stimulus. The delay is frequency dependent. We have modified the continuously sweeping tone procedure developed in our lab for evaluating DPOAE to separate the stimulus and SFOAE. The LSF procedure was modified to permit dynamic fitting of the delay of a given signal component permitting the extraction of the time varying SFOAE.

11:20

2aPP11. Orthotropic material properties of the gerbil basilar membrane. Shuangqin Liu and Robert D. White (Dept. of Mech. Eng., Tufts Univ., Medford, MA 02155)

Material properties of the cochlear partition play an important role in understanding the cochlear hydrodynamics. In this paper, two sets of experimental results are analyzed to extract the orthotropic material properties of the gerbil cochlear partition at different longitudinal locations. In order to accomplish this, a geometrically nonlinear composite orthotropic plate model is employed. The model is used to predict both out-of-plane and in-plane motion of the partition under a static finite area distributed load. Both in-plane and out-of-plane motion are needed for comparison with recent experimental results. It is shown that the spatial decay rate (the space constant) for the in-plane deflection is different from the out-of-plane deflection, which has a significant effect on the derived partition properties. The size of the probe tip is shown to have little influence on the results. Results are presented for two types of boundary conditions. Orthotropy ratios (the ratio of the two elastic moduli) determined from the experimental data are found to vary with longitudinal position and choice of boundary conditions. Orthotropy ratios vary from 65 close to the base to 10 in the upper middle turn of the cochlea.

WEDNESDAY MORNING, 28 NOVEMBER 2007 NAPOLEON C3, 8:00 TO 10:05 A.M.

Session 2aSCa

Speech Communication: Role of Attention in Speech Perception

Alexander L. Francis, Chair
Purdue Univ., Speech Language and Hearing Science, Heavilon Hall, West Lafayette, IN 47907-1353

Chair's Introduction—8:00

Invited Papers

8:05

2aSCa1. What is attention, and what is it doing in speech perception? Alexander L. Francis (Speech, Lang. & Hear. Sci., Purdue Univ., West Lafayette, IN 47907)

The word attention is beginning to appear more frequently in speech research, but closer examination suggests that this term can refer to a variety of potentially unrelated phenomena. The goal of this talk is to begin a discussion of these various phenomena, in the form of a preliminary, and subjective, overview of some areas of speech perception research in which ''attention'' is starting to play a more prominent role, including experimentally and contextually determined changes in cue weighting, foreign language acquisition, speech perception by elderly and/or hearing-impaired listeners and cochlear implant users, normal and disordered language development, perception of spoken discourse, and multimodal speech perception. I will also discuss some areas of basic attention research in which speech research may be able to make a unique contribution, including theories of the nature and role of auditory ''objects'' in spatial and object-based attention, load-based theories, and research on the interaction of attention and long- and short-term (working) memory and perceptual learning. [Work supported by NIH-NIDCD R03DC006811.]

8:25

2aSCa2. Mechanisms of auditory selective attention. Robert Melara (City College, City Univ. of New York, 138th St. and Convent Ave., New York, NY 10031, rmelara@ccny.cuny.edu)

Recent research on auditory attention addresses three main questions: (1) How early in the processing stream does attention influence the analysis of auditory signals?; (2) are the distinct features of an auditory object processed together or separately?; and (3) how does the attentional system enable task-relevant information to be analyzed separately from task-irrelevant information? The three questions are answered here using behavioral and electrophysiological evidence. First, psychophysical evidence suggests that attention operates early, modulating auditory filters (critical bands). Yet, little physiological evidence has been found for attentional effects on a peripheral auditory system. Such evidence suggests instead that the effects of attention are confined mainly to cortical regions outside the primary auditory cortex. Second, information theory and signal-detection models suggest that auditory dimensions are integral, with features quickly bound into auditory objects. However, physiological evidence reveals that separate auditory features

are processed in parallel and exhaustively from the earliest stages. Finally, electrophysiological evidence indicates that auditory selection involves a duality of processes: Excitation of task-relevant features and inhibition of task-irrelevant features. The magnitude of selection failure is mainly determined by the degree of psychophysical change in distractors (tectonic theory), rather than their physical similarity to targets (biased competition and attentional-trace theory).

8:45

2aSCa3. Selective attention to time scales and time points during speech perception. Lisa D. Sanders (Dept. of Psych., Univ. of Massachusetts-Amherst, Tobin Hall, 135 Hicks Way, Amherst MA 01003, lsanders@psych.umass.edu)

Selective attention provides a mechanism by which people preferentially process subsets of stimuli when faced with overwhelming amounts of information. Spatially selective attention is particularly important for perceiving complex scenes in which multiple objects are presented simultaneously at distinct locations. Recent evidence indicates that temporally selective attention plays an important role in perception when more information than can be processed in detail is presented rapidly. Specifically, listeners can direct selective attention to both different levels of hierarchically organized rhythms and to distinct time points. Doing so modulates auditory evoked potentials at early perceptual stages. Additional studies show that directing temporally selective attention to hierarchical levels and specific time points affects perception of the rapidly changing acoustic features that constitute speech. Language experience and proficiency affect the skill with which adult listeners select the most relevant features and differentially process attended and unattended information in speech streams. Furthermore, developmental deficits in selective attention have been implicated in specific language impairment. These deficits can be reduced with interventions designed to improve speech perception and are accompanied by increases in receptive language scores. [Work supported by NIDCD grants R03DC008684 and F32DC005291 to LDS, R01DC005660 to David Poeppel, and R01DC00128 and R01DC00481 to Helen Neville.]

9:05

2aSCa4. Attending to what is said or who said it. Kimberly Fenn and Howard Nusbaum (The Univ. of Chicago, Dept. of Psych., 5848 S. University Ave., Chicago, IL 60637)

Research on selective attention in speech perception has examined how listeners process source information or linguistic information. Much of this work has focused on specifying the boundary conditions under which perceptual processing is limited, and how attention to talker and linguistic information can compete for limited processing resources. In general, listeners can only attend to a single stream of speech, and changing the talker increases demands on attention. The studies that support this claim usually use target detection or speeded classification tasks, activities we seldom engage in during normal conversation which is the protypical form of spoken language use. Is attention to speech as limited in real conversations as it is in stylized laboratory tasks? In a phone survey paradigm, participants engaged in naturalistic conversation rarely detect changes in their interlocutor. Furthermore, when participants expect that a change in interlocutor will occur, they largely believe that they have spoken to two different people even when the same person conducted the entire experiment. During a conversation, attention is directed to speech consistent with listening goals, and perception of a talkers voice is not an automatic consequence of attending the linguistic message.

9:25

2aSCa5. Selective perception, perceptual modes, and automaticity in first- and second-language speech processing. Winifred Strange (City Univ. of New York, Grad. School and Univ. Ctr., 365 Fifth Ave., New York, NY 10016-4309)

A common phenomenon reported by experienced, late learners of a second language (L2) is that comprehending L2 speech, especially under non-optimal conditions (e.g., in noisy rooms, when driving, when the speaker talks rapidly) is more effortful than processing L1 input. A variety of paradigms have documented this phenomenon experimentally. Many current theories of L1 and L2 speech perception invoke concepts of learned (language-specific) patterns of selective attention or attunement to characterize the processes by which native speakers rapidly and efficiently extract sufficient phonetic information from complex acoustic signals in order to recover phonological sequences (words or word-forms). It is suggested here that there are two modes of language-specific speech processing available to adult listeners, a phonetic mode (requiring attentional resources) and a phonological mode (automatic), that are tapped in the laboratory to different degrees as a function of complex interactions of subject's linguistic experience, stimulus characteristics, and task demands. Exemplary experiments of L1 and L2 listener's perception, using perceptual assimilation (cross-language identification) and speeded discrimination tasks, as well as electrophysiological indices of discrimination, illustrate some of these interactions, within the framework of the automatic selective perception model of speech perception. [Work supported by NIH-NIDCD and NSF.]

9:45

2aSCa6. Determinants of auditory cue weighting and selective attention. Lori L. Holt (Carnegie Mellon Univ., Dept. of Psych., 5000 Forbes Ave., Pittsburgh, PA 15213) and Andrew J. Lotto (Univ. of Arizona, Speech, Lang. & Hearing Sci., 1131 E. 2nd St., Tucson, AZ 85721-0071)

The categorization of speech sounds into phonetic units has been characterized as involving the differential weighting of perceptual cues. Cue weighting has been often explicitly or implicitly associated with the construct of selective attention. We will present data from speech and non-speech categorization studies that attempt to illuminate the factors that influence relative weighting of perceptual dimensions. Of particular interest is the role of variability across the different auditory dimensions. The results of these experiments will be related to work on cue weighting in other psychophysical tasks and the usefulness of the attention construct will be explored. [Work supported by NSF BCS0345773].

Session 2aSCb

Speech Communication: Speech Perception Issues (Poster Session)

Alexander L. Francis, Chair
Purdue Univ., Dept. of Speech Language and Hearing Science, Heavilon Hall, West Lafayette, IN 47907-1353

Contributed Papers

All posters will be on display from 10:30 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 10:30 a.m. to 11:15 a.m. and contributors of even-numbered papers will be at their posters from 11:15 a.m. to 12:00 noon.

2aSCb1. Levels of selective attention in speech perception. Alexander L. Francis (Speech, Lang. & Hear. Sci., Purdue Univ., W. Lafayette, IN 47907)

Listeners' ability to attend to a specific talker while ignoring simultaneous speech from another talker (the ''cocktail party'' problem) is not yet well understood. Lavie's load theory [Trends Cogn. Sci. **9**(2), 2005], originally developed in the visual modality, conceptualizes selective attention as a limited-capacity resource and provides a method for investigating levels of attentional operation. Selection may arise at a cognitive level, when intruding irrelevant information is explicitly inhibited, or at a perceptual level when limited capacity is overwhelmed and task-irrelevant information cannot be fully processed. Under high cognitive load, insufficient processing capacity is available to reject intruding distractors, and interference increases. Under high perceptual load, irrelevant stimuli intrude less because fewer processing resources are concentrated on them, decreasing interference. The present study evaluates the predictions of this theory for speech perception. Listeners identified words produced by a central talker while ignoring a peripheral one. Perceptual load was manipulated by presenting a concurrent (non-speech) response cue, with response contingent upon a single feature of the cue (low load) or a conjunction of features (high load). Cognitive load was manipulated using a concurrent digit pre-load task with one (low load) or six digits (high load). [Work supported by NIDCD No. R03DC006811.]

2aSCb2. Phonetic cue use and integrality. Grant McGuire (Univ. of California at Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, grantmcguire@berkeley.edu)

Phonetic categories are generally assumed to have multiple cues available, which listeners can reliably use for identification. Moreover, there is considerable evidence that such cues are perceived integrally as a unit [Best et al. (1981), Kemler Nelson (1993)]. This paper presents results from two experiments exploring cue perception by comparing Mandarin and American English listeners' perception of the Polish alveopalatal ~ retroflex sibilant contrast. This contrast is native to Mandarin listeners [Stevens et al. (2004)] and is difficult for English listeners to discriminate [Lisker (2001)]. In the first experiment, listeners discriminated pairs of syllables having fully correlated cues, pairs differing in a single dimension (fricative noise, formant transition), or fully conflicting cue pairs. Results demonstrate that only Mandarin listeners are negatively affected by conflicting cue information. In a second experiment, the same listeners labeled a two-dimensional continuum varying in fricative noise and formant transition cues. Mandarin listeners reliably used both cues and showed an effect of integration. English listeners relied solely on vocalic information and ignored fricative variation. Together these results provide more evidence that cue use is language specific and suggests that listeners do not automatically generalize cue information used in other contrasts to a new context.

2aSCb3. The behavior of affricates under restricted lengthening conditions. Anne Pycha (UC Berkeley, 2724 Prince St, Berkeley, CA 94705)

This study uses affricates as a test case for lengthening processes in speech. In general, consonants can lengthen as the result of phonological factors, such as a geminating suffix, or as the result of phonetic factors such as position within a phrase (Keating et al. 2003, Byrd et al. 2000, 2005). In Hungarian and many other languages, however, the presence of an additional consonant restricts phonological lengthening because CCC clusters containing geminates are ill-formed (Kenesei et al. 1998). Acoustic duration measurements were used to examine whether this restriction also holds for phonetic lengthening. Initial results from three male Hungarian speakers indicate that it does not. Consonants in phrase-final position are significantly longer than their counterparts in phrase-medial position, whether they stand alone or in a cluster. A focus on lengthened affricates, however, reveals that the ''shape'' of lengthening differs in these two conditions. Affricates in the stand-alone condition lengthen in a lopsided fashion, showing increases primarily in frication duration; those in the cluster condition lengthen more symmetrically, with increases in both stop closure and frication durations. These findings provide new support for the idea that affricates possess different structures at different levels of analysis (Lombardi 1990, Clements 1999).

2aSCb4. Intelligibility of information in temporally desynchronized bands of speech. Christian E. Stilp, Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53706), Joshua M. Alexander (Boys Town Natl. Res. Hospital, Omaha, NE 68131), and Michael Kiefte (Dalhousie Univ., Halifax, Canada, NS, B3H 1R2)

[Stilp et al., [J. Acoust. Soc. Am. **121**, 3195 (2007)] demonstrated that acoustic measures of potential information, absent explicit linguistic information, are closely related to intelligibility of locally time-reversed sentences. Here, this finding is replicated in two experiments in which synchronization of speech energy within four spectral bands was perturbed in different ways. An ensemble of seven-syllable sentences was synthesized at three different speaking rates (2.5, 5.0, 10 syllables/s) and filtered by one-third-octave Butterworth filters centered at 335, 850, 2135, and 5400 Hz. In Experiment 1, onset of the 335- and 5,400-Hz bands was delayed by multiple durations (0, 25, 50, 100, 200, 400, 600 ms). Across speaking rates, intelligibility was a nonmonotonic function of delay with local minima corresponding to roughly one syllable. Smaller and larger delays yielded better performance. In Experiment 2, bands were nonlinearly scaled in time (jittered) to preserve overall duration with local rate adjusted to preserve the same overall degree of desynchronization as Experiment 1. Performance monotonically decreased with increasing jitter. Overall, results closely mimic previous findings for locally time-reserved

sentences, and provide support that physical acoustic potential information closely relates to intelligibility of temporally distorted speech. [Work supported by NIDCD.]

2aSCb5. Acoustic measures of sound change in language obsolescence. Charles B. Chang (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, cbchang@berkeley.edu)

Following recent acoustic investigations of obsolescing languages such as Babel (2007), I present an acoustic study of phonetic change in Southeastern Pomo (Northern Hokan, Pomoan) based upon recordings of two generations of speakers. This paper focuses on the realization of two phonological contrasts: A velar/post-velar contrast and a dental/alveolar contrast. First, acoustic measurements show that both contrasts have undergone phonetic change convergent with English, the dominant contact language, which contains neither contrast. The contrast between velars and post-velars is considerably diminished in Generation 2 with respect to Generation 1, as seen in the significantly smaller difference between the second formant resonances of the velar and post-velar fricatives. The dental/alveolar contrast undergoes a similar change: The difference between the burst peak frequencies of the dental and alveolar ejectives and the difference between the third formant resonances of the following vowels both shrink. However, the data also indicate that although the acoustic distance between the members of each contrast decreases, the contrast between the two is still maintained in one or more of the above acoustic dimensions. Thus, this study provides a demonstration of how acoustic data can be used to illuminate the status of a sound change in progress.

2aSCb6. Perception of phonemically-identical true prefixes and pseudo-prefixes in real words heard in noise. Rachel Baker, Sarah Hawkins (Dept. of Linguist., Univ. of Cambridge, Sidgwick Ave., Cambridge CB3 9DA, U.K., rb373@cam.ac.uk), and Rachel Smith (Univ. of Glasgow, 12 University Gardens, Glasgow G12 8QQ, U.K.)

This experiment assessed the perceptual salience of acoustic-phonetic differences measured in the first syllables of words with initial syllables that differ in morphological, but not phonemic structure [Baker et al., Proc. 16th ICPhS (2007)]. Words began with true prefixes (Pr: e.g., *mistimes, displease*) or pseudo-prefixes (Ps: e.g., *mistakes, displays*). The fourth phoneme was a stop. A woman recorded sentences, identical except for the critical word, e.g., *I think she distrusts/destroyed them*, in scripted dialogs for naturalness. Critical syllables (mis-, dis-), up to but excluding the stop burst, were cross-spliced either into the ''wrong'' context (mismatched conditions, PrPsPr, PsPrPs) or into another instance of the same sentence (controls, PrPrPr, PsPsPs). In a counterbalanced design, four groups of ten listeners each heard eight mismatched and eight control sentences in cafeteria noise (SNR 2 dB), and wrote what they heard. Errors were greater for critical words in mismatched than control sentences, though the difference only achieved significance for dis- words. The data suggest that perceptual disruptions reflect the difference between the spectrotemporal pattern heard and that expected on the basis of the critical syllable's morphological status (true- or pseudo-prefix), the sentence prosody, and possibly the statistical distribution of the critical syllable's duration.

2aSCb7. Determining and interpreting acoustic landmark sequences in American English. Chiyoun Park (Speech Comm. Group, RLE, MIT, Cambridge, MA 02139, goodie@mit.edu), Nancy Chen, and Youngsook Jung (Harvard-MIT Div. of Health Sci. and Tech., Cambridge, MA 02139)

Acoustically-significant time points (i.e., landmarks) can be used in knowledge-based speech recognition. Consonant landmarks represent three types of speech production processes (glottal vibration, sonorant, and obstruent consonant [K.N. Stevens, J. Acoust. Soc. Am. **111**, 1872–1891 (2002)]) and are classed as ''onset'' and ''offset'' subtypes. A useful characteristic of landmarks is that their sequential combinatorial properties are limited: Given physiological constraints, only 44% of potential landmark pairs are legal. In this study, bigrams are used to quantitatively model the grammar of consonant landmarks. This work expands Liu's method [S.A. Liu, J. Acoust. Soc. Am. **100**, 3417–3430 (1996)], which detects the spectral abruptness of the speech signal, to a probabilistic framework, and then bigrams are applied to determine the most likely landmark sequences. Results tested on TIMIT are consistent across gender and several dialects of American English. Though lower thresholds than Liu's are used to allow more insertions in detecting landmark candidates, the insertion errors are reduced by 46% due to bigram constraints. Many of the errors are systematic, reflecting variants such as flaps or syllabic nasals; they can be used to improve detection methods, and help us further understand acoustic variation in speech. [Work supported by NIH/NIDCD DC02978, DC00075, and T32DC00038.]

2aSCb8. Investigating the consonant-vowel boundary: Perceptual contributions to sentence intelligibility. Daniel Fogerty and Diane Kewley-Port (Dept. of Speech and Hearing, Indiana Univ., Bloomington, IN 47405, dfogerty@indiana.edu)

Speech sounds divide into two broad categories: vowels and consonants. Vowels have a distinct perceptual advantage over consonants in determining sentence intelligibility [Kewley-Port, Burkle, and Lee, J. Acoust. Soc. Am. (in press)]. However, specifying the segmentation of consonants and vowels is problematic because of their extensive acoustic overlap due to coarticulation. The current study used TIMIT sentences to investigate perceptual contributions of consonants and vowels across systematic changes in the consonant-vowel (C-V) boundary. Sentences were presented to listeners with either consonant or vowel information replaced by noise. The C-V boundary was shifted by six specific proportions of the vowel, such that consonant duration increased while vowel duration decreased, yielding 12 different noise replacement conditions. The percent of words listeners repeated correctly was scored. A two-to-one vowel advantage for intelligibility at the traditional C-V boundary was confirmed. Initial results from ten listeners suggest that functions of the shifted C-V boundaries for consonants and vowels differentially contribute to intelligibility. Preliminary comparison of these functions suggests that both consonants and vowels contribute equally (50%) to intelligibility at about a 20% proportional boundary shift. Results will be interpreted in terms of phonological versus acoustic accounts of speech perception. (Supported by NIH)

2aSCb9. Perception of word boundaries in English speech by Japanese L2 learners of English. Kikuyo Ito and Winifred Strange (Ph.D. Program in Speech and Hearing Sci., City Univ. of New York-Grad. Ctr., 365 5th Ave, New York, NY 10016, kito@gc.cuny.edu)

This study investigated the perception of aspiration and glottal stop phonetic cues for word juncture in English phrases by Japanese L2 learners of English, replicating a study of Spanish L2 learners [E. P. Altenberg, Second Language Research. 21, 325–358 (2005)]. Thirty Japanese and 10 American listeners responded to 84 phrases in randomly ordered lists containing 42 contrasting pairs, e.g., *keeps talking* versus *keep stalking* (aspiration-cued contrasts) and *a nice man* versus *an ice man* (glottal stop-cued contrasts) in a two-choice task with confidence ratings. Results indicated that aspiration cues were less available to Japanese listeners (mean=73% correct) than glottal stop cues (mean=91% correct); Americans performed near ceiling on both types of contrasts (96% and 97% correct). Japanese listeners had the most difficulty perceiving aspiration-cued contrasts containing consonant clusters at word junctures, e.g., *keeps canning* versus *keep scanning, tops crawled* versus *top scrawled*. JapaneseEconfidence ratings were high even when they erred, while those of native listeners better reflected response accuracy. The results support the notion that L1/L2 differences in both phonetic and phonotactic structure significantly affects perception of L2 word boundaries.

2aSCb10. Effects of syllable-final duration variation on the identification of synthetic /ba/-/wa/ and /bas/-/was/ speech continua by humans and birds. Thomas E. Welch, Micheal L. Dent, and James R. Sawusch (Dept. of Psych., 206 Park Hall, Univ. at Buffalo, the State Univ. of New York, Buffalo, New York 14260, tewelch2@buffalo.edu)

In an attempt to test whether experience with or knowledge of language is necessary to show typical human only duration-based tendencies in speech perception tasks, budgerigars (Melopsittacus undulatus) and zebra finches (Taeniopygia guttata) were trained to categorize the endpoint stimuli of four synthetic continua consisting of CV and CVC stimuli, with both short and long syllable-final durations. Humans and birds were then tested on the exact same sets of full ten-step continua and their performance was compared in terms of their identification responses and of phoneme boundaries along the /b/-/w/ stop-glide continua. Operant conditioning was used to train the birds to respond differentially to /b/ and /w/ stimuli. Overall, zebra finches, who unlike budgerigars, cannot mimic human speech, were poor at categorizing the human speech tokens and very few could even categorize two endpoints above criterion levels to be moved onto testing. The birds that were able to learn the task showed similar identification results to humans for the continua. The implications of these results and comparisons of the effects of phoneme duration (as in human speaking rate) between humans and birds will be discussed. For Speech Communication Best Student Paper Award Best ''Outstanding Paper By A Young Presenter'' Awards.

2aSCb11. Characteristics of /h/ in American English: Effects of speaker and phonetic context. Laura Koenig (Haskins Labs, 300 George St., New Haven, CT 06511)

This work explores the factors associated with variation in /h/ realization in American English, with particular focus on the presence or absence of phonation. Traditional phonetic accounts have suggested that voiced /h/ occurs most commonly in unstressed, intervocalic positions, but limited quantitative data exist. In an earlier study, we assessed /h/ production in the Texas Instruments/MIT [TIMIT] database, considering one utterance produced by all speakers (N=630), along with a few additional utterances. Those results suggested that /h/ voicing varied considerably depending on phonetic context. This study extends that work by measuring all /h/-containing utterances in the TIMIT database that were produced by multiple speakers. Each possible /h/ was inspected to determine whether the /h/ was deleted, showed visible noise, and/or had a phonation break (defined for each speaker as more than the duration of two pitch periods as measured in an adjacent stressed vowel). The utterances were then coded for position (utterance-initial, word-initial, intervocalic); preceding and following phonemes; stress (full or reduced vowel following); and speaker (gender and dialect region). These data will provide a more complete picture of the environments that favor phonation and aspiration noise in an open vocal tract in running speech.

WEDNESDAY MORNING, 28 NOVEMBER 2007 GRAND BALLROOM D, 8:30 TO 10:05 A.M.

Session 2aUWa

Underwater Acoustics: Marine Sediments and Bottom Interactions

William M. Sanders, Chair
Seafloor Sciences Branch, Naval Research Lab., Stennis Space Center, MS 39529

Chair's Introduction—8:30

Contributed Papers

8:35

2aUWa1. Sub-bottom profiling using time-frequency analysis. William Sanders, Dale Bibee (Seafloor Sci. Branch, Naval Res. Lab., Stennis Space Ctr., MS 39529), and Edit Kaminsky (Univ. of New Orleans, New Orleans, LA 70148)

Methods for characterizing the returns from acoustic scatterers can roughly be separated into spectral or imaging methods. If the wavelength of the acoustic signal is small compared to the size of the target (geometrical scattering), the object may be imaged to such a fine resolution that its shape or structure may be revealed. This is often achieved using a wide bandwidth (high temporal resolution) signal. Otherwise, if the wavelength of the acoustic signal is large compared to the target size, spectral methods are deemed superior. These utilize relatively narrowband methods (with good spectral resolution) designed to identify characteristic resonances or nulls in spectra. However, practical factors place sub-bottom profiling systems squarely in the middle of these two domains. Hence time-frequency analysis becomes a viable option for characterization of scattering from buried objects. This study utilized simulated and tank-collected data to demonstrate the utility of Wigner-Ville distributions in analyzing and characterizing specific acoustic interactions. Salient features of scattering are revealed in the Wigner-Ville distributions that are not easily identifiable in either time domain or frequency domain only analysis. [This work was supported in part by the Naval Research Laboratory under program element PE62782N.]

8:50

2aUWa2. High frequency scattering from water-saturated granular sediments: Scaling effects. Anatoliy Ivakin (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105) and Jean-Pierre Sessarego (CNRS/LMA, 13402 Marseille cedex 20, France)

Sound backscattering from water-saturated granular sediments at frequencies from 150 kHz to 8 MHz at normal and oblique incidence was studied in controlled laboratory conditions. Two kinds of sediments, moderately sorted medium and coarse sands, were chosen for the study, degassed, and their surface was flattened. At oblique incidence, intrinsic scattering due to the sediment granular structure can be considered as a dominating mechanism of backscatter. Comparison of frequency dependencies of the backscattering strength for the two sediments with different mean grain size shows the existence of a persistent scaling effect that allows a description of the backscattering strength as a unique scaling function of one parameter, the mean grain size/wavelength ratio. For a wide range of this ratio, 0.15 to 1.3, the scaling function has been mea-

sured and presented, which is the major result of this work. Possible applications to remote sensing of marine sediments, and particularly determining the sediment mean grain size, are discussed. [Work supported by ONR and CNRS.]

9:05

2aUWa3. Comparison of finite element model of scattering with experimental results from the Experimental Validation of Acoustic modeling techniques (EVA) sea test. Marcia Isakson, Nicholas Chotiros, James Piper (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713), and Mario Zampolli (NATO Undersea Res. Ctr., La Spezia, Italy)

Reflection coefficient measurements from 5 50 kHz at grazing angles from 10–80 degrees were taken as part of the Experiments for Validation of Acoustic modeling techniques (EVA) sea test on the north shore of Isola dElba, Italy. The microtopography of the seafloor was also measured using an ROV mounted laser line scan system. The seafloor roughness measurements were used as inputs in a finite element model to calculate the effects of scattering on the reflection coefficient measurement. The variance of the calculated reflection coefficient is compared with the measured variance to determine the contribution from interface scattering at different frequencies and angles. [Work supported by ONR, Ocean Acoustics.]

9:20

2aUWa4. Numerical study of the influence of roughness on coherence of signals reflected by the seafloor. Laurent Guillon (Ecole navale, IRENav, BP 600, 29240 Brest armees, France, guillon@ecole-navale.fr) and Charles W. Holland (The Penn State Univ., State College, Pennsylvania 16804)

The context of this work is the study of broadband acoustic signals reflected by the seafloor and recorded on a 15-hydrophone vertical array. Two different aspects of the coherence of these signals were examined: in the temporal domain with the maximum of the cross-correlation coefficients and in the frequency domain with the phase of the cross-spectra. In previous work, it was shown that these coherence parameters are sensitive to the geoacoustic nature of the seafloor which may open the door for a new method of probing the seabed properties on both large scales (i.e., deterministic) and small scales (i.e., statistical). The aspect examined in this present work is the influence of the seafloor roughness on the coherence. To do so, the time series are simulated by a three-step process: 1) computation of the plane-wave reflection coefficient, 2) computation of the spherical wave reflection coefficient by a numerical integration of the Sommerfeld integral, and 3) Inverse Fourier Transform. Using these simulations, with the measured data as a guide, the potential for detecting and extracting roughness estimates from the coherence in complex multilayered seabed environments is explored.

9:35

2aUWa5. Extension of the grain-shearing theory of wave propagation in marine sediments to include pore-fluid viscosity. Michael J. Buckingham (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238)

Buckingham's grain-shearing (GS) theory of wave propagation in saturated granular materials predicts a sound speed showing weak dispersion and an attenuation that scales essentially linearly with the frequency. Measurements of the dispersion curves made during the SAX99 experiment in the Gulf of Mexico match the GS theory at higher frequencies, above 10 kHz, but show a lower sound speed and a higher attenuation than predicted by the GS theory at lower frequencies. A generalized version of the GS theory has now been developed, designated the VGS theory, which takes into account the effective viscosity of the molecularly thin layer of pore fluid between contiguous grains. In effect, the viscosity of the pore fluid limits the degree strain-hardening that occurs as grains slide against one another. The resultant VGS dispersion curves are the same as those of the GS theory, except for the appearance of a simple algebraic function that has an effect only at frequencies below about 10 kHz. This low-frequency function gives rise to VGS compressional-wave dispersion curves that match the sound speed and attenuation measurements made during SAX99 across the frequency band from 1 kHz to 400 kHz. [Research supported by ONR.]

9:50

2aUWa6. A finite-element analysis of reflection coefficient variability due to interface roughness. R. Abe Yarbrough and Marcia J. Isakson (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, abe@arlut.utexas.edu)

The effect of sea floor roughness on the reflection of sound diverging from a spherical source is investigated. An ensemble of water-sediment interface realizations are modeled, each with height statistics matching those of a rough surface prepared in a laboratory experiment. Using the finite-element method to solve the Helmholtz equation, time harmonic solutions from 40 to 60 kHz are calculated. From these frequency domain solutions, the band limited impulse response time domain solution is constructed via Fourier synthesis. The first two moments of the spherical wave reflection coefficient are determined as a function of grazing angle, and results are compared to measurements from the laboratory experiment. [Work supported by ONR, Ocean Acoustics.]

Session 2aUWb

Underwater Acoustics: Uncertainty and Fluctuations

Richard A. Keiffer, Chair
Acoustics Div., Naval Research Laboratory, Stennis Space Center, MS 39529-5004

Chair's Introduction—10:25

Contributed Papers

10:30

2aUWb1. Application of generalized polynomial chaos method to describe uncertainty in acoustic scattering from random rough interfaces. Richard Keiffer (Naval Res. Lab., Stennis Space Ctr., MS 39529, keiffer@nrlssc.navy.mil)

Recently, the transformed field expansion (TFE) boundary perturbation method [D. P. Nicholls and F. Reitich, J. Comput. Phys. **170**, 276–298 (2001)] has been extended to random rough interfaces using the method of generalized polynomial chaos (gPC) [D. Xiu and J. Shen, Comm. in Comp. Physics **2**, 54–72 (2007)] to compute statistics of the scattered field. The TFE boundary perturbation method is similar to the usual small wave height approximation approach, but starts with a coordinate transformation that avoids ill-conditioning in the perturbation series and maps the original deterministic Helmholtz equation on a random domain to a stochastic Helmholtz equation on a deterministic domain. When dealing with large multiscale rough interfaces like the ocean surface, the number of expansion terms in the gPC representation of the field can grow large. One strategy to keep the number of random variables manageable is to pose the scattering problem in the time domain, allow for the excitation of the rough interface by an impulse, and consider only those random surface features that can be causally linked to the received field. [This work was supported by ONR through NRL.]

10:45

2aUWb2. Response uncertainty using the stochastic finite-element method. John Burkhardt (U.S. Naval Acad., Annapolis, MD 21402)

Acoustic field predictions in shallow water are complicated by irregular bottom profiles and composition, sea-surface effects, as well as poorly characterized sound speed fields. To model the natural uncertainty of the problem, a variety of stochastic methods have been used. This presentation will consider the use of the stochastic finite-element method for the solution of the problem. The presented approach differs from previous spectral approaches in that the spatial problem, isolated by use of a polynomial chaos expansion, will be solved using the finite-element method, i.e., using a finite expansion of piecewise linear shape functions. Use of the finite-element method allows for the natural incorporation of irregular boundaries, as well as spatially and temporally dependent material properties. To demonstrate the technique, a one-dimensional elastic rod with spatially varying stiffness (sound speed) was considered. Specifically, the mean value of the rod stiffness and the standard deviation of the rod stiffness were varied in space. The technique was successfully implemented and confirmed using Monte Carlo techniques.

11:00

2aUWb3. An experimental test of a model for scattered envelope statistics. Anthony P. Lyons, Susan E. Parks (Appl. Res. Lab., The Penn State Univ., State College, PA 16804), and Thomas C. Weber (Univ. of New Hampshire, Durham, NH 03824)

Co-located acoustic and high-resolution environmental data were collected at Seneca Lake in the summers of 2005 and 2006. The first acoustic data set was acquired in August of 2005 near Dresden, New York, over an approximately 5 square km section of Seneca Lake near the Naval Undersea Warfare Center's System Measurement Platform and consisted of large numbers of high bandwidth high-frequency pings (tens of kHz). Analysis of the acoustic data showed areas of distinctly nonRayleigh statistics. In mid-July, 2006, high-resolution environmental ground truth data were collected including bathymetry and scattering strength data obtained using a RESON 8101 SeaBat 240 kHz calibrated multibeam system, video from ROV, and sediment grab samples. The ground truth data showed the dominant scattering mechanism to be large numbers of Quagga mussels. The spatial distribution of the mussels can be used to predict the broadband statistics using a recently developed model [D. A. Abraham and A. P. Lyons, IEEE J. Ocean. Eng., **27**, 800–813, (2002)], providing an excellent opportunity to test the model. Results and model/data comparisons are presented in this paper. [Work performed under ONR Grant Nos. N00014-04-1-0013 and N00014-06-1-0245.]

11:15

2aUWb4. Uncertainty and coherence function for the scattering by a time-varying, two-dimensional, rough surface. Robert W. Scharstein (Elec. Eng. Dept., Univ. of Alabama, 317 Houser Hall, Tuscaloosa, AL 35487-0286, rscharst@bama.ua.edu)

A perturbative analysis is used to construct and evaluate the mutual coherence function of the scattered field due to point-source excitation of a time-varying, rough surface. The wide-sense stationary (statistically homogeneous) rough surface is characterized by a simple time-space autocorrelation function that displays features of the moving sea surface. Doppler effects and evanescent wave contributions are explicit in the plane-wave spectral terms in the perturbation series, where the small parameter is the mean-square surface height.

11:30

2aUWb5. Sound fluctuation in a non-Gaussian random field: A wavelet-based Wiener chaos expansion approach. Hao Xing and Li Ma (Inst. of Acoust., Chinese Acad. of Sci., 21 Bei Si Huan Xi St., Beijing 100080, P. R. China)

Recently Wiener chaos derived stochastic expansion methods, especially polynomial chaos (PC), have caught great attention in the underwater acoustics community for its success in modeling random properties of underwater acoustic field. However, most research until now concern a Gaussian random field, while fluctuations in ocean medium are not always Gaussian. For non-Gaussian probability density function (PDF) with a finite support, it is reported that a Wiener chaos decomposition, which use wavelets as basis functions instead of orthogonal polynomials, is more suitable than PC. This paper concentrates on non-Gaussian high frequency internal waves [T. Wang and T. Gao, Chin. J. Ocean. Limnol. **20**, 16–21 (2002)] and presents a multi-resolution stochastic expansion approach to handle acoustic field fluctuation in a non-Gaussian internal wave field. The depth dependent PDF of sound speed fluctuation is determined empirically from experiment results. Spatial coherence function and scintillation index of acoustic field were calculated for Gaussian and non-Gaussian cases. In addition, a comparison between both situations was made, which shows that in the non-Gaussian case, the acoustic field has a more complex coherence structure. [Work partially supported by NSFC.]

Meeting of the Standards Committee Plenary Group

to be held jointly with the meetings of the
ANSI-Accredited U.S. Technical Advisory Groups (TAGs) for:
ISO/TC 43, Acoustics,
ISO/TC 43/SC 1, Noise,
ISO/TC 108 Mechanical vibration, shock, and condition monitoring,
ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles, and structures,
ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring instruments,
ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock,
ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines,
ISO/TC 108/SC 6, Vibration and shock generating systems,
and
IEC/TC 29, Electroacoustics

P. D. Schomer, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics and ISO/TC 43/SC 1, Noise
2117 Robert Dr., Champaign, IL 61821

D. J. Evans, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108, Mechanical vibration shock and condition monitoring, and ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring devices
National Institute of Standards and Technology (NIST), 100 Bureau Dr., Stop 8220, Gaithersburg, MD 20899

A. F. Kilcullen, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles, and structures
734 Walden Rd., Hedgesville, WV 25427

R. Taddeo, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
1333 Isaac Hull Ave., SE, Washington Navy Yard, Washington, DC 20376

D. D. Reynolds, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock
3939 Briar Crest Ct., Las Vegas, NV 89120

D. J. Vendittis, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
701 Northeast Harbour Ter., Boca Raton, FL 33431

R. Taddeo, Vice Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
1333 Isaac Hull Ave., SE, Washington Navy Yard, Washington, DC 20376

G. Booth, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 6, Vibration and shock generating systems
44 Bristol St., Branford, CT 06405-4842

V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics
National Institute of Standards and Technology (NIST), Sound Building, Rm. A147, 100 Bureau Dr., Stop 8221, Gaithersburg, MD 20899-8221

The reports of the Chairs of these TAGs will not be presented at any other S Committee meeting.

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S2, S3, and S12, which are scheduled to take place in the following sequence:

ASC S12 Noise	**28 November 2007**	**10:45 a.m. to 12:00 noon**
ASC S1 Acoustics	**28 November 2007**	**1:30 p.m. to 2:30 p.m.**
ASC S3 Bioacoustics	**28 November 2007**	**2:45 p.m. to 3:45 p.m.**
ASC S2 Mechanical Vibration and Shock	**28 November 2007**	**4:00 p.m. to 5:00 p.m**

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees and U.S. TAGs.

The U.S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3, and S12 are as follows:

U.S. TAG Chair/Vice Chair	TC or SC	U.S. Parallel Committee
ISO		
P. D. Schomer, Chair	**ISO/TC 43** Acoustics	S1 and S3
P. D. Schomer, Chair	**ISO/TC 43/SC1** Noise	S12
D. J. Evans, Chair	**ISO/TC 108** Mechanical vibration shock and condition monitoring	S2
A. F. Kilcullen, Co-Chair R. Taddeo, Co-Chair	**ISO/TC 108/SC2** Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures	S2
D. J. Evans, Chair	**ISO/TC 108/SC3** Use and calibration of vibration and shock measuring instruments	S2
D. D. Reynolds, Chair	**ISO/TC 108/SC4** Human exposure to mechanical vibration and shock	S3
D. J. Vendittis, Chair R. F. Taddeo, Vice Chair	**ISO/TC 108/SC5** Condition monitoring and diagnostics of machines	S2
G. Booth, Chair	**ISO/TC 108/SC6** Vibration and shock generating systems	S2
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	S1 and S3

WEDNESDAY MORNING, 28 NOVEMBER 2007 SOUTHDOWN ROOM, 10:45 A.M. TO 12:00 NOON

Meeting of Accredited Standards Committee (ASC) S12 Noise

R. D. Hellweg, Chair S12
Hellweg Acoustics, 13 Pine Tree Rd., Wellesley MA 02482
W. J. Murphy, Vice Chair S12
NIOSH, 4676 Columbia Pkwy., Mail Stop C27, Cincinnati, OH 45226

Accredited Standards Committee S12 on Noise. Working Group Chairs will report on the status of noise standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 43/SC 1 Noise, take note - that meeting will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Wednesday, 28 November 2007.

Scope of S12: Standards, specifications, and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort and physical acoustics as related to environmental and occupational noise.

Session 2pAA

Architectural Acoustics: Impact and Footfall Noise II

David Braslau, Chair
David Braslau Associates, Inc., 1313 5th St., SE, Ste. 322, Minneapolis, MN 55414

Contributed Papers

1:20

2pAA1. Sound transmission loss of nontraditional building materials and redundancies. Michael Ermann (Virginia Tech. School of Architecture + Designs, 201 Cowgill Hall (0205), Blacksburg, VA 24061, mermann@vt.edu), James Carneal, Daniel Mennitt, Christopher Jackson, Bharti Karmarkar, Matthew Helveston, and Patrick Clay (Virginia Tech., Blacksburg, VA 24061)

This line of inquiry aims to provide sound transmission loss data, both lab- and field-tested, for novel transparent and translucent wall assemblies. Special attention is paid to: (1) Redundant assemblies featuring two layers of material with a large gap or small room between them, (2) assemblies featuring emerging materials, and (3) assemblies with high thermal resistance values. To that end, composite configurations of acrylic, polycarbonate, aerogel, and glass were tested in a laboratory, and an acrylic room and glass vestibule were tested in the field. Results will be presented with an eye toward future use of the data. Indeed, it was the unavailability of the data in existing publications that initially drove the research.

1:35

2pAA2. Impact noise in a prefabricated custom modular condominium. Michael Ermann, Gret Tew (Virginia Tech. School of Architecture + Design, 201 Cowgill Hall (0205), Blacksburg, VA 24061, mermann@vt.edu), and Matthew V. Golden (Kinetics Noise Control)

This line of inquiry aims to provide field-measured impact noise data for ''box-on-box'' construction, as was found in field tests. This type of construction holds the potential to increase impact isolation class (IIC) because the very process of building discrete boxes off site, transporting them, and stacking them on-site, establishes an airspace between the floor joist of one box and the ceiling joist of another. This partially decouples the floor and ceiling (a ridged connection remains at the perimeter of the boxes). A preliminary test of wood finished flooring yielded an FIIC of 42. Ongoing supplemental tests with floor underlayments and resilient seismic connections at the perimeter of the boxes will also be presented.

1:50

2pAA3. Sound insulation in residential buildings. Ballard W. George (1367 Bobolink Circle, Sunnyvale, CA 94087-3312)

This paper considers similarities and differences experienced with sound insulation in various types of residential buildings, such as townhouses and condominiums. Typically with two adjacent dwelling units, there is a primary inter-unit sound path, such as a party wall, and other paths that are categorized as flanking paths. In typical condominiums and apartments, the floor/ceiling is also a primary path. Implications of the plywood subfloor extending across the party wall are discussed. Avantages along with potential limitations of double stud party walls in controlling noise flanking transmission by the subfloor path are shown. Diagonal noise transmission between units is a factor that is common to both townhouse and apartment and condominium types of buildings. An example is given of the conversion of a single famiy house to a vertically-separated duplex, with implications regarding the closing off of the stairs and isolation of the garage from the neighbor unit. The exterior wall, due to its proximity to the neighboring building, acted to some extent as a ''party wall'' assembly. An example is also given of a ''defacto'' townhouse where settling had caused the exterior wall of one single dwelling to butt against another house.

2:05–2:50
Panel Discussion

Issues related to footfall noise
Warren E. Blazier, Jr., Moderator

Session 2pAB

Animal Bioacoustics: Sound Source Localization

Eduardo Mercado, III, Cochair
Univ. at Buffalo, Dept. of Psychology, Buffalo, NY 14260-4110

Micheal L. Dent, Cochair
Univ. at Buffalo, Dept. of Psychology, Buffalo, NY 14260-4110

Chair's Introduction—1:00

Invited Papers

1:05

2pAB1. Evolution of sound localization in lizards. Catherine Carr, Yezhong Tang (Dept. Biol., Univ. of Maryland, College Park, MD 20742-4415), and Jakob Christensen-Dalsgaard (Univ. of Southern Denmark, DK-5230 Odense M, Denmark)

Tympanic hearing may have developed independently in four major tetrapod groups, the anurans, lepidosaurs, archosaurs, and mammals. The emergence of a tympanic ear would have increased the frequency range and sensitivity of hearing. Furthermore, early tympana were acoustically coupled through the mouth cavity and therefore must have acted as pressure difference receivers, which are inherently directional, obviating a need for central computation of ITD and ILD. In lizards, this acoustical coupling generates a large directional difference at the tympanum and recordings from auditory nerve are strongly ITD- and ILD-dependent. To separate neural from acoustical interactions, interaural transmission in a gecko was blocked to reveal true binaural neurons in nucleus magnocellularis, nucleus angularis, and superior olive. The function of these neurons with naturalistic, i.e., free-field stimulation may be to sharpen the directional response, a mechanism similar to that suggested for teleost fish. The closure of the middle ear cavity in the mammals and some birds is a derived condition, and may have profoundly changed the operation of the ancestral tetrapod ear by decoupling the tympana, improving the low-frequency response of the tympanum, and leading to a requirement for neural computation of directionality in the central nervous system.

1:25

2pAB2. Are binaural cues available for sound source localization in teleost fishs? Peggy Edds-Walton and Richard Fay (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 North Sheridan Rd., Chicago, IL 60626)

When reading the review literature regarding directional hearing or sound source localization by fishs, one usually sees a list of reasons why fish cannot use binaural cues. The assumptions are that teleost fishs cannot detect left versus right differences in arrival times (due to the speed of sound in water), amplitude differences (due to the closeness of the ears and the acoustic transparency of the tissues), nor phase differences (due to the long wavelengths of the frequencies most fish can hear best). However, there is some behavioral evidence that fishs can distinguish the direction of a sound source, if not the actual location. We will present evidence that there are potential binaural cues that might be available given left versus right differences in auditory input to the brain that are the result of both anatomical organization and physiological mechanisms. In addition, sharpened directional responses are present in the medulla and the midbrain, which are likely to be due to binaural computations. [Work funded by the National Institutes of Health, NIDCD.]

1:45

2pAB3. Sound localization by echolocating dolphins. Brian K. Branstetter (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., Univ. of Hawaii, Honolulu, HI 96822 branstet@hawaii.edu) and Eduardo Mercado III (Univ. of Buffalo, State Univ. of New York, Buffalo, NY)

Dolphins (and other odontocetes) rely on echolocation to aid in detecting, identifying, localizing, and capturing prey. Dolphin sonar signals are well suited for localizing multiple, small targets in a three-dimensional aquatic environment by utilizing highly directional (3 dB beamwidth of about 10 deg), broadband (3 dB bandwidth typically of about 40 kHz; peak frequencies between 40 kHz and 120 kHz), short duration clicks (about 40 μs). Dolphins can localize sounds both passively and actively (echolocation) with a resolution of about 1 deg. Cross-modal matching (between vision and echolocation) suggests dolphins perceive the spatial structure of complex objects interrogated through echolocation, a feat that likely requires spatially resolving individual object features and integration into a holistic representation of object shape. Although dolphins are sensitive to small, binaural intensity and time differences, mounting evidence suggests dolphins employ position-dependent spectral cues derived from well developed head-related transfer functions, for sound localization in both the horizontal and vertical planes. A very small temporal integration time (264 μs) allows localization of multiple targets at varying distances. Localization adaptations include pronounced asymmetry of the skull, nasal sacks, and specialized lipid structures in the forehead and jaws, as well as acoustically isolated middle and inner ears. [Financial support was provided by the National Science Foundation under Grant No. 234046.]

2:05

2pAB4. Sound localization by echolocating bats: Are auditory signals enough? Murat Aytekin, Cynthia F. Moss (Dept. of Psych. and Inst. of Systems Res., Univ. of Maryland, College Park, MD 20742), and Jonathan Z. Simon (Univ. of Maryland, College Park, MD 20742)

Echolocating bats localize objects by producing ultrasound vocalizations and listening to sonar returns. Interaural differences in the echoes arriving at the bats' ears contribute to the process of object localization. The spatial properties of sonar emissions and the mobility of the head and pinna also contribute to the process of sound localization in bats. Measurements of sonar emission patterns from unrestrained, freely vocalizing bats have shown that sonar beam shape varies across individual vocalizations. Consequently, the bat auditory system needs to adapt its computations to accurately localize objects using changing acoustic inputs. Extra-auditory signals that carry information about pinna position and beam shape are required for auditory localization of sound sources. The auditory system must learn associations between extra-auditory signals (e.g., head and pinna position) and acoustic spatial cues. These demands on the nervous system suggest that sound localization is achieved through the interaction of behavioral control and acoustic inputs. A sensorimotor model that demonstrates how an organism can learn space through auditory-motor contingencies is presented. The model also reveals how different aspects of sound localization, e.g., experience-dependent acquisition, adaptation, and extra-auditory influences, can be joined under a comprehensive framework.

2:25

2pAB5. The acoustics and psychoacoustics of sound localization in the cat. Daniel Tollin (Univ Colorado Hlth. Sci. Cntr, Dept. Physiol., RC1-N, 8307, Box 6511, 12800 E 19th Ave, Aurora, CO 80045, daniel.tollin@uchsc.edu)

Cats are a common model system for the anatomy, physiology, and psychophysics of sound localization. There are three main cues to location: Interaural differences in time (ITD) and level (ILD) and monaural spectral shape. Here, the cues were derived from measurements of directional transfer functions, the directional components of head-related transfer functions. In the frontal hemisphere, spectral notches were present for frequencies from ~8–16 kHz; the notch frequency increased with increases in source elevation and in azimuth towards the ipsilateral ear. The maximum ITDs were ~380 s. Maximum ILDs depended on source azimuth and frequency, but were as large as 35 dB at some frequencies. For the psychophysics of localization, cats were trained using operant conditioning to indicate the apparent locations of sounds via gaze shift (combined eye and head movements); eye and head position were measured using the search coil technique. Auditory targets consisted of noise bursts of varying duration, intensities, and spectral content. Localization of long duration, moderately intense, and broadband targets were accurate and precise, but deteriorated for sounds that were short-duration, narrow-band, or that were very intense or faint. Cat sound localization capabilities are in many ways similar to humans. [NIH grants DC006865 and DC02840.]

2:45

2pAB6. Training-induced improvements on interaural level difference (ILD) and interaural time difference (ITD) discrimination in human adults. Beverly A. Wright (Dept. of Comm. Sci. and Dis., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208)

The two primary cues to the location of sound sources on the horizontal plane are interaural level differences (ILDs) and interaural time differences (ITDs). The malleability of the processing of these two cues in humans was investigated by examining how multiday practice affects the discrimination of different values of ILDs and ongoing ITDs presented over headphones in adults with normal hearing. On average, the listeners improved on both ILD and ITD discrimination, but the learning patterns differed between the two cue types. Improvement was initially rapid for both cue types and appeared to generalize broadly across conditions, suggesting procedural learning. However, a subsequent slower-improvement stage occurred solely for the ILD cue and showed some specificity to the stimulus used during training. Interestingly, for ILD discrimination, both the best and worst daily threshold estimates decreased with multiday training, indicating an improvement in fundamental processing capacity. In contrast, for ITD discrimination, the best threshold estimates remained unchanged but the worst decreased, suggesting an increased ability to access already existing capacities. One interpretation of these data is that training mediates the processing of ILDs and ITDs at a stage in the auditory pathway, where these two cues are processed separately. [Work supported by NIH/NIDCD.]

3:05–3:25 Break

3:25

2pAB7. Neural computations in the owl's sound localization. Jose L. Pena (Dept. of Neurosci., Albert Einstein College of Medicine, Kennedy Ctr. Rm 529, Bronx, NY 10463)

Neural models predict mathematical operations taking place in systems of neurons. The plausibility of such operations is a matter of debate. The owl's auditory system offers a good substrate to approach this question. To determine what and where acoustic objects are, the brain must compare and combine noisy information coming from each ear. Comparison, combination, and noise reduction can be represented by equations. A common tool used to extract signals from variable sets of stimuli is averaging. Similarly, cross-correlation and multiplication are mathematical correlates of comparison and combination selectivity, respectively. In the owl's auditory pathway, we can find neurons that effectively perform these types of processing. The nucleus Laminaris neurons perform a running cross-correlation that indicates the degree of similarity between right- and left-side inputs. Midbrain neurons become selective to combinations of spatial cues by effectively multiplying inputs from independent processing pathways. In the inferior colliculus, an average-like process produces a tuning to interaural time difference that would require at least five repetitions of the same stimulus one stage before. Thus, the owl's auditory system appears to perform the computations that models predict, representing an interesting example of how close to the nature of neural processing mathematical operations are.

3:45

2pAB8. Recovery from localization dominance in the barn owl. Brian S. Nelson and Terry T. Takahashi (Inst. of Neurosci., Univ. of Oregon, 1254 Univ. of Oregon, Eugene, OR 97403, bsnelson@uoregon.edu)

Barn owls, like humans, localize sounds that arrive directly from a source and are less influenced by reflections that follow after a short delay. Neural responses in the owls' auditory space map are diminished when a sound duration exceeds that of the reflection's delay and firing recovers only when the delay exceeds ~10 ms. Neither the onset- nor the ongoing-delay causes this diminishment. Instead, recovery corresponds with the emergence of a response at the stimulus end, when only the reflection is present. Head-saccades were measured while manipulating the duration of the reflection-alone sound segment. The proportion of saccades that each owl made toward the lagging source depended on the duration of the reflection-alone segment and was independent of the stimulus onset-delay. The perceptual salience of a reflection, or other lagging sound, may thus depend on the space map's response to this reflection-alone segment. [Work supported by NIH, Grants F32 DC008267-01A1 to B.S.N and DC 03925.]

4:05

2pAB9. The building blocks of ranging: What laboratory studies of distance perception can tell us. Leslie Phillmore (Dept. of Psych., Dalhousie Univ., Halifax, NS, Canada, B3J 4H1, Leslie.Phillmore@dal.ca), Christopher Sturdy (Univ. of Alberta, Edmonton, AB, Canada T6G 2E9), and Ronald Weisman (Queen's Univ., Kingston ON, Canada K7L 3N6)

Ranging (estimating the distance to) conspecifics is an important skill for songbirds. For example, territorial males must determine, often by acoustic information alone, whether a rival male is either within territorial boundaries, requiring an aggressive response or outside boundaries, requiring him to withhold response to conserve energy. Most field studies of ranging in songbirds measure males' behavioral responses to playback of vocalizations seeming to originate from within or outside an established territory. However, this approach is inadequate for investigating ranging abilities in species that do not defend territories or for assessing the extent to which ranging abilities are dependent on early experience with distance. In a series of studies, a go/no-go operant task requiring birds to discriminate vocalizations recorded at various distances was developed to ask some of these comparative questions about auditory distance perception. Results showed that a territorial species, black-capped chickadees (Poecile atricapillus), learned to discriminate more quickly than a nonterritorial species, zebra finches (Taeniopygia guttata), that both species learned to discriminate chickadee vocalizations more quickly, and that chickadees raised without experience with ranging could perform the distance cue discrimination task as well as field-reared birds. [Research completed at Queens University, Kingston, Canada; supported by NSERC.]

4:25

2pAB10. Why humpback whales learn songs: Courting or ranging? Eduardo Mercado III (Dept. of Psych., Park Hall, Univ. at Buffalo, The State Univ. of New York, Buffalo, NY 14260, emiii@buffalo.edu)

Singing humpback whales within a population appear to continuously modify the content of their songs such that at any given time, most whales in a particular region are singing stereotypically-structured songs. This phenomenon is widely regarded as evidence that humpback whales learn songs from conspecifics, and is often cited as evidence that humpback whale songs are a reproductive acoustic display comparable to the songs of songbirds. An alternative possibility, however, is that humpback whales copy the songs of their neighbors because this maximizes their ability to passively or actively track the positions of other whales. In particular, comparing incoming sounds to memories of self-produced sounds may enhance a whale's capacity to judge the distance that the sound has traveled by increasing the whale's ability to identify degradation cues that vary systematically as a function of propagation distance. An increased capacity to accurately estimate the position of sound sources might in turn improve a whale's ability to segregate songs when multiple singers are audible. This alternative perspective suggests that song learning by humpback whales is an adaptation for auditory spatial perception, rather than the result of sexual selection for a communal vocal display that increases the responsiveness of females.

Contributed Paper

4:45

2pAB11. Acoustic radiation patterns of mating calls of the túngara frog: Implications for multiple receivers. Ximena E. Bernal, Rachel A. Page, Michael J. Ryan (Section of Integrative Biol., Univ. of Texas, 1 University Station C0930, Austin, TX 78712), Theodore F. Argo IV, and Preston S. Wilson (Univ. of Texas, Austin, TX 78713-8029)

In its simplest form, communication is an interaction between a signaler and a receiver in which the signal has some probabilistic influence on the behavior of the receiver. For communication to proceed, the signal must be detected and perceived by the receiver; that is, the receiver must be within the active space of the signal. If patterns of sound radiation are not omni-directional, the position of the receiver relative to the sender is critical. In previous measurements of the horizontal directivity of mating calls of frogs and toads, the acoustic signals were analyzed using peak or RMS analysis [Gerhardt, J. Comp. Physiol. A **102**, 1–12 (1975)] and the resulting broadband directivities ranged from negligible (omnidirectional) to a maximum of about 5 dB. We report new laboratory measurements of both the horizontal and vertical frequency-dependent directivity of the mating calls of túngara frogs (*Physalaemus pustulosus*). These band-limited directivities are significantly greater than broadband directivities, with a maximum directivity of 20 dB in the vertical plane for harmonics near 6 kHz. Ultrasonic frequencies, recently found in the calls of some frogs, are absent in the mating call of túngara frogs. The implications with regard to mating and predator-prey interactions are discussed.

Session 2pAO

Acoustical Oceanography and Underwater Acoustics: Deep and Shallow Seismic Sensing of Geological Structure in the Ocean Bottom

Altan Turgut, Cochair
Acoustics Div., Naval Research Laboratory, 4555 Overlook Ave., SW, Washington, DC 20375

Warren T. Wood, Cochair
Naval Research Lab., Stennis Space Center, MS 39529

Chair's Introduction—1:10

Invited Papers

1:15

2pAO1. Seismic investigation of natural methane hydrate deposits. Warren Wood (Naval Res. Lab., Code 7432, Stennis Space Ctr., MS 39529)

The world's continental margins are home to the very high pressures and low temperatures where methane combines with water to form the ice-like methane hydrate. Only in the last couple of decades has it been discovered that methane hydrate is distributed over virtually every continental margin, as well as Arctic and Antarctic regions, and as such, may play a significant role as a future energy source, an agent of global climate change, or a factor in marine slope stability. Because the acoustic impedance of gas phase methane is significantly less than sediment, water, or methane hydrate, the phase boundary between gas and gas hydrate is frequently quite strong, and easily discernible in seismic sections. Further, because pressure variations in this part of the seafloor are typically small, the phase boundary can be used to determine sediment temperature. In certain circumstances, lateral variations in the depth to this phase boundary can be used to constrain the flux of heat, pore-water, and methane responsible for concentrated methane hydrate deposits and chemosynthetic communities found at seafloor seeps.

2:00

2pAO2. High-resolution marine seismic reflection profiling: Systems and capabilities. Patrick E. Hart (U.S. Geological Survey, 345 Middlefield Rd., Menlo Park, CA 94015)

Seismic reflection profiling is the most effective and commonly used geophysical tool for sub-seafloor geologic investigations. A broad variety of seismic data acquisition systems are available, depending on the imaging depth and resolution required. High-resolution methods utilize controlled sound sources that generate signals in the frequency range of approximately 1,000 to 10,000 Hz. Data acquired with piezoelectric swept-frequency (chirp) transducers, mini-sparkers, and boomer-plate sources routinely image up to 100 m subbottom in marine sediments, with vertical resolution better than 50 cm. Small air-gun systems with frequencies between 100 Hz and 500 Hz achieve as much as 1 km of penetration and resolution of a few meters. Imaging depths are highly dependent on the geologic environment; surveys in fine-grained marine sediments usually produce the best results. Free gas from decaying organic matter in shallow sediments can dramatically degrade data quality. High-resolution methods are contrasted by current hydrocarbon exploration industry systems that utilize large arrays of air guns to generate high-energy signals between l0 Hz and 100 Hz recorded by multiple hydrophone streamers as much as 10 km in length. The resulting three-dimensional data volumes can provide detailed images to depths greater than 15 km.

2:20

2pAO3. Prestack full waveform inversion for estimating elastic properties of marine sediments. Mrinal Sen (UT Austin, J. J. Pickle Campus Bldg. 196, 10100 Burnet Rd., Austin, TX 78758)

Prestack seismic waveform inversion is a highly challenging task. The nonlinearity and nonuniqueness together with computed intensive forward modeling make the problem intractable. Here we report on a new approach to seismic waveform inversion based on a gradient descent scheme that makes use of innovative gradient calculation, adaptive regularization, and an efficient conjugate gradient algorithm. We exploit sparsity of the gradient matrix based on physics in designing an algorithm that is computationally very efficient. The results from our approach are highly encouraging in that we are able to invert large 3-D dataset with nearly 1,000 model parameters fairly rapidly. Not only that, because of the use of adaptive regularization, we are able to track smooth as well as sharp variations in the impedance, Vp and Vs profiles when realistic starting solutions are used. We have applied the technique successfully to numerous 3-D exploration and academic datasets. Of particular interest is our ability to estimate density from seismic data that can be used to map porosity in an effective manner. Application to seismic data collected over offshore Oregon resulted in estimates of Vp and Vs that could be used in estimating hydrate concentration in the shallow sediment.

2:40

2pAO4. Synthetic aperture processing of reflection profiling data. Steven Schock (Dept. of Ocean Eng., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431, schock@oe.fau.edu)

Chirp sonar data are processed using synthetic aperture processing to improve the spatial resolution and subsurface penetration of 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 reflectors are summed coherently. The changes in projector and hydrophone positions between successive transmissions are calculated based on Doppler velocity log and inertial measurement unit measurements of vehicle motion. The pulse repetition interval can exceed half-wavelength spacing by increasing the number of physical hydrophone channels. Imagery generated from reflection profiles of seabeds generated in several depositional environments are compared, and used to quantify the improvement in the spatial resolution and the SNR of reflection profiles as a function of the length of the synthetic aperture.

3:00–3:20 Break

3:20

2pAO5. Chirp seismic reflection data on the New Jersey middle and outer shelf: The geologic response to 40,000 years of sea level change. John Goff, James Austin, Jr. (Univ. of Texas Inst. for Geophys., Jackson School of Geosciences, Univ. of Texas, 10100 Burnet Rd., Austin, TX 78758), Beth Christensen (Adelphi Univ.), and Altan Turgut (Naval Res. Lab.)

Chirp seismic reflection data have been collected extensively on the New Jersey continental shelf in support of Office of Naval Research (ONR)-sponsored research. This composite data set includes surveys in 2001 and 2002 for the Geoclutter program, and in 2006 for the Shallow Water 06 program. Chirp uses a swept-frequency source (1–15 kHz), matched filtered to provide decimeter vertical accuracy and penetration up to 30 m in coarse-grained sediments. These data have been used to map out the complex shallow stratigraphy of this region, which includes: (1) a regional reflector formed about 40,000 years ago; (2) a layered shelf-edge wedge deposited during sea level fall; (3) fluvial channels formed at lowstand, and later filled with an estuarine sequence during sea level rise; (4) a ravinement surface created by shoreface erosion, which forms a lag of coarse material; and (5) the Holocene sand sheet, formed into sand ridges up to 10 m thick. Modern erosion has exposed many of the older units at or near the seafloor. Seabed sediments along acoustic propagation pathways are therefore highly heterogeneous. Vibracoring is planned for August, 2007. Analysis will include geoacoustic logging and geologic characterization. Preliminary results from these cores will also be presented.

3:40

2pAO6. Mobile parametric sub-bottom profiler system for shallow and medium depth applications. Jens Wunderlich (Innomar Technologie GmbH, Schutower Ringstr. 4, D-18069 Rostock, Germany, jwunderlich@innomar.com)

Sub-bottom profilers based on parametric acoustics offer many advantages, especially for applications with the focus on high resolution. They have a narrow beam in spite of small transducers and virtually no side lobes. They transmit very short sound pulses at high pulse repetition rates. Excellent vertical, as well as horizontal, resolutions at reasonable penetrations are proven benefits by using this technique. Electronic steering of focussed sound beams allows the compensation of movements of the survey platform. The paper presents a parametric system with low frequencies in the range of 4 kHz to 15 kHz. The transducer has an active area of 22 cm $\times$ 22 cm only and provides a stabilized sound beam of less than 4 deg. Such small size and no need for an additional receiver gives high mobility for the operation on different platforms. Sound beam stabilization guarantees the operation from offshore vessels under ocean conditions. Due to a special transmit regime, high ping rates of up to 30 pings per s are maintained, independent of water depth. Some more design parameters will be discussed and illustrated, with results from shallow water and offshore applications.

Contributed Papers

4:00

2pAO7. Extrapolation of geoacoustic properties using seismic reflection data: A shallow water example. Charles W. Holland (The Penn State Univ., Appl. Res. Lab., State College, PA), Allen Lowrie (238 F. Z. Goss Rd. Picayune, MS), Jan Dettmer and Stan E. Dosso (Univ. of Victoria, Victoria, BC, Canada)

In areas of the world that are bottom limited, sonar performance models require accurate seabed geoacoustic properties. Geoacoustic properties, e.g., sound speed, density, and attenuation as a function of depth, frequency and geographic position, are expensive to acquire over large areas. Most geoacoustic techniques, including direct sampling (e.g., cores, probes) or acoustic inversion methods (e.g., matched field methods) spatially sample either in 1-D or 2-D. Conducting these measurements in a dense grid over large areas is not feasible and as a consequence seabed database developers rely on seismic reflection data coupled with geologic models to spatially interpolate and extrapolate the geoacoustic properties. Seismic reflection data is attractive for this purpose because it has much greater geographic coverage and contains information on underlying geologic processes that ultimately control the geoacoustic properties. The inherent strengths and weaknesses of seismic data for geoacoustic extrapolation are explored using seismic reflection transects in the Straits of Sicily together with ground-truth geoacoustic properties and their uncertainties (from wide-angle reflection measurements) at each end of the transect. The uncertainties of the extrapolated geoacoustic properties are determined with and without the ground-truth data. [Work supported by the Office of Naval Research OA321 and the NATO Undersea Research Centre.]

4:15

2pAO8. Performance of a chirp sonar sub-bottom profiler for different bottom types. Altan Turgut (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

Performance of a 0.5–12 KHz chirp sonar sub-bottom profiler is evaluated for different bottom types. The bottom types at different survey sites include mud (Malta Plateau and Gulf of Mexico), silt (Santa Barbara Channel), sand (New Jersey Shelf), and mud stone (Oregon Shelf). Excellent vertical resolution and penetration revealed several geological features (e.g., buried river channels, mud slides, seep alterations, etc.) at muddy, silty, and sandy sites. Sub-bottom imaging at the Oregon Shelf site was relatively poor due to insufficient penetration and interference from side-lobes over a rough terrain of mud-stone bottom. Several improvements on the instrument reponse and source-receiver beam patterns are discussed. Methods for estimation of bottom types and measurements of reflection and attenuation coefficients are also presented. [Work supported by ONR.]

Session 2pEA

Engineering Acoustics and ASA Committee on Standards: Infrasonic Instrumentation II

Qamar A. Shams, Cochair
NASA Langley Research Center, 4 Langley Blvd., Hampton, VA 23681

Allan J. Zuckerwar, Cochair
4909 Camberley Cir., Williamsburg, VA 23188-8801

Chair's Introduction—1:00

Invited Papers

1:10

2pEA1. Infrasonic instrumentation. Gunnar Rasmussen (G.R.A.S. Sound & Vib. A/S, Skovlytoften 33, DK-2840 Holte, Denmark, gras@gras.dk)

Infrasonic measurements and calibration. Infrasonic measurements call for special instrumentation in order to cover the low frequency range. Below 100 Hz the microphones present a special problem. The microphone performance is very different between free field conditions with wave motion and closed conditions without propagating waves. The primary method for pressure calibration of standard microphones by the reciprocity technique described in IEC 1094-2 assumes that the microphone is only exposed to the pressure variation on the front of the diaphragm. A true pressure calibration will therefore present an increase in sensitivity below the air equalization time constant. In a free field the sound pressure has access to the back of the diaphragm through the air equalization vent and the sensitivity of the microphone will go towards zero at low frequencies. The uncertainty therefore at low frequencies goes towards infinity, when standards for free field sensitivity by comparison with a laboratory standard microphone or working standard microphone as presented in IEC 61094, part 3-7-8, are applied. A low frequency calibrator which will present a constant pressure, independent of frequency and adiabatic-isothermal conditions at frequencies lower than 250 Hz, will be discussed.

1:35

2pEA2. Comparison of measurement and theory for the acoustic pressure field in an infrasonic calibrator through the isentropic-to-isothermal transition. Timothy M. Marston and Thomas B. Gabrielson (Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

In a previous paper [T. M. Marston and T. B. Gabrielson, J. Acoust. Soc. Am. **119**, 3378 (2006)], a simple total-immersion calibrator for infrasonic microphones was described. This calibrator has been improved and so has the correspondence with theory. The Rott equations for thermoviscous acoustics have been incorporated in a stepwise solution for the acoustic field. These equations account for the transition between boundary-layer flow in which the main field is almost isentropic, and the very-low-frequency approach to isothermal pressure oscillations. The entire microphone body is placed in the calibrator so that the pressure-equalization leak functions normally. The calibrator operates from 0.02 Hz to 20 Hz and includes a DC-coupled piezoresistive differential pressure sensor for comparison calibration; however, the model, which incorporates the geometry of the calibrator pipe, the temperature, and the piston stroke, permits response estimation without dependence on the reference sensor. [Research funded by the Applied Research Laboratory Educational and Foundational Fund.]

2:00

2pEA3. Meteorological instrumentation for infrasonic wind noise studies. Richard Raspet and Jeremy Webster (Natl. Ctr. for Physical Acoust. and Dept. of Phys. and Astron., Univ. of Mississippi, University, MS 38677)

Recent research on the prediction of wind noise contributions from the measured turbulent velocity spectrum has led to a better understanding of what meteorological variables are necessary to analyze and compare wind noise measurements on infrasound arrays. In this talk, we will describe the variables necessary to correlate different contributions to the pressure spectrum, depending on the sensor configuration and arrangement. The instrumentation necessary to study surface pressure fluctuations on wind screens, the intrinsic turbulence-turbulence pressure fluctuation, the mean shear-turbulence fluctuation pressure, and the surface pressure fluctuations above flat surfaces will be discussed. The case of elevated shears above vegetation layers is particularly important to infrasound arrays. A theoretical investigation has identified what meteorological parameters must be evaluated in order to compare theory to measurement. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

2:25

2pEA4. Decoupling of seismic and infrasonic sources based on the orientation of a low-frequency electret condenser microphone. Qamar A. Shams, Allan J. Zuckerwar, Lloyd W. Binger (NASA Langley Res. Ctr., M.S. 238, Hampton, VA 23681), and John W. Stoughton (Retired NASA, Suffolk, VA 23434)

Membrane-based microphones are subject to seismic excitation, which interferes with signals from propagating infrasound. An experiment was designed to decouple seismic and infrasonic signals. The microphone was rigidly attached to a base plate on the ground with a fixture that permitted the orientation of the microphone axis to be varied. An accelerometer was mounted to respond to seismic stimuli normal to the base plate. Subsequent field testing included the dropping of various weights from a height of 2–3 m vertical drop at distances of 2 to 15 m from the system. In addition, a low frequency Helmholtz resonator (17 Hz) was used to provide an acoustic test signal. The intent of the experiment was to identify the best microphone orientation to reduce the effects of local seismic disturbances in the presence of low-frequency acoustic monitoring. Additional experiments were performed in the laboratory using a vibration exciter. The results of decoupling of seismic and infrasonic sources, based upon laboratory and field experiments, will be presented.

2:50–3:05 Break

3:05

2pEA5. Sources of background noise in low-frequency condenser microphones. Allan J. Zuckerwar (Analytical Services and Mater., 107 Res. Dr., Hampton, VA 23666, ajzuckerwar@yahoo.com), Qamar A. Shams (NASA Langley Res. Ctr., Hampton, VA 23681), and Jeremy A. Cooper (Virginia Commonwealth Univ., Richmond, VA 23284)

The background noise of a condenser microphone originates from external sources (electromagnetic interference, vibration) and internal sources (self-noise). The leading internal sources include the membrane damping resistance, the polarization voltage resistance (air condenser microphone only), the input resistance to the preamplifier, the capillary vent resistance, and a source of 1/f noise. These sources are modeled in a noise-equivalent circuit, from which an expression for the noise power spectral density is derived. This expression is compared to measurements on a low-frequency electret condenser microphone taken in an acoustic isolation vessel. The capillary vent and 1/f noises vary inversely with frequency and, thus, impact the background noise levels in the infrasonic region.

Contributed Papers

3:30

2pEA6. Acoustical properties of porous hose wind noise filters. Wheeler Howard, Kevin Dillion, and F. Douglas Shields (Mil-tec Res. Technol., Oxford Enterprise Ctr., Oxford, MS 38655)

Observation of infrasound signals in the atmosphere is often masked by wind noise. A common means of filtering out the wind noise is to connect a commercial porous (or soaker) hose to the manifold of a micro barometer. The filtering effect is attributed to the ability of the porous hose to average the pressure variations over its length. The pressure variations due to the turbulent wind field are incoherent on a scale equal to the hose length and, therefore, are reduced in the averaging process. Infrasound signals, with wavelengths much longer than the hose, are reduced little in the averaging process. There remains the question, ''How does the porous hose respond to infrasound signals that have wavelengths comparable to the hose length?'' To answer this question, measurements have been made of infrasound signals radiated from a jet engine during takeoff using an infrasound sensor with a porous hose connected to its port, and an array of infrasound sensors distributed along the length of the porous hose. Measurements are reported for the hose oriented both in the direction and perpendicular to the direction of the sound source. [Work supported by the Space and Missile Defense Command is gratefully acknowledged.]

3:45

2pEA7. Simple calculation models for wind noise spectral contributions in outdoor turbulent flows. Jiao Yu, Jeremy Webster, and Richard Raspet (Natl. Ctr. for Physical Acoust. and Dept. of Phys. and Astron., Univ. of Mississippi, University, MS 38677)

We have previously reported on the predictions of the stagnation pressure fluctuations, turbulence-turbulence interaction pressure fluctuations and turbulence-mean shear interaction pressure fluctuations in outdoor turbulent flows [J. Acoust. Soc. Am. **120**(5), 3032 (2006)]. The predictions of these contributions to the full pressure spectrum involve double integrals and are computationally complex. In this talk, we present simple algebraic fits to the predictions, and use these fits to discuss dominant terms under atmospheric conditions different from the measurement conditions reported earlier. Turbulence percentage and scale are shown to be the determining factor influencing the relative contributions to the wind noise outdoors. The simple calculation models provide a quick way to predict wind noise contributions from wind velocity measurements. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

4:00

2pEA8. An investigation of the correlation of infrasound signals as a function of sensor separation and wind velocity. Wheeler Howard, Kevin Dillion, F. Douglas Shields (Mil-tec Res. Technol., Oxford Enterprise Ctr., Oxford, MS 38655), and Claus Hetzer (Univ. of Mississippi, University, MS 38677)

Wind noise is a common obstacle to the detection and analysis of infrasonic signals. The masking caused by the wind noise is often reduced by spatial averaging via pipe or porous hose arrays. An alternate method is to average the signals from multiple sensors in distributed arrays. In order to reduce the wind noise, the sensors in the distributed array must be far enough apart so that their wind noise signals are incoherent, but not so far apart that the sound loses its coherence. Few measurements have been reported of the infrasound correlation distance in the presence of wind. This paper reports results of measurements of this infrasound correlation when the infrasound sensors are separated by distances up to hundreds of meters. Techniques for combining infrasound signals from widely separated sensors and measuring their correlation in the presence of wind noise are discussed.

4:15

2pEA9. Microphone array images of low-frequency wind noise. D. Keith Wilson, Roy Greenfield (U.S. Army Engineer Res Dev Ctr., 72 Lyme Rd., Hanover, NH 03755-1290, D.Keith.Wilson@erdc.usace.army.mil), and Michael J. White (U.S. Army Engineer Res Dev Ctr., Champaign, IL 61822)

Transient, propagating pressure disturbances associated with turbulent eddies (wind noise) and sound waves are imaged by wavelet processing of data from a horizontal, 49-microphone array. It is found that the choice of analyzing wavelet has a strong effect on the appearance of the pressure

images; because of the rapid spectral roll-off of the wind noise, there exist important trade-offs between temporal/spatial resolution and frequency resolution. For a given time scale, the wind disturbances are much smaller and less spatially coherent than sound waves. Since the wind disturbances move much more slowly than the sound waves, the spatial pattern they exhibit is independent of the look direction of a conventional acoustic beamformer. The wind disturbances also do not resemble plane waves. Conventional beamforming techniques are particularly sensitive to wind noise when sensor separations are small compared to the acoustic wavelengths of interest.

Session 2pED

Education in Acoustics: Professional Development Programs for K-12 Teachers of Science

James M. Sabatier, Cochair
National Center for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Corinne Darvennes, Cochair
Tennessee Technological Univ., Box 5014, Cookeville, TN 38505

Chair's Introduction—1:00

Invited Papers

1:05

2pED1. Preparing and retaining high quality science and mathematics teachers: Research, programs and results. Nicole Gillespie (Knowles Sci. Teaching Foundation, 1000 N. Church St., Moorestown, NJ 08057)

A growing body of research is revealing that the much-publicized lack of sufficient high-quality teachers in U.S. schools is a problem of retention rather than insufficient production, particularly in the case of secondary math and science teachers. Across all subjects, nearly 50% of teachers tend to leave the profession within the first 5 years of beginning teaching, so the attrition problem appears to be one of novice teachers rather than veterans. Furthermore, there is a common assumption in the U.S. that math and science teachers leave at a higher rate than teachers of other subjects because more, and more lucrative, career opportunities are available to individuals with science and math backgrounds. In response to a growing need for highly effective math and science teachers, the Knowles Science Teaching Foundation (KSTF) provides 5-year teaching fellowships to beginning secondary mathematics and science teachers across the United States. In this presentation, I will discuss the structure and rationale of the fellowship in light of national trends in math and science teacher qualifications, effectiveness and retention as well as preliminary results of the fellowship program evaluation research.

1:25

2pED2. Teachers in industry: A model for K-12 professional development. Sandra H. Harpole (Office of Res. and Economic Development, Mississippi State Univ., P.O. Box 6343, Mississippi State, MS 39762)

The Center for Science, Mathematics and Technology (CSMT) at Mississippi State University, with funding from the National Science Foundation and the Appalachian Regional Commission, has developed a model for professional development that utilizes industry-education partnerships, critical in educating students to enable them to compete for jobs in today's knowledge-based economy. Through quality learning experiences that result from research/industry experiences, teachers can better prepare students for the 21st century workforce, become more aware of career opportunities, and emphasize to their students the importance of preparation in science, technology, engineering, and mathematics. Professional development programs offered by the CSMT focus on utilizing partnerships to bridge the gap between the classroom and the workplace by placing K-12 teachers in industry to learn the relationship between academic content and real world applications, as well as the basic skills and competencies required by employers. Components of successful CSMT professional development programs include multi-year summer workshops providing industry experience, academic year follow-up sessions, and local industry internships. Partner industries include Northrop Grumman Ingalls Ship Systems, Tennessee Valley Authority, North American Coal, Nissan, and Stennis Space Center, as well as a number of industries located in participants local area. [Work is supported by NSF and ARC.]

1:45

2pED3. Issues of equity in physics access and enrollment in New York City public high schools. Angela Kelly (Dept. of Mathematics, Sci., and Technol., Teachers College, Columbia Univ., 525 W. 120th St., Box 210, New York, NY 10027)

Despite reports to the contrary, the availability of physics as a course for high school students is not equitably distributed throughout the U.S. While some schools provide physics for all, a more common scenario is limited availability, often open only to the relatively few students who have completed the necessary prerequisites or are deemed capable of understanding physics. Some schools do not offer physics at all. This is particularly true in urban districts, where science resources and instruction are chronically inadequate. This study looks at the access to and availability of physics courses for high school students in New York City. The

schools were surveyed to determine where physics was offered and how many students were enrolled in the course. Statistics were performed to compare differences between schools that offered physics and those that did not. Additionally, factors were examined that relate to physics availability, such as the magnet school configuration, the AP Physics option, science curricular sequence, and the conceptual physics option. Overall, it was determined that physics availability is very limited in NYC schools, a serious inequity that disproportionately affects students of color and poor children. Strategies for improving access and enrollment will be discussed.

2:05

2pED4. Revising New Mexico's science standards in 2003. Sharon B. Dogruel (The Education Ctr., Santa Fe, NM), Steven A. Sánchez (Las Cruces Public Schools, Las Cruces, NM), and Gregory W. Swift (Los Alamos Natl. Lab., Los Alamos, NM, swift@lanl.gov)

Partly in response to the No Child Left Behind Act of 2001, most states use ''standards'' to guide the creation of K–12 curriculum in school districts. The authors of this talk were centrally involved in the 2003 revision of New Mexico's science standards. According to the Fordham Foundation's 2005 review of 50 state science standards, the 2003 revision took New Mexico's standards from among the worst in the nation to seventh best. Summarizing that revision, this talk will challenge the audience to consider what every American should know about physical acoustics, science, and scientific methods, and how government-established standards for K–12 education can support such goals. Examples of the relationship between core science standards and specialized subjects, such as acoustics, will be given, and the measurable effect of standards-based curriculum development on student achievement in one New Mexico school district will be described. The influence of those who mix science and religion will also be discussed, with particular attention to testable hypotheses in science and to lessons learned while revising the New Mexico science standards.

Contributed Papers

2:25

2pED5. Introduction of therapeutic ultrasound projects into undergraduate biomedical engineering curriculum at the George Washington University. Vesna Zderic (Dept. of Elec. and Comput. Eng., The George Washington Univ., 801 22nd St. NW, Washington, DC 20052)

Therapeutic ultrasound examples and projects have been successfully incorporated in both formal classes and undergraduate research at the George Washington University. Topics appropriate for various levels of experience and knowledge had to be carefully selected. These included: General demonstrations for summer high school students from local public schools, well-defined sophomore research projects, and finally capstone projects that focus on device design and development. Demonstrations of therapeutic ultrasound ranged from application of high-intensity focused ultrasound (HIFU) to make precise holes in plastic containers and cook meat (showing thermal effects), to production of water fountains (showing streaming effects) and production of dancing (standing wave pattern) bubbles that outline the shape of focus (showing cavitation effects). While very entertaining for younger students, these demonstrations also allow an explanation of basic ultrasound theory and potential clinical applications. Some examples of sophomore research in the therapeutic ultrasound area are: Ultrasound-enhanced delivery of ocular antibiotics, HIFU treatment of deep seated tumors, and studying effects of the presence of air or bone in the post-focal HIFU region. Examples of capstone projects include: Portable ultrasound driving system, cooling system for HIFU transducers, robotic HIFU control, flow phantom for testing of ultrasound-guided HIFU treatments, and system for measurements of ultrasound powers.

2:40

2pED6. My voice looks like that? Analysis and synthesis of vowel sounds in a PC-based lab. Andrew A. Piacsek (Dept. of Phys., Central Washington Univ., 400 E. Univ. Way, Ellensburg, WA 98926)

How is it that we can recognize familiar vowel sounds regardless of who is speaking? How does my voice compare to your voice? The acoustics laboratory exercise described here appeals to our natural curiosity about everyday phenomena and about ourselves. In the first part, students use a PC equipped with a built-in microphone and easy-to-use sound analysis software to record themselves ''singing'' a variety of vowel sounds, each with the same steady pitch. With the spectrum of each vowel displayed, students initially confirm the presence of the harmonic series and identify the fundamental frequency, reinforcing the association of these characteristics with the pitch of the vowel sound. Working in small groups, students then investigate a correlation between the specific vowel and its spectrum envelope. By comparing the spectra of similar vowels sung by different students, as well as the spectra of different vowels, students learn to recognize formants within the spectrum. In the second part, students apply their newfound understanding of the role of formants by attempting to synthesize a vowel sound by adding harmonic components. The sound synthesis is also done with commonly available PC software.

2:55–3:35
Panel Discussion
Peggy B. Nelson, Moderator

Session 2pMU

Musical Acoustics and Speech Communication: Musical Pitch Tracking and Sound Source Separation Leading to Automatic Music Transcription II

James W. Beauchamp, Chair
Univ. of Illinois Urbana-Champaign, School of Music, Dept. of Electrical and Computer Engineering, Urbana, IL 61801

Chair's Introduction—1:30

Invited Papers

1:35

2pMU1. Active music listening interfaces based on sound source separation and F0 estimation. 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 describes research aimed at building *"active music listening interfaces"* to demonstrate the importance of music understanding technologies, including sound source separation and F0 estimation, and the benefit they offer to end users. Active music listening is a way of listening to music through active interactions. Given polyphonic sound mixtures taken from available music recordings, our active music listening interfaces enrich end-users' music listening experiences. For example, by suppressing drum sounds and adding other drum sounds, *Drumix* [Yoshii et al., IPSJ Journal, **48**(3), (2007)] enables a user to change the volume and timbre of drum sounds separated from the mixtures and rearrange rhythmic patterns of these drum sounds during playback. In addition, during the playback of a song, *LyricSynchronizer* [Fujihara et al., Proc. of IEEE ISM 2006] displays scrolling lyrics and highlights the phrase currently sung. Because lyrics are automatically synchronized with vocal melodies (separated from polyphonic sound mixtures using our predominant-F0 estimation method *PreFEst*), a user can easily follow the current playback position and click on a word in the lyrics to listen to it. These interfaces can also be regarded as *"augmented music-understanding interfaces"* that facilitate deeper understanding of music by end users.

1:55

2pMU2. Timbre modeling for the purpose of sound source recognition and separation in music. Anssi Klapuri (Inst. of Signal Processing, Tampere Univ. of Technol., Korkeakoulunkatu 1, 33720 Tampere, Finland)

Timbre models of musical instruments were studied in order to recognize sound sources in music and to organize sounds to their sources. The log-power spectrum of the sounds was modeled using a combination of linear bases, one defined as a function of frequency, and another as a function of harmonic index. Algorithms for initializing, learning, and sequentilly adapting this model are described. In simulations, the model had a benefit over conventional Mel-cepstral representations, especially in the case of certain instruments, including the electric guitar, the organ, the mallet percussions, and the clarinets. The improvement was observed in supervised instrument classification, but not so clearly in the unsupervised clustering of sounds.

2:15

2pMU3. Supervised separation of musical sources. Paris Smaragdis (Mitsubishi Electric Res. Labs, 201 Broadway, 8th Fl., Cambridge, MA 02139, paris@media.mit.edu)

Sounds, especially in the musical context, are often presented in mixtures. Given that traditional signal processing theory is not well equipped to work with concurrent sound sources, this limits the analysis and creative options on musical audio. As it was recently shown, sound mixtures can be analyzed with latent variable models of time-frequency distributions and, thus, reveal the additive structure of acoustic scenes. Unfortunately, such processes separate acoustic elements in an indiscriminant manner that does not always result in extracting the desired sources. However, in the case where a user can provide examples of the sources to extract, the aforementioned algorithms can become powerful tools for supervised source separation. Performing experiments on real single-channel recordings, it was shown that by simply indicating the regions of time where a particular source was active, it is possible to extract that source with minimal distortion, even though it may constantly be part of a sound mixture. Since this is an example-based procedure, it is able to deal with arbitrary sources, regardless of their acoustic properties, without requiring any tedious heuristic modeling. This extraction method facilitates the analysis of musical data, and also allows creative manipulation of music.

2:35

2pMU4. A framework for sound source separation using spectral clustering. George Tzanetakis, Mathieu Lagrange, Luis Gustavo Martins, and Jennifer Murdoch (Dept. of Comput. Sci., Univ. of Victoria, Canada, gtzan@cs.uvic.ca)

Clustering based on the normalized cut criterion, and more generally, spectral clustering methods, are techniques originally proposed to model perceptual grouping tasks, such as image segmentation in computer vision. In this work, it is shown how such techniques can be applied to the problem of dominant melodic source separation in polyphonic music audio signals. One of the main advantages of this approach is the ability to incorporate mutiple perceptually-inspired grouping criteria into a single framework without requiring multiple processing stages, as many existing computational auditory science analysis approaches do. Experimental

results for several tasks, including dominant melody pitch detection, are presented. The system is based on a sinusoidal modeling analysis front-end. A novel similarity cue based on harmonicity (harmonically-wrapped peak similariy) is also introduced. The proposed system is data-driven (i.e., requires no prior knowledge about the extracted source), causal, robust, practical, and efficient (close to real-time on a fast computer). Although a specific implementation is presented, one of the main advantages of the proposed approach is its ability to utilize different analysis front-ends and grouping criteria in a straightforward manner.

2:55

2pMU5. Methods for stereo music source separation. Andreas Ehmann (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 1406 W. Green St., Urbana, IL 61801)

Musical source separation from stereo signals has been growing in popularity recently. Algorithms (including the author's) and the results of a recent stereo audio source separation evaluation campaign (SASSEC) will be presented. In general, such approaches involve extracting mixing parameters from time-frequency (T-F) analyses of both channels. Each corresponding T-F point of the two channels is compared and interchannel intensity differences (IID) and (in the case of stereo mixing), interchannel phase differences (IPD) are computed. A histogram of signal power in the IID and IPD space is constructed, with sources showing strong peaks at the locations of their mixing parameters. Masks are then created for each potential source, corresponding to T-F points that share common mixing parameters. These masks are applied to the T-F analysis and resynthesized as probable sources. The underlying assumption of such systems is that of W-disjoint orthogonality, that is, T-F points of the sources do not overlap. In music, however, this is rarely the case, as common musical intervals often contain overalapping harmonic ''collisions.'' The use of an exponentially damped sinusoidal model, that in limited cases has the ability to resolve very closely spaced, beating, sinusoids is explored as a means of separating harmonic collisions.

3:15–3:25 Break

3:25

2pMU6. Algorithm for separating vocals from polyphonic music. Tuomas Virtanen and Matti Ryynänen (Tampere Univ. of Technol., P.O. Box 553, FI-33101 Tampere, Finland, tuomas.virtanen@tut.fi)

An algorithm for the separation of vocals from polyphonic music is described. The algorithm consists of two stages, which first estimate the predominant melody line, and then the sinusoidal modeling parameters corresponding to the melody line. The melody line is estimated using a hidden Markov model where the output of a multiple fundamental frequency estimator is used as a feature set. The states of the hidden Markov model correspond to musical notes having different fundamental frequencies, and the state transition probabilities are determined by a musicological model. The sinusoidal modeling stage estimates the frequency, amplitude, and phase of each overtone of the predominant melody line in each frame. The sinusoidal modeling stage can also include mechanisms which reduce the effect of interfering sound sources on the estimated parameters. The resulting separation algorithm is independent of singer identity, musical genre, or instrumentation. Simulation experiments on real polyphonic music show that the algorithm enables separation of vocals from the accompaniment, providing robust results on various musical genres.

3:45

2pMU7. Separation of singing voice from music accompaniment for monaural recordings. Yipeng Li and DeLiang Wang (Dept. of Comput. Sci. and Eng., The Ohio State Univ., Columbus, OH 43210)

Separating singing voice from music accompaniment is useful in a wide range of applications, such as lyrics recognition and alignment, singer identification, and music information retrieval. Compared to speech separation, which has been extensively studied for decades, singing voice separation has been little explored. We have developed a pitch-based system to separate singing voice from music accompaniment for monaural recordings. Our system consists of three stages. The singing voice detection stage partitions and classifies an input into vocal and nonvocal portions. For each vocal portion, the predominant pitch detection stage detects the pitches of the singing voice. Finally, the separation stage uses the detected pitch to group the time-frequency segments of the singing voice. Quantitative results show that the system performs the separation task successfully. [The work is supported by AFOSR and AFRL.]

4:05

2pMU8. Statistical models for music signal analysis and transcription. A. Taylan Cemgil (Dept. of Eng., Univ. of Cambridge, Trumpington St., CB2 1PZ Cambridge, UK)

In recent years, there has been an increasing interest in statistical approaches and tools from machine learning, for the analysis of audio and music signals. The application of statistical techniques is quite natural: Acoustical time series can be conveniently modeled using hierarchical signal models by incorporating prior knowledge from various sources: From physics or studies of human cognition and perception. Once a realistic hierarchical model is constructed, many audio processing tasks, such as coding, restoration, transcription, separation, identification, or resynthesis can be formulated consistently as Bayesian posterior inference problems. This contribution illustrates various realistic generative signal models for audio and music signal analysis. In particular, factorial switching state space models, Gamma-Markov random fields, and point process models will be discussed. Some models admit exact inference, otherwise, efficient algorithms based on variational or stochastic approximation methods can be developed. We will illustrate the approach on music transcription, restoration, and source separation applications. [Work supported by EPSRC.]

4:25

2pMU9. Approaching polyphonic transcription of piano sounds. Luis I. Ortiz-Berenguer (Audiovisual and Commun. Eng. Dept., EUITT, Polytechnic Univ. of Madrid (UPM), Ctra Valencia km7, Madrid, Spain) and Francisco J. Casajus-Quiros (Polytechnic Univ. of Madrid (UPM), Ciudad Universitaria s/n. Madrid, Spain)

Polyphonic transcription of piano is a challenging task due to the specific characteristics of its sound spectrum. Pattern matching method, which compares the spectrum with a set of spectral patterns, has proven to give good results, although some limitations still exist mainly when analyzing notes of the lower octaves. Present research is oriented to improve both the accuracy of the spectral patterns, especially for lower octaves, and the matching-score calculation. The former is being carried out by developing a model of the vibration of piano. The last is aimed to decide whether a simple internal product or a more complex MSE calculation are needed and how the spectral pattern has to be modeled regarding partial levels. The model of the vibration permits calculating the partial frequencies, for which it takes into account the vibration of the stiff string and the effect of the soundboard impedance. The soundboard impedance modeling is under a deeper study using both real measurements and FEM modeling. As further work, once the model is completed, it is expected to generate the patterns by using a waveguide synthesis method. [Work supported by Spanish National Project TEC2006-13067-C03-01/TCM.]

Contributed Paper

4:45

2pMU10. Employing the mutual phase coherence of the overtones in musical sounds for audio source separation and reconstruction. Gordana Velikic and Mark F. Bocko (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Rochester, NY 14627)

Source separation from a monaural mixture of musical sounds is challenging. When only a single mixture of multiple sources is available, the method of independent component analysis may not be straightforwardly applied, and proposed source separation methods have relied on previous knowledge of the independent sources. Our method employs mutual correlations of the overtone phases of musical sounds to ascribe spectral features to individual sources. The phase versus time is computed from the analytical signal representation of the output of a set of adaptive band-pass filters that perform initial spectral separation. A small number of cleanly separable partials, usually the fundamentals, are then employed as references to which the remaining filter-bank outputs are correlated. To separate overlapping spectral features and ascribe the proper amount of power to each source, the phase of the reference signal is scaled to the harmonic frequency, and the amplitude of each overtone is estimated by computing the cross-correlation of the phase trajectories. To group sonic events in time, we employ the history of the signal itself, thus eliminating the need for prior information. We discuss this separation method and the audio quality of the reconstructed source signals, with examples and a quantitative perceptual metric.

WEDNESDAY AFTERNOON, 28 NOVEMBER 2007 GRAND BALLROOM E, 3:00 TO 4:40 P.M.

Session 2pNS

Noise: Rain Noise

Robert C. Coffeen, Chair

Univ. of Kansas, School of Architecture and Urban Design, Lawrence, KS 66045

Chair's Introduction—3:00

Invited Papers

3:05

2pNS1. A standard way of measuring rainfall noise. Richard J. Peppin (Scantek, Inc., 7060 Oakland Mills Rd., #L, Columbia, MD 20146) and Robert C. Coffeen (Univ. of Kansas, Lawrence, KS 66045)

About 10 years ago at an ISO TC 29 meeting, the concept of the measurement of rainfall noise was suggested. At the time it made no sense to me. Who cares about this? Not me, since I liked the sound of rain on my roof, and the louder the better. The concept was to evaluate roofs to assess the insulation capabilities from sound (vibration?) produced by rainfall. But in spite of my skepticism, wiser heads prevailed, and in the end, the approach was drafted into a work item, a draft, and a final draft standard. This paper is a summary of ISO FDIS 140-18, ''Acoustics—Measurement of sound insulation in buildings and of building elements—part 18: Laboratory measurement of sound generated by rainfall on building elements.'' I discuss the equipment needed, the test method, and the uncertainty.

3:25

2pNS2. Rain noise level measurements in a performing arts center. Robert C. Coffeen (School of Architecture and Urban Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

Noise level measurements made in a performance hall and in related spaces with subjectively described light, moderate, and heavy rainfall on the facility roof are presented in octave frequency bands, and as described by associated A weighted sound pressure levels and noise criteria (NC) ratings. Rain noise levels are also compared with HVAC noise and ambient noise without the HVAC system operating.

3:50

2pNS3. Waterfall noise. Jerry Lilly (JGL Acoust., Inc., 5266 NW Village Park Dr., Issaquah, WA 98027)

Noise generated by natural and man-made waterfalls is of considerable interest to acoustical consultants, who are looking for reliable sources of outdoor masking noise. Being able to estimate noise levels from water features prior to construction is also important. This presentation discusses field measurements taken at waterfalls with flow rates ranging from 0.2 to 20,000 cubic ft per s and drop heights ranging from 1 to 270 ft. Octave band spectra and time variations will be shown. Empirical expressions for estimating octave band and *A*-weighted sound power levels will also be presented.

4:15

2pNS4. Comparison of interior noise levels produced by rain impinging on several commercial roof constructions. Robert C. Coffeen (School of Architecture & Urban Planning, Univ. of Kansas, 1465 Jawhawk Blvd., Lawrence, KS 66045, coffeen@ku.edu)

Noise produced within architectural spaces by rain on roofs is difficult to quantify, due to varying rain intensity, water drop size and velocity, roof construction, and interior acoustical characteristics. However, it is possible to compare interior rain noise levels when rain conditions and the interior space conditions are constant. This paper compares rain noise levels for several commercial roof constructions with noise level measurements and rain generation in general conformance with ISO 140-18 ''laboratory measurement of sound generated by rainfall on building elements.'' [This research was conducted through the School of Architecture and Urban Planning at the University of Kansas.]

WEDNESDAY AFTERNOON, 28 NOVEMBER 2007 BORGNE, 1:30 TO 4:30 P.M.

Session 2pPA

Physical Acoustics: Bubbles and Cavitation

Ronald A. Roy, Chair

Boston Univ., Dept. of Aerospace and Mechanical Engineering, Boston, MA 02215

Contributed Papers

1:30

2pPA1. Characterization of a large volume spherical resonator for studies of acoustically induced cavitation in liquids. Jason L. Raymond, Kenneth Bader, Joel Mobley (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Univ., MS 38677), D. Felipe Gaitan, Robert A. Hiller, and Ross A. Tessien (Impulse Devices, Inc., Grass Valley, CA 95945)

This paper describes a spherical acoustical resonator system used to study acoustic cavitation phenomena in liquids as part of an effort to scale up the energy density of collapse of transient cavitation. The resonator is formed by a stainless steel spherical shell 24.1 cm in diameter (OD) and either 1.27 cm or 1.90 cm thick designed for generating transient cavitation at high static pressures. An external transducer attached to the surface of the resonator was used to excite an acoustic standing wave in the liquid in order to generate a pressure maximum near the center of the liquid. We will present the results of our characterization of this device, including hydrophone measurements of the acoustic pressure generated in the liquid and vibration analysis on the surface of the resonator carried out using laser Doppler vibrometry. The resonance frequency spectrum and modal structure are compared to numerical predictions using theory developed by (Mehl, JASA **78**(2), 782–788 (1985)) [Work supported by SMDC Contract No. W9113M-07-C-0178.]

1:45

2pPA2. Optical spectroscopy of transient cavitation at high pressure. Robert Hiller, Felipe Gaitan, Ross Tessien, and Jeff Alstadter (Impulse Devices, Inc., 13366 Grass Valley Ave., Unit H, Grass Valley, CA 95945, hiller@impulsedevices.com)

The UV-VIS spectrum of light emitted from transient cavitation events in a high-pressure acoustic resonator is measured to determine the contents and state of matter in the plasma hot spot at the collapse. Spectra are measured for spontaneous cavitation events in water and other liquids at static pressure up to 300 bar, both degassed and with various gases dissolved in the liquid. Spectra are captured with a grating monochromator using a gated detector, or multiple detectors with bandpass filters, to partially resolve the temporal structure of the emission event. These data are used to estimate the temperature and duration of the implosion. [Work supported by SMDC Contract No. W9113M-07-C-0178.]

2:00

2pPA3. Computer simulations of cavitation collapses at high static pressure using hydrocode HYADES: Plasma conditions and shock waves in the liquid. D. Felipe Gaitan, Ross A. Tessien, Robert A. Hiller, and Jeff Alstadter (Impulse Devices, Inc., 13366 Grass Valley Av., Unit-H, Grass Valley, CA 95945)

Previously (JASA **121** pt.2, p. 3181) transient cavitation in high pressure spherical resonators was described. The phenomenon is characterized by cavitation events lasting a few ms (100 s acoustic cycles) in which a cavity (or cavities) collapse violently, emitting flashes of light of widths $\sim$1–40 nsec and spherical shock waves with amplitudes $\sim$1–100 bars (depending on the static pressure) at a distance of 10 cm from the collapse. Both SL and shock amplitudes are proportional to the static pressure. The goal of these experiments is to investigate how the static pressure increases the intensity of acoustic cavitation collapse and whether thermonuclear fusion reactions are possible this way. The determination of plasma temperatures, pressures, and densities is therefore critical in making progress. Furthermore, the amplitude and velocity of the shock waves can be used to determine the conditions at stagnation (end of collapse), assuming the cavity size at that time is known. To this end, results of numerical simulations using HYADES plasma physics hydrocode will be presented, with emphasis on the plasma conditions versus the amount of gas in the cavity and the static pressure, as well as the amplitude and velocity of the outgoing shock waves generated after the collapse [Work supported by SMDC Contract No. W9113M-07-C-0178.]

2:15

2pPA4. Measurement of shock velocity from cavitation collapse in high static pressure liquids using Schlieren and shadowgraphy. James Schaefer, Ross A. Tessien, and D. Felipe Gaitan (Impulse Devices Inc., 13366 Grass Valley Ave., Unit H, Grass Valley, CA 95945, jschaefer@impulsedevices.com)

High amplitude shock waves due to collapse of a cavitation bubble were photographed using a high power pulsed laser diode driven with an avalanche transistor pulser. The strong implosions of high pressure bubble cavitation travel initially at several times the local speed of sound. This method was used to measure shock wave velocity in water at static pressures up to 300 bars. The photographs were taken between 30 and 60 ns of the bubble implosion. Schlieren and shadowgraphy techniques were used and compared. The purpose is to characterize the pressure conditions that exist shortly after the implosion of a high pressure sonoluminescence bubble. [Work supported by SMDC Contract No. W9113M-07-C-0178.]

2:30

2pPA5. Time of flight measurements for strong shocks emitted from transient cavitation in 300 bar resonator. Ross A. Tessien, James Schaefer, Robert A. Hiller, D. Felipe Gaitan, and Henry Tardif (Impulse Devices Inc., 13366 Grass Valley Ave., Unit H, Grass Valley, CA 95945)

The behavior of shocks emitted from transient cavitation in acoustic cavitation systems with static pressures up to 300 bar is explored using photomultiplier tubes, hydrophones, Schlieren, and Shadowgraphic methods. Shocks emitted from strong implosions travel initially at supersonic velocities, leading to a shortening of the time of flight (TOF) from implosion site to detector as compared to sound speed motions. A larger difference between expected sonic and actual supersonic translation times is indicative of higher stagnation pressures. Experimental results are indicative of possible stagnation pressures in the Mbar range and place constraints on computer models. [Work supported by SMDC Contract no. W9113M-07-C-0178.]

2:45

2pPA6. Morphology of transient cavitation in 300 bar resonator. Ross A. Tessien, Henry Tardif, Robert A. Hiller, James Schaefer, and D. Felipe Gaitan (Impulse Devices Inc., 13366 Grass Valley Av., Unit-H, Grass Valley, CA 95945)

Details of the observed evolution of a high pressure transient cavitation event are described. The combination of photomultiplier tubes, hydrophones, fiber optic hydrophone, laser to photo diode light blocking methods are used to explore the evolution of a transient bubble cloud event from ns to ms time scales. [Work supported by SMDC Contract No. W9113M-07-C-0178.]

3:00–3:15 Break

3:15

2pPA7. A model for the interaction of clustered bubbles and particles under acoustic excitation. Todd A. Hay and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

An analytical model has been presented previously for the interaction between a spherical bubble and a spherical particle [e.g., Hay and Hamilton, J. Acoust. Soc. Am. **120**, (2006)]. The present work extends this model to the case of a cluster of spherical bubbles and particles that are free to pulsate and translate. The bubbles and particles may be of arbitrary size, the particles may be elastic or rigid, and the external acoustic source excitation is included to an order consistent with the accuracy of the model. Effects of liquid compressibility are included in the equations of radial motion, and bubble coalescence is taken into account according to a set of conservation relations. Numerical simulation results will be presented for a number of cluster configurations and source excitations relevant to practical applications, such as shock wave lithotripsy. [Work supported by the NSF Graduate Research Fellowship.]

3:30

2pPA8. Acoustic and vibratory emissions from bubbles collapsing near boundaries. Nicholas J. Manzi, Sean Geary, Allison Squires, Parag V. Chitnis, Robin O. Cleveland, R. Glynn Holt, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The collapse of transient bubbles in the vicinity of a boundary was investigated. Transient bubbles were created using a focused laser, and a shock-wave lithotripter. Two boundary types were investigated, rigid and smooth, and rigid and porous. The porous boundary was either gas-saturated, or actively bubbling. Acoustic emissions from the collapsing transient bubbles were measured with a hydrophone. Boundary vibrations were (in some cases) measured using a laser vibrometer. Images of the collapse were also obtained from a high-speed camera. These diagnostics were used to investigate the threshold for and dynamics of the well-known asymmetric jetting collapse of a bubble near a boundary, by varying the distance of the bubble from the boundary. The possibility of using an acoustic or vibration-based measurement as a diagnostic for boundary damage from bubble collapse is considered. [Work supported by ORNL.]

3:45

2pPA9. Resonant bubble effects in void fraction measurements from modal frequencies of a bubbly impedance tube. Christopher E. Ormonde, Seth B. Cohen, Parag V. Chitnis, Robin O. Cleveland, Ronald A. Roy, and R. Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The effects of resonant bubbles on the measurement of void fraction in a cylindrical waveguide are examined. Air bubbles are introduced into a steel cylindrical waveguide containing a Xanthan gum gel. The frequency response of the system is obtained. Sound speeds are inferred by assuming a linear dependence of axial mode number on mode frequency, and void fractions are calculated assuming a mixture sound speed for a bubble population much smaller than the resonant sizes in the modal frequency range (Woods limit). Results from experiments with various bubble sizes and varying number of bubbles are presented, including single bubbles, two differently sized bubbles, and multiple bubbles of roughly the same size. Local dispersion effects are exhibited in the region of bubble resonance for bubbles that are large enough to violate the Woods limit assumption. The effect of such rogue bubbles on the accuracy of the void fraction measurement in a system with unknown bubble size is considered. [Work supported by ORNL.]

4:00

2pPA10. Acoustic phase speed measurements in a bubbly liquid. Stanley A. Cheyne and Robert H. Hembree (Dept. of Phys. and Astron., Hampden-Sydney College, Hampden-Sydney, VA 23943)

Acoustic phase speed measurements of a bubbly liquid have been made using hydrophones. Standing wave and time-of-flight methods were used in both a cylindrical tube and a larger tank. A standing-wave technique similar to [Silberman, J. Acoust. Soc. Am. **29** (1957)] was used to measure sound velocity at low void fractions ($<2\%$). Above 2% the bubble size distribution was not uniform, making it more difficult to get reliable results. Bubble clouds were produced using a variety of arrays of hypodermic needles.

4:15

2pPA11. Correlation of vortex-induced cavitation bubble dynamics and acoustic emissions. Natasha A. Chang, Ryo Yakushiji, David R. Dowling, and Steven L. Ceccio (Univ. of Michigan, 2350 Hayward St., 2212 G. G. Brown Lab., Ann Arbor, MI 48109)

The bubble dynamics and corresponding acoustic emission from low-event rate cavitation inception due to the interactions of a pair of parallel counter-rotating vortex was studied experimentally in a water tunnel. The measured bandwidth of the acoustic pulse from the growth/collapse of a

small isolated cavitation bubble is more than 200 kHz, and the measured pulse duration is ~15–20 μs. The underlying vortical flow, static pressure, and nuclei distribution were characterized and are reported. These fluid and flow parameters influenced the acoustic signal, and dynamics of the cavitation bubbles. Details of the acoustic signature of the cavitation bubble were investigated during its inception, growth, splitting, and collapse. In the chosen flow field, it was found that during bubble growth, the acoustic signal is the strongest with the bulk of the signal energy in frequencies between 1 kHz and 6 kHz. Here, the frequency content of the acoustic signal during inception and growth was related to the volumetric change of the bubble measured from images taken with a high-speed video camera with a correlation of 84%. [Work supported by ONR.]

WEDNESDAY AFTERNOON, 28 NOVEMBER 2007 NAPOLEON D2/D3, 1:30 TO 4:30 P.M.

Session 2pSC

Speech Communication: Consonant and Vowel Issues in Communication (Poster Session)

Amee P. Shah, Chair
Cleveland State Univ., Dept. of Speech and Hearing, Cleveland, OH 44115

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

2pSC1. Second mention reduction in Indian, English, and Korean. Rachel E. Baker and Ann R. Bradlow (Dept. of Linguist, Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208-4090, r-baker2@northwestern.edu)

Phonetic reduction of the second relative to the first mention of a word in a discourse (Second Mention Reduction, or SMR) is a well-documented feature of spoken American English (AE). This phenomenon could be mediated by discourse-level prosodic structure: second mentions are reduced because they are less likely than first mentions to be accented in AE; or it could occur independently of prosodic structure. Impressionistic studies indicate that Indian English (IE) speakers do not deaccent second mentions like AE speakers [J. Gumperz, Discourse Strategies (1982)]. Korean marks information structure through phrasing and word order rather than pitch accents. Therefore, SMR in these languages would support a direct link between lexical probability and duration. Six IE speakers read five paragraphs containing 59 repeated words, and five Korean speakers read a Korean translation of one of these paragraphs. IE speakers were significantly less likely to deaccent second mentions than AE speakers (sign-test, $p<.05$), and four IE speakers had identical accenting rates for both mentions, providing empirical evidence that IE speakers do not deaccent second mentions. Moreover, significant SMR was found in both Korean ($W=43$, $p<.05$) and IE ($W=1340$, $p<.001$), suggesting that probabilistic effects on word duration are not mediated by prosodic prominence.

2pSC2. Effects of word length and presentation order in the perception of non-native sequences. Lisa Davidson (Dept of Linguist., New York Univ., 726 Broadway, 7th Fl. NY, NY 10003, lisa.davidson@nyu.edu)

Previous research in the perception of non-native phonotactics indicates that listeners have difficulty distinguishing certain sequences like [ebzo]-[ebuzo] (Japanese listeners) or [zgamo]-[zəgamo] (English listeners). Whereas discrimination of phonemes in monosyllables (e.g. [k'a]-[q'a]) is diminished but better than chance [J. Werker and R. Tees J. Acoust. Soc. Am. 75, 1984], discrimination of non-native sequences in longer words is often below chance. In this study, the effect of length and presentation order on the discrimination of non-native phonotactics is investigated. English speakers ($n=30$) were presented with pairs of the type CCVCV-CəCVCV (e.g., /tkapu-təkapu/) and CCV-CəCV (/tka-təka/) in an AX discrimination task. Half of the ''different'' pairs were presented with the ə word first, and half with the cluster word first. Results indicate that English listeners are significantly less accurate on the discrimination of longer words both in ''same'' and ''different'' trials. Presentation order was also significant: listeners are more accurate when the cluster word is presented first than when the ə word is. These results suggest that while adults can discriminate non-native sounds or sequences under optimal contexts, this ability is not necessarily helpful in situations with more realistic input, such as language acquisition. [Research supported by NSF.]

2pSC3. Sensitivity to phonetic detail in perceptual assimilation: The perception of Korean stops by native speakers of Hindi. Jenna Silver (Dept. Comm. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611), Ratree Wayland (Program in Linguist., Univ. of Florida, Gainesville, FL 32611), and James Harnsberger (Inst. for Adv. Study of the Commun. Proc., Univ. of Florida, Gainesville, FL 32611)

The study examines the perceptual assimilation of six Korean stop contrasts (/t/-/th/, /t/-/t*/, /th/-/t*/, /k/-/kh/, /k/-/k*/, /kh/-/k*/) by native speakers of Hindi to explore the question: Does variable assimilation of non-native speech sounds to multiple native perceptual categories reflect 1) an insensitivity to non-native phonetic detail or 2) a residual sensitivity that results in better (although not nativelike) perception? The Korean series was presented in both a categorial AXB discrimination test and an orthographic classification task. The Korean consonants were variably identified by Hindi listeners both in terms of place of articulation (Korean dentialveolar stops were variably assimilated to the dental and retroflex places in Hindi) as well as voicing (Korean tense stops were variably judged as either voiced or voiceless unaspirated; Korean lax stops were judged as either voiceless unaspirated or voiceless aspirated). A comparison with the discrimination scores supports the residual sensitivity hypothesis: The most variably assimilated Korean contrasts also elicited discrimination scores that were significantly above chance. Such residual sensitivity reflects an adult perceptual system in which sensitivity to non-native phonetic detail is not lost, but maintained and available to aid in second language learning.

2pSC4. A methodological note on signal-to-noise ratios in speech research. Kristin J. Van Engen (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

In speech-in-noise tests that utilize a range of signal-to-noise ratios (SNRs), SNR is usually manipulated either by changing the noise level relative to a constant signal level or by changing the signal level relative to a constant noise level. In the HINT, for example, noise level is held constant while target sentence level changes. Such manipulations entail differences in overall energy delivered to the ear across SNR conditions. It is possible, then, that overall energy changes are partially responsible for behavioral differences observed across SNRs. For example, [Van Engen and Bradlow (2007)] showed that English 2-talker babble was more detrimental than Mandarin 2-talker babble for English sentence recognition by native English listeners at difficult SNRs, where SNR was varied by raising the noise level with respect to the target sentence level. In the present study, it is shown that releveling the mixed signal and noise files to a single output level across SNRs yields similar results. Thus, the interaction of the noise language effect with SNR observed in the previous study and the general difference across easy (0 dB) and hard (-5 dB) SNRs cannot be attributed to overall energy changes across SNRs, but rather reveal speech-in-noise processing effects.

2pSC5. Word frequency and neighborhood density effects on L2 listeners: English listeners learning Korean. Sung-A Kim (School of General Education, Kyung Hee Univ., 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Korea, sakim@khu.ac.kr)

This study investigated the word frequency and neighborhood density effects on native and non-native speakers' word recognition. There has been a disagreement among previous studies on word frequency and neighborhood density effects on non-native listening (Bradlow and Pisoni, 1999; Imai et al., 2005; Yoneyama, 2006). In the present study, Korean and English test stimuli with varying word frequency and neighborhood density were presented to English speakers learning Korean, who differed in Korean pronunciation proficiency. Two experiments were conducted: In the first experiment, each language was produced by its native speaker, while in the second one, the Korean stimuli were produced by an English-Korean bilingual, in order to color the stimuli with an English accent. The experimental results revealed a strong influence of word frequency and neighborhood density on word recognition in non-native listening of the English speakers. The participants' pronunciation proficiency turned out to affect word recognition in the non-native listening as well. Yet, English-accented stimuli showed no significant effects on the word recognition of the Korean-learning English speakers across the proficiency levels. These results are discussed in relation to the exemplar model (Bybee, 2000; 2001).

2pSC6. Release direction in clicks correlates with coarticulation patterns. Amanda Miller (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853)

Electropalatograhic data from one speaker's productions of Khoekhoe palatal and alveolar clicks is presented. In alveolar clicks, the anterior part of the posterior constriction lowers first, increasing lingual cavity volume for rarefaction. The apical alveolar anterior constriction is abrupt. These clicks occur more frequently with [ai] than with [i]. Palatal click production exhibits apico-laminal anterior contact from the dental to post-alveolar regions, velar contact, and a small palatal opening. The anterior release starts with dental separation, followed by alveolar separation. Post-alveolar contact is released last, simultaneously with velar contact forming the front edge of the posterior constriction. The palatal click anterior release is simultaneous with the posterior release, involving a larger contact area and a smaller cavity volume than the alveolar click. Anterior constriction contrasts in these clicks are apical alveolar versus laminal post-alveolar. The post-alveolar laminal click occurs freely with front vowels, while the apical alveolar click occurs freely with [ai], and less frequently with [i]. Anterior and posterior click releases follow the direction of airflow: Front to back. Release direction correlates with posterior constriction coarticulation patterns. [Work supported by NSF.]

2pSC7. Are types of filled pause discrete classes? Kyle Gorman (Dept. of Linguist., Univ. of Pennsylvania, Philadelphia, PA 19104)

Clark and Fox Tree hypothesize that filled pauses ''uh'' and ''um'' in English serve the discourse function of alerting the interlocutor of production difficulties. They demonstrate that the mean duration of pause immediately after ''um'' is significantly longer than after ''uh.'' This account suggests that fillers are categorically differentiated from each other; however, it fails to predict a cross-linguistic generalization. Across languages, fillers tend to have centralized vowels [Vasilescu et al. (2005)]. When there are two realizations of filled pause within a language, one tends to consist of a vowel followed by a nasal, as in ''um'' [Clark and Fox Tree (2002)]. These observations suggest filled pause is a phonetic process and, therefore, may apply gradiently. English filled pauses are revisited, showing that the discourse effect identified is not categorically distributed, but rather defined acoustically on a gradient scale. Acoustic models of fillers are trained on the identities of ''uh'' and ''um'' in a corpus. These models are then applied to predict the duration of pause. This phonetic account of disfluency performs better than the categorical account in predicting the duration of non-null pause, showing that the discourse effects may be caused by a gradient strategy for disfluency realization.

2pSC8. Articulatory coordination and real-time spoken word recognition. Susie Choi and Dani Byrd (Dept. of Linguist., Univ. of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693)

This study investigates how listeners use information regarding articulatory timing in consonant sequences during lexical access. Research in spoken word recognition of assimilated consonant sequences has shown that listeners make use of detailed variation in the acoustic signal and knowledge of systematic phonological variability [e.g., Gaskell and Marslen-Wilson (1996), Gow and McMurray (to appear)]. However, no earlier work has reported on real-time processing of consonant sequences for which articulatory data is available. This study combines articulator movement-tracking with real-time eye-movement monitoring to relate the articulatory timing properties associated with assimilation with use of the resulting acoustic signal in word recognition. Using magnetometry, words to be used as stimuli in an eye-tracking task are recorded. These contain consonant sequences with a continuum of intergestural overlap obtained by varying prosody and speech rate. In the eye-tracking experiment, listeners hear two-word stimuli such as ''bad ban,'' some of which are perceived with an assimilated final consonant in the first word, and select a picture that matches the stimulus. The recorded eye movements provide a window into real-time processing [see Altmann and Kamide (2004)], allowing us to investigate listener behavior in using articulatory timing cues for lexical access. [Supported by NIH.]

2pSC9. Italian geminates: Interarticulator programming and articulatory dynamics. Anders Lofqvist (Haskins Labs., 300 George St., New Haven, CT 06511)

This study examined lip and tongue coordination and tongue movement kinematics in Italian long and short consonants with native speakers of Italian as subjects. The linguistic material consisted of Italian word pairs that only differed in the duration of a labial or lingual consonant, which was either long or short. Recordings were made of lip and tongue movements using a magnetometer system. There was a very robust difference in the duration of the oral closure. The duration of the tongue body movement between the two vowels in a sequence of vowel-bilabial nasal-vowel was longer, and its movement path longer, for a long than for a

short consonant; the average speed of the movement showed no reliable difference. The onset of the tongue movement relative to the oral closure started earlier for a short than for a long consonant. In lingual stop and fricative consonants, the tongue movement during the closure was longer, and its average speed slower, in the long than in the short consonants. Overall, these results for Italian are comparable to those found in long and short Japanese consonants. In particular, they provide additional evidence for task-specific modulation of tongue movement speed. [Work supported by NIH.]

2pSC10. The interaction of respiratory drive and linguistic structure. Grant McGuire and Sam Tilsen (Univ. of California at Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, grantmcguire@berkeley.edu)

During high respiratory drive competition between the need for oxygen and the needs of speech breathing are in conflict. This conflict provides a way of exploring linguistic structure in two domains. Syntactically, talkers must breath more frequently such that inhalations occur within sentences, a rarity in normal speech (Bunn & Mead 1971, Grosjean & Collins 1979, Doust & Patrick 1981, Bailey & Hoit 2002). Phonetically, high respiratory drive provides competition between between the high expiratory airflow rates needed to purge carbon dioxide and the high subglottal pressure necessary for phonation. This paper presents acoustic and physiological data from a study where American English talkers are exercised until in high respiratory drive and asked to read a syntactically-controlled passage. Preliminary results suggest that even in high respiratory drive, subjects respect constituency structure such that higher level phrase boundaries are correlated with more frequent inhalations. Further, subjects show less control over vocal fold control and must periodically exhale excess carbon dioxide, resulting in acoustic effects on the preceding word.

2pSC11. Accuracy and response times in identifying foreign accents: Are listeners all over the map? Amee Shah (Dept. of Health Sci., Cleveland State Univ., 2121 Euclid Ave., MC 424, Cleveland, OH 44115)

The present study aims to increase the understanding of how listeners process different foreign accents. Specifically, listeners were presented with 50 varieties of foreign accented speech samples and asked to point to the general area on a world map where they thought the accent was from, and also provide a label/name for the accent. Accuracy and response times were obtained for both methods of accent-identification, and a correlation between these two methods was performed. Errors in identification were further analyzed to investigate the consistent patterns of confusion across listeners' scores. Prosodic similarities across certain types of accents, or geographical proximities across these confusable accents were further studied as potential sources of the consistent patterns of confusion across some accents. Furthermore, listener-related variables, including age, education level, travel experience, linguistic and phonetic experience, and musicality were analyzed further to determine each of their roles in influencing the accuracy in identification of foreign accents. Results help to understand the variables involved in processing foreign accented speech, and crucially, to determine the kinds of segmental, prosodic, and/or extra-linguistic information listeners use to represent and store the different varieties of foreign accents in their memory.

2pSC12. The role of foreign-accentedness in lexical processing. Amee Shah (Dept. of Health Sci., Cleveland State Univ., 2121 Euclid Ave., MC 424, Cleveland, OH 44115) and Conor McLennan (Cleveland State Univ., Cleveland, OH 44115)

Previous evidence indicates that foreign-accented speech influences spoken word recognition (accented stimuli are processed with reduced accuracy and slower response times). The present study aims to extend and further explore the relationship between foreign-accentedness and spoken word recognition. Specifically, the present study strives to determine whether word recognition will modulate the ratings of foreign accentedness. Spoken words are easier to recognize in sentences that are in an appropriate context versus an inappropriate context. In order to determine whether listeners attribute the relatively slow and difficult spoken word processing in an inappropriate context to a relatively strong foreign accent, participants were asked to rate the strength of an accent for the final word in a sentence, as spoken by a native or a non-native speaker. Crucially, half of the words were in an appropriate and the other half in an inappropriate sentence context. Ratings of accent and response times were obtained as a function of appropriate and inappropriate sentence context for both the native and non-native speakers. Results provide new information regarding the relationship between foreign-accented speech, indexical variability, and lexical access.

2pSC13. Language contexts in speech categorization: Testing the double phonetic standard in bilinguals. Adrian Garcia-Sierra, Nairan Ramirez-Esparza, Craig A. Champlin (Univ. of Texas, CMA A2-200, 1 Univ. Station A1100, Austin, TX 78712), and Maritza Rivera-Gaxiola (Univ. of Washington, I-LABS Box 357988, Seattle, WA 98195-7988)

It is widely known that one's native language influences where phonetic boundaries are placed. However, it is less understood how bilinguals categorize speech sounds. The goal of this investigation was to test if bilinguals' phonetic boundary shifts in accordance with the language they use at the moment. In order to accomplish this goal, Spanish-English bilinguals ($N=27$) and English monolinguals ($N=27$) were asked to identify the speech sound "ta" from a 10 token speech continuum varying in VOT in two language contexts, while electrophysiological measures were recorded. It was expected that: (a) Bilinguals, but not monolinguals, would show a shift in their phonetic boundary according to the language context, and (b) bilinguals, but not monolinguals, would show different ERP amplitudes (N2/MMN components) in accordance with the language context. The behavioral results showed that bilinguals, but not monolinguals, perceived specific speech sounds—in the ambiguous zone—differently among language contexts. The electrophysiological results showed that bilinguals', but not monolinguals', ERP components differed across language contexts. Interestingly, behavioral measures correlated significantly with electrophysiological measures only in bilinguals. The challenges of testing the double phonetic boundary are discussed, along with the limitations of the methodology used in this study.

2pSC14. Enhancing learning on foreign-accent adaptation with a combination of active training and passive stimulus exposure. Melissa Baese, Ann Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208), and Beverly A. Wright (Northwestern Univ., Evanston, IL 60208)

Improvement on most perceptual tasks requires active performance of the target task during training. However, there has been little consideration of whether that active performance is necessary throughout the entire practice period. Suggesting that it is not, there is recent evidence that active performance on a simple frequency-discrimination task, interleaved with passive stimulus exposures, can yield more learning than the active performance alone. The current study examines whether these findings extend to a speech-learning task, specifically foreign-accent adaptation. Native-English speakers transcribed sentences spoken in English by non-native speakers. Participants who performed the target task for half the training period and received passive stimulus exposures while performing a distracter task for the other half out-performed participants who received the same amount of active training, but no passive exposures during the distracter task, suggesting that passive exposures enhanced learning. The active-passive group also out-performed participants who received only passive exposures—and performed as well as a group who performed only the target task—for the entire training period. Taken together, these results

suggest that the combination of active training and passive stimulus exposure may play a key role in real-world perceptual learning for speech. [Work supported by NIH/NIDCD.]

2pSC15. A speech audiogram of Setswana consonants. Sinah Y. Seoke [a], Mini N. Shrivastav, and James D. Harnsberger (Univ. of Florida, Gainesville, FL 32611, mnarendr@csd.ufl.edu)

Setswana is a Bantu language used in several regions of Southern Africa, and its phonemic inventory includes a diverse range of consonants. To date, very little information exists about the acoustic correlates of Setswana speech sounds. The present study was aimed at analyzing these speech sounds in order to develop a clinically useful speech audiogram for Setswana. Three native speakers of Setswana produced 45 target phonemes in word initial and medial positions, in three vowel contexts (/a/, /i/, and /u/). The stimuli were then analyzed for peak frequency and intensity. A preliminary analysis revealed an effect of word position and vowel context on peak frequency, and hence, only phonemes in the word medial position and in the /a/ vowel context were used to plot the speech audiogram. The speech audiogram hence developed was then compared to speech audiograms of American English phonemes commonly employed in clinical settings. Consonants common to both Setswana and American English inventories were used to establish the validity of the study's procedures. The speech audiogram developed can potentially be used to provide partial estimates of the audibility of Setswana phonemes to native speakers with various types of hearing loss. [a] Currently at the Dept. of Communication Pathology, Univ. of Pretoria, South Africa

2pSC16. A comparison of speech rate of speakers of Dominican and Puerto Rican Spanish and of New York City English. Leticia Nunez-Agosto, Fredericka Bell-Berti, Altagracia Jorge, and Deborah Reynolds (Dept. Speech, Comm. Sci., & Theatre, St. John's Univ., 8000 Utopia Pkwy., New York, NY 11439)

When we hear speakers of other languages or of other dialects of our own language, we perceive them as speaking more rapidly (or, occasionally, more slowly) than do speakers of our own language/dialect. The purpose of this study was to compare speech rates across dialects and languages to begin to understand these perceived differences. To this end, we compared the speech rate across three groups: 25 speakers each of Dominican and of Puerto Rican Spanish and 25 speakers of English from New York City; all were monolingual speakers of their languages. We measured speech rate for a habitual/automatic speech task, reciting the Lord's Prayer; each speaker repeated the prayer for at least one minute. We measured the durations of the speech and pauses for one repetition of the Lord's Prayer (usually the second) for each speaker. Results will be presented in terms of syllables per second. Preliminary results suggest that speakers of Puerto Rican Spanish and New York English do not have different habitual/automatic speech rates, and that speakers of Dominican Spanish have slower habitual speech rates. [Supported by St. Johns University.]

2pSC17. Augmented visual feedback in second language learning: Training Japanese post-alveolar flaps to American English speakers. June S. Levitt and William F. Katz (Callier Ctr. for Commun. Disord., Univ. Texas at Dallas, 1966 Inwood Rd., Dallas, TX, 75235)

Recent studies suggest beneficial effects of kinematic feedback in the articulatory training of individuals with speech disorders [Katz et al. (2002)]. To better understand the nature of speech motor learning in healthy individuals, the present study examines how adult monolingual speakers of American English learn to produce a non-native consonant: The Japanese post-alveolar flap. Eight participants were randomly assigned to one of two training groups: (1) Conventional L2 training, and (2) conventional L2 training with electromagnetic articulography (EMA)-based kinematic feedback on tongue tip position. Participants completed three baseline recording sessions, eight training sessions, and two post-training sessions (four weeks after training). During training, participants practiced a total of 16 disyllabic non-words containing the post-alveolar flap in word-initial or word-medial position. Participants practiced trained words 20× in blocked fashion, with summarized oral feedback given after every five repetitions. The data were analyzed in terms of flap consonant duration. Preliminary results indicate: (1) Before training, participants produced longer flap durations than native Japanese speakers; (2) with training, flap durations approximated native norms; and (3) kinematic feedback with EMA appears to facilitate the acquisition of novel speech sound production.

2pSC18. The relation between focus effects in production and exemplar locations in perception for stop types in English, Japanese, and Korean. Eric Oglesbee and Kenneth de Jong (Dept. of Linguist., Indiana Univ., 322 Memorial Hall, Bloomington, IN 47405, eoglesbe@indiana.edu)

This paper examines focus-induced variability in the production of labial stop contrasts in English, Japanese, and Korean, pursuing differences between the languages in how voice onset time (VOT) and fundamental frequency (F0) are used to indicate differences between stop types. Language differences are compared to differences in the perceptual usage of these acoustic cues using AMBEL, an adaptive tracking algorithm that allows subjects to search multi-dimensional stimulus arrays for best exemplars. All three languages exhibit category differences in VOT, though Japanese and Korean exhibit overlap for two of the categories. As found previously, for these categories, both Japanese and Korean exhibit large differences in F0 in the following vowel. Except for the smaller F0 differences in English, all of these VOT and F0 differences are systematically expanded in a focus condition. In particular, VOT increases for the more aspirated stops in all three languages, and substantial rising F0 patterns are observed in lenis Korean and zero-lag Japanese stops. Perceptual data match these focus effects; best exemplars differ in VOT in all three languages, and in Japanese and Korean, where substantial F0 focus effects were observed, F0 differences affect the identification of best exemplars. [Work supported by NSF.]

2pSC19. The effect of language experience on the perception of coronal stop bursts. Sameer ud Dowla Khan and Molly Shilman (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095)

The difficulty of perceiving a phonetic contrast not present in one's native language is well known. However, there is also research showing that discrimination of non-native sounds that are sufficiently far from a listener's language is possible even in adulthood, as in the case of English speakers discriminating Zulu clicks [Best et al. (1988)]. This raises the question of whether listeners' ability to discriminate contrasts can hinge on whether listeners are perceiving sounds as speech or non-speech. To test whether language experience affects all auditory perception or only linguistic perception, the current study investigates the ability of Bengali and English speakers to identify isolated Bengali dental and post-alveolar stop bursts in an ABX task, where the bursts are first described as sounds in nature (i.e., non-linguistic context) and later as sounds in language (i.e., linguistic context). It is predicted that, due to the phonemic contrast of dental and post-alveolar stops in their language, Bengali speakers will perform better than English speakers in the linguistic context, and that English speakers will perform better in the non-linguistic context than in the linguistic context.

2pSC20. Effects of native language and training on speaker normalization in lexical tone perception. Ratree Wayland, Edith Kaan, Mingzhen Bao, and Christopher Barkley (Program in Linguist., Univ. of Florida, 4131 Turlingtong Hall, Box 115454, Gainesville, FL 32611-5454)

Tone languages, such as Thai, use pitch differences to distinguish lexical meaning. In this study, we investigated the effects of native language and training on speaker normalization in perception of lexical tones in Thai. Three tokens each of the Thai high, mid, and low tones were generated to simulate three female speakers with three relatively different F0 pitch ranges (high, mid, and low). The stimuli were presented to native Mandarin Chinese (NC), American English (NE), and Thai (NT) for categorization before and after a two-day perceptual training. Results obtained thus far from 32 NC, 30 NE, and 11 NT suggested that, overall: (a) NC outperformed NE both before and after training, (b) NT numerically outperformed NC before training, but not after training, and (c) performance of all three groups of participants improved after training. Interestingly, it was also found that participants' performance varied from tone to tone, depending on the tonal environment in which it was presented for discrimination. These results suggest that knowledge of a tonal language and perceptual training facilitates the speaker normalization process in lexical tone perception, and that tonal contexts affect discrimination of tones.

2pSC21. Mirroring, not imitation, for the early learning of L1 pronunciation. Piers Messum (Dept. of Phonet. and Linguist., UCL, Gower St., London WC1E 6BT, p.messum@gmail.com)

To learn to pronounce, a child must learn to produce speech sounds that listeners take to be equivalent to those that they themselves produce. Once this ''correspondence problem'' is solved, he can learn the pronunciation of any word by imitating the sequence of its constituent sounds. Learning to pronounce has also been assumed to be imitative, but in the different sense of being a mimicry or matching-to-target process applied to the speech signal after some form of normalization for vocal tract size. However, attending to ephemeral words and speech sounds as meaningful unit bars attending to the latter as a signal to be mimicked. Since a child must recognize what is said, basic perceptual psychology provides one reason among several that preclude the possibility of mimicry in the early learning of speech sound qualities. Instead, the normal pattern of mother-infant vocal interaction points to the mother's reformulation of the child's output within ''imitative'' contexts (i.e., vocal mirroring) as enabling the child to establish the correspondences needed. The nature of the representations thus formed resolves, inter alia, the mystery of fast shadowing and the longstanding question of whether speech is fundamentally an acoustic or gestural code.

2pSC22. Incomplete neutralization in Dutch opacity. Marc Ettlinger (Dept of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720)

In Dutch, syllable-final plosives are devoiced so /bad/ ''bathe'' is pronounced [bat]. However, despite being resyllabified to syllable-initial position by a clitic in /bad+er/ ''bathe her,'' the /d/ is still devoiced: [ba.ter]. This is referred to as phonological opacity because a generalization syllable-final devoicing is obscured or rendered opaque by another process—resyllabification. In this experiment, the acoustics and perception of this opaque relationship will be investigated. The voice onset time, release burst, and consonant duration of plosives for native Dutch speakers in phrases such as [ba.ter] will be compared to those in a phrase such as [ba.der] ''bather'' and /prat+er/ ''speaker'' [pra.ter]. Also, a perception experiment will be conducted testing whether native Dutch speakers are able to distinguish these phones. Based on a similar study of German [R. Port and M. O'Dell, J. Phonetics **13**, 455–471 (1985)], it is expected that the [t] in [ba.ter] is not identical to the [t] in [pra.ter]. If, indeed, this is the case, and the [t] in [ba.ter] instead bears some of the acoustic properties of [d] in [ba.der], this would provide a useful set of perceptual cues enabling speakers to observe the devoicing generalization through the opacity.

WEDNESDAY AFTERNOON, 28 NOVEMBER 2007 GRAND BALLROOM D, 1:00 TO 3:45 P.M.

Session 2pUWa

Underwater Acoustics, Acoustical Oceanography, and Signal Processing in Acoustics: Design of Distributed Surveillance and Oceanographic Monitoring Systems

Arthur B. Baggeroer, Cochair
Massachusetts Institute of Technology, Dept. of Mechanical Engineering, Cambridge, MA 02139

Juan I. Arvelo, Jr., Cochair
Johns Hopkins Univ., Applied Physics Lab., National Security Tech. Dept., Laurel, MD 20723-6099

Chair's Introduction—1:00

Invited Papers

1:05

2pUWa1. Distributed remote surveillance systems: An overview of the Naval Studies Board panel report. Arthur B. Baggeroer (Massachusetts Inst. of Technol., Cambridge, MA 02139)

The Chief of Naval Research requested the Naval Studies Board of the Academies to review distiributed remote surveillance (DRS) systems in April 2005. The tasking was to ''conduct an assessment of distributed remote sensing concepts and related technologies'' for improving Navy capability to conduct undersea warfare missions. The panel reported its findings in November 2006. The technical issues covered included: (i) System concepts, (ii) detection and localization, (iii) classification and automation, (iv) communications, (v) power and processing, (vi) distributed network control, and (vii) survivability. Six findings and recommen-

dations were distributed in an NSB report publicly; however, the USN determined that the detailed technical content should be classified. Since DRS is now thought to be a solution for many Navy missions, an enumeration of the technical issues in any successful DRS was felt to be a useful system. The presentation covers these issues. [Work sponsored by N81, USN.]

1:25

2pUWa2. Nested autonomy—A paradigm for distributed, undersea acoustic sensing and surveillance. Michael R. Benjamin, Arjuna Balasuriya, and Henrik Schmidt (MIT, 77 Massachusetts Ave., Rm 5-204, Cambridge, MA 02139)

Undersea surveillance networks using acoustic communication between nodes are severely bandwidth-limited. Equally critical, operational constraints often require a gateway vehicle to surface and quickly relay messages to and from the operator, with typical latencies of 10–30 min. Therefore, in contrast to above-sea systems, the data processing and control of the underwater assets cannot be performed centrally but must be distributed, and the individual and clusters of assets must be capable of autonomously exploiting the environmental acoustics and completing the mission objectives, such as detecting, classifying, localizing, and tracking (DCLT) of a target. The resulting target solution must be assembled in a track report, which must be transmitted back to the operators within the constraint of the communication system. Constrained by small individual acoustic apertures, the net-centric systems may exploit adaptive and collaborative sensing and control for enhancing system performance. Under the ONR PLUSNet program, a nested, autonomous command and control architecture is being developed, with fully integrated sensing, modeling, and control within each autonomous node and clusters of nodes. It enables autonomous adaptation to the environmental and tactical situation, and collaborative target tracking by multiple platforms. The operator controls the autonomous mission behaviors by very simple commands tailored to the limited communication bandwidth. The performance of the nested autonomy paradigm will be demonstrated by examples using a high-fidelity simulation environment, and by results from the PN07 field experiment. [Work supported by ONR.]

1:45

2pUWa3. National Defense Industrial Association distributed netted anti-submarine warfare systems study. Thomas A. Casey (Lockheed Martin Maritime Systems and Sensors, 9500 Godwin Dr., Manassas, VA 20110-4157, thomas.casey@lmco.com)

The NDIA Undersea Warfare Division was tasked by the US Navy, Director Sea Shield, OPNAV, N87 to formulate new DNS concepts to hold enemy forces at risk, to secure a friendly maneuver area, and to assure access world-wide. The new DNS concepts are focused on the Port Egress and Sea Base scenarios, end-points in the spectrum of ASW operations. The threat comprises a large number of widely dispersed, quiet, diesel-electric, AIP, and nuclear submarines in the year 2015. The study team established that new DNS concepts derive advanced capability from expansion of four critical DNS functions: Deployment, sensing, communications, and C2. Seven Port Egress and five Sea Base concepts are each described including: An operating concept, identification of critical components, description of capabilities that each concept offers, and a functional binning diagram. The new DNS concepts were ranked and scored on a set of ASW utility attributes specific to either the port egress or sea base scenario. Concepts were evaluated for potential reductions in the detect to engage sequence and enabling technologies were identified. The study summary, conclusions, and recommendations were presented to the Chairman, ASW Cross-Functional Board, RADM C. Van Mauney on 5 June 2007.

2:05

2pUWa4. Undersea distributed networked systems: A development framework. Raymond J. Christian (Naval Undersea Warfare Ctr., Code HQ10A, 1176 Howell St., Newport, RI 02841, christianrj@npt.nuwc.navy.mil)

Since the early 1990s, the military and commercial undersea communities have experienced dramatic changes in operational, technical, and business climates driven by the confluence of technology advances and increased civilian and military interest for operating in the undersea domain. Undersea distributed networked systems (UDNS) consisting of sensors, unmanned vehicles, platforms, influencers, command/control and, most important, human systems networked to create effects that can be summoned for advantage promise to be a key enabler for military and civilian undersea operations. Presented is a description of the nature and development framework of UDNS, including fundamental functionalities, user paradigm shifts, and operational capabilities tradespace. Traditional system engineering processes are seen to be necessary but not sufficient for engineering UDNS as a complex system; new methods and processes of evolutionary engineering applied to UDNS will be discussed.

2:25

2pUWa5. Achieving a balance in distributed sensor fields. Robert Henrick (Johns Hopkins Univ. Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

The challenges for DNS processing result in a balance between the laws of Moore, Shannon, probability, and physics, and the key limitation of wireless systems: Energy. Moore's law supported by the demand for increasingly sophisticated portable electronics indicates that the amount of computations per watt of power continues to decrease. This not only supports increasingly sophisticated sensor processing, but also sophisticated routing such as ad hoc networking and RF components. Shannon's law provides an invariant set of communication limits, from the amount of power required to send a bit of information to the amount of information in a shared or dedicated RF link. The laws of probabilities imply that to retain a constant field false report rate, false alarm rates per sensor resolution cells (e.g., range, Doppler, . . .) must decrease with the number of sensors, requiring per sensor OR collaborative behavior exploiting differences in target and clutter correlation statistics. Finally, the laws of physics impose constraints key to the system, such as sensor detection ranges, RF Line of site, and physical sensor size. This talk examines these issues in the context of generic distributed systems and recommends areas of promising research.

2:45

2pUWa6. The predictability of signal time modulation, induced by nonlinear internal waves for sound transmissions in the South China Sea. Alban Simon and John Colosi (Dept. of Oceanogr., Naval Postgrad. School, 833 Dyer Rd., Monterey, CA 93943)

Acoustical remote sensing in the shallow water environment must contend with several confounding factors, one of which is scattering from nonlinear internal waves (NLIWs). An important consequence of sound interaction with NLIWs is the time modulation of the signal as the NLIW propagates across the transmission path. Observations and modeling studies from the shallow water acoustic random medium experiment suggest that this modulation may be more deterministic than random. Using observation from the 2001 South China Sea Experiment in which distinct trans-basin NLIWs were observed propagating across the transmission path, the time modulation of low order mode phase and energy is quantified. The observed time modulation of the modes is compared to a simple analytic model to establish the predictability of ISW-induced modulation. If the scattering by NLIWs is not too strong, the model could be a useful tool in surveillance and ocean-monitoring systems.

3:00

2pUWa7. Design criteria for in-buoy geoacoustic inversion using ship noise. Altan Turgut (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

Recent analysis of the ship noise data from R/V Oceanus, collected during the RAGS03 winter and base 2007 summer experiments, showed that robust geoacoustic inversions are possible under both winter and summer conditions. In littoral waters, fast-traveling surface ships provide an excellent (broadband) source of opportunity to characterize the sea floor with a lightly populated vertical array. A well-mixed and stable water-column during the winter provides robust geoacoustic inversions up to 20 km in range. Under summer conditions, robust inversions are only possible up to 5 km in range, due to the near-surface mixed-layer and highly-dynamic water-column effects of internal waves. Matched-field inversion results from both RAGS03 and Base07 experiments are presented using noise data from different ships at different ranges. Design criteria of a distributed system using single-element receiver nodes for geoacoustic inversion and oceanographic monitoring are also discussed. [Work supported by ONR.]

3:15

2pUWa8. Acoustical observation of randomly fluctuating interior of shallow oceans. Tokuo Yamamoto (RSMAS, Univ. of Miami, 4600 Virginia Key, Miami, FL 33149) and Haruhiko Yamaoka (JCWPL, Tokyo, Japan)

Acoustic reciprocal transmission experiments are conducted at two locations (90 m and 120 m in depth) on the outer continental shelf near New York Byte. The transmission ranges between three transceivers are 3, 4, and 5 km. The moderate-frequency acoustic sources (5.5 kHz) are used with the source level of 192 dB re 1 mcPa at 1 m. Environmental data of fine scale currents and temperature are measured by CTD and inverted ADCP. Characteristics of hydrodynamic fluctuations are determined by transmission data and simulation by a PE code. The sound transmission data and inversion results are severely randomized by the spatial and temporal fluctuations due to fluid dynamic processes, such as turbulent flows, internal waves, and internal solitons. The vertical and temporal coherence analyses are conducted as the techniques for the waveform analyses. These techniques separate the received signals into the coherent component corresponding to the mean state of current velocity and sound speed, and the incoherent component due to fluid dynamic fluctuations. [Work supported by ONR.]

3:30

2pUWa9. Distributed-system detection in correlated clutter. Douglas A. Abraham (CausaSci LLC, PO Box 5892, Arlington, VA 22205)

Distributed sonar systems decide if a target is present through use of a rule that combines information from each individual sensor. The most basic of these are the ''and'' and ''or'' data fusion rules. Analysis of these rules has typically been restricted to cases where the target is detected on some number of sensors, and the false alarms are assumed to arise from processes that are statistically independent from sensor to sensor. In active sonar systems operating in clutter-dominated areas, this assumption may be far from the truth. Active sonar clutter often has its origins in scattering from physical objects (e.g., shipwrecks, fish schools, rock outcroppings, or mud volcanoes). Particularly when the scatterers are small in terms of wavelengths, this leads to false alarms that can be highly correlated from sensor to sensor. In this work, clutter scattering is assumed to arise from a finite number of scatterers with an exponentially-distributed size, which would lead to a K-distributed matched-filter envelope. The performance of the ''and'' and ''or'' fusion rules is then evaluated for two sensors with partially overlapping resolution cells. The resulting detection statistics from the two sensors are not independent, necessitating approximation of the joint distribution function.

Session 2pUWb

Underwater Acoustics: Active Sonar

Shawn Johnson, Chair
Pennsylvania State Univ., Applied Research Lab., 117 Applied Science Bldg., State College, PA 16801

Chair's Introduction—4:00

Contributed Papers

4:05

2pUWb1. Effect of Resolution on Synthetic Aperture Sonar Image Statistics. Shawn F. Johnson, Anthony P. Lyons (Appl. Res. Lab.-Penn State Univ., 117 Appl. Sci.e Bldg., State College, PA 16804), and Douglas A. Abraham (CausaSci LLC, Arlington, VA 22205)

Synthetic Aperture Sonar (SAS) images often exhibit large inter-pixel intensity fluctuations owing to the small resolution scales of such systems, with the reverberation envelope probability density potentially being significantly heavier-tailed than the Rayleigh distribution traditionally assumed for lower-resolution systems. The K-distribution has been shown to well describe sonar reverberation over a wide range of resolutions, with the shape parameter being proportional to the resolution cell size or inversely proportional to the bandwidth. The resolution cell size can be increased by restricting the bandwidth of the data in post processing, which in turn reduces the tails of the reverberation envelope for a higher matched K-distribution shape parameter. Broadband acoustic data suitable for SAS processing collected as part of the SAX04 experiment has been analyzed, and results concerning image resolution and statistics of such data will be presented.

4:20

2pUWb2. An experiment to investigate aural classification of coherent-source active sonar data. Victor W. Young, Paul C. Hines, and Jeff Scrutton (Defence R&D Canada-Atlantic, 9 Grove St., Box 1012, Dartmouth, NS, Canada, B2Y 3Z7, victor.young@drdc-rddc.gc.ca)

Active sonar performance is degraded by the presence of environmental clutter objects, which lead to false alarm echoes. Earlier research suggests that perceptual signal features similar to those employed in the human auditory system can be used to automatically discriminate between impulsive-source target and clutter echoes, thereby improving sonar performance by reducing the number of false alarms [V.W. Young et al., Can. Acoust. 34(3), 50–51 (2006)]. The present work attempts to extend these findings to the realm of coherent-source active sonar echoes. To this end, an experiment was conducted on the Malta Plateau using a cardioid towed-array and broadband FM sweeps running from 500 Hz up to 3.5 kHz. The data set consists of hundreds of pulse-compressed echoes from several targets, shipwrecks, and clutter objects. It includes both monostatic and bistatic echoes, as well as echoes recorded at different aspect angles and at different propagation ranges between roughly 5 km and 30 km. In this paper the experiment is reviewed and preliminary results are presented for an automatic classifier that employs features based on a simple auditory model.

4:35

2pUWb3. Statistical analysis and modeling of active sonar echoes from a shipwreck. Douglas A. Abraham (CausaSci LLC, P.O. Box 5892, Arlington, VA 22205)

A statistical analysis of active sonar echoes from a shipwreck at 36° 18.804′ north by 14° 41.136′ east (Malta Plateau region in the Mediterranean Sea) is performed as a function of azimuth at both 2 km and 10 km range for broadband transmit waveforms in low- and mid-frequency bands. The data were taken in May 2007 during the NATO Undersea Research Centre (NURC) Clutter 2007 sea trial. The data are compared with standard probability density function models and seen to be extremely heavy-tailed with respect to Rayleigh and even K envelope distributions. Models containing a coherent scattering component (e.g., Rician and generalized K) are seen to provide a better fit to the data. Statistical measures from the two different ranges and two different frequency bands are seen to be correlated as a function of azimuth indicating minimal changes in the statistics of the scattering. [This work was sponsored by the Office of Naval Research under contract number N0001407C0092 and made possible by the CLUTTER Joint Research Project (JRP), including as Participants the NURC, Pennsylvania State University-ARL-PSU (USA), Defence Research and Development Canada-DRDC-A (CAN), and the Naval Research Laboratory-NRL (USA).]

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

P. Battenberg, Chair S1
Quest Technologies, Inc., 1060 Corporate Center Dr., Oconomowoc, WI 53066-4828
R. J. Peppin, Vice Chair S1
Scantek, Inc., 7060 #L Oakland Mills Rd., Columbia, MD 21046

Accredited Standards Committee S1 on Acoustics. Working Group Chairs will report on the status of standards currently under development in the areas of physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound, etc. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note—those meetings will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Wednesday, 28 November 2007.

Scope of S1: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics, including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S3 Bioacoustics

C. A. Champlin, Chair S3
Univ. of Texas, Dept. of Communication Sciences & Disorders, CMA 2-200, Austin, TX 78712
D. A. Preves, Vice Chair S3
Starkey Laboratories, Inc., 6600 Washington Ave., S., Eden Prairie, MN 55344

Accredited Standards Committee S3 on Bioacoustics. Working Group Chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note—those meetings will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Wednesday, 28 November 2007.

Scope of S3: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S2 Mechanical Vibration and Shock

R. L. Eshleman, Chair S2
Vibration Institute, 6262 Kingery Hwy., Ste. 212, Willowbrook, IL 60527
A. T. Herfat, Vice Chair S2
Emerson Climate Technologies, Inc., 1675 West Campbell Rd., P. O. Box 669, Sidney, OH 45365-0669

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working Group Chairs will report on the status of various shock and vibration standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 108, Mechanical vibration and shock, and its five subcommittees, take note—that meeting will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Wednesday, 28 November 2007.

Scope of S2: Standards, specifications, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

Session 3aAB

Animal Bioacoustics: Passive and Active Marine Bioacoustics

Juliette W. Ioup, Chair
Univ. of New Orleans, Dept. of Physics, New Orleans, LA 70148

Chair's Introduction—8:00

Contributed Papers

8:05

3aAB1. A simple ocean noise exposure metric based on naturally occurring noise levels and biological thresholds. Michael Stocker, Tom Reuterdahl, Libbie Horn, and Gail Hurley (Ocean Conservation Res., P.O. Box 559, Lagunitas, CA 94938)

Anthropogenic noise is compromising the habitat for marine mammals, fish, and, potentially, other marine organisms. Determining acceptable thresholds is confounded by the fact that marine animals have adapted to some exceedingly loud naturally occurring sounds, whereas exposure to certain anthropogenic noises at equivalent or lower amplitudes causes harm. It is clear that mitigation levels cannot be established by signal amplitude alone. This proposed metric helps establish exposure levels based on broadband and temporal representation of a subject noise compared to a set of spectral curves based on ambient noise levels and biological thresholds.

8:20

3aAB2. Ocean acoustic effects of explosions on land: Evaluation of Cook Inlet beluga whale habitability. Sara K. Tremblay (Animal Sci. Dept., Univ. of Connecticut, Storrs, CT 06269), Thomas S. Anderson (U.S. Army Corps of Engineers, ERDC, Hanover, NH 03755), Erin C. Pettit (U.S. Army Corps of Engineers, CRREL, Fairbanks, AK 99703), Peter M. Scheifele (Univ. of Cincinnati, OH 45221), Gopu R. Potty, and James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882)

The Eagle River Flats is an impact region for artillery at Fort Richardson, Alaska. Adjacent to the Flats is the Knik Arm of Cook Inlet which is the habitat for a distinct population of beluga whales (Delphinapterus leucas). In order to assess the effects of 155 mm artillery explosions on the habitat of these whales, a series of 6.8 kg C4 plastique charges were detonated on land 500 meters from the waters edge. In addition to land seismic and acoustic arrays, hydrophones were deployed in the Knik Arm at high and low tide. This paper discusses the ocean acoustic measurements. The received signal 30 meters from the shore in water depths of 8 meters was more intense at high tide, with broadband peak levels of approximately 180 dB *re* 1 microPa. The dominant frequency was about 20 Hz and most of the received acoustic energy was below 500 Hz. The geology and oceanography of the area were used to model the acoustic time series. Modeled and measured time series are compared to validate the geophysical model and provide estimates of peak pressure and energy flux density over the near shore habitat. [Work supported by U.S. Army Corps of Engineers, CRREL.]

8:35

3aAB3. Acoustic environments of bottlenose dolphins (*Tursiops truncatus*) in the Big Bend region of Florida. Athena Rycyk and Douglas Nowacek (Florida State Univ. Dept. of Oceanogr. 105 N. Woodward Ave., P.O. 3064320 Tallahassee, FL 32306-4320, Rycyk@ocean.fsu.edu)

The researchers examined the acoustic environments of coastal bottlenose dolphins (Tursiops truncatus) in the Big Bend region of Florida, one of the most pristine coastal environments in the state, by means of remote autonomous acoustic recorders deployed at eight sites. The rates and types of dolphin vocalizations (whistles, echolocation, burst-pulse calls, and pops) differed between the eight sites. Dolphins were not the only source of noise in their marine habitat; other biological and anthropogenic sounds occurred. Each site was found to have a unique soundscape with respect to dolphin, fish, snapping shrimp, and anthropogenic sound. Toadfish (Opsanus beta), silver perch (Bairdiella chrysoura), and sea catfish (Arius felis) were the only identifiable fish species to produce sound and each caused notable increases in sound levels at low frequencies where they vocalized. Locations exhibited different time peaks in fish produced sounds. Overall, anthropogenic noise was uncommon and only found in the form of boat noise. Surprisingly, dolphin vocalizations were not found at all locations only at Alligator Harbor, Dog Island, Carabelle River, and West Pass. The lack of dolphin vocalizations does not imply dolphins were not present in these regions only that no dolphins were vocalizing at the time of recordings.

8:50

3aAB4. Preliminary acoustic data analysis from Littoral Acoustic Demonstration Center 2007 (LADC07) passive acoustic experiment. Natalia A. Sidorovskaia (Dept. of Phys., UL Lafayette, UL BOX 44210, Lafayette, LA 70504-4210), George E. Ioup, Juliette W. Ioup, Arslan M. Tashmukhambetov (Univ. of New Orleans, New Orleans, LA 70148), Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS 39406), Joal J. Newcomb, Robert Field, and Guy Norton (Naval Res. Lab., Stennis Space Ctr., MS 39529-5004)

The Littoral Acoustic Demonstration Center conducted a passive acoustic experiment in the Northern Gulf of Mexico, about 120 nmi south of the Mississippi delta between July 5 and July 17, 2007. The primary objective of the experiment was to make probably the first passive acoustic recordings of beaked whales in the Gulf of Mexico. Six environmental acoustic recording system (EARS) buoys were deployed in two triangular configurations in about 1,500 m water with the hydrophones located at a 1,000 m depth. The distance between the two triangles was about 20 nmi. Each EARS buoy recorded in the frequency band of 100 Hz–96 kHz for ten consecutive days, with a total data amount collected of about 3 Tbytes. More than ten different marine mammal species were identified in the vicinity of the buoys during visual observations, which supported the experiment including three different types of beaked whales. Preliminary environmental and acoustic data analysis, including average noise curves and different species identification by using newly developed frequency-domain detectors, is presented. Assessment of the density of recorded acoustic phonations of different species is discussed. [Research supported through SPAWAR PMW-180 PE63207N.]

9:05

3aAB5. Sperm whale tracking in the Bahamas: One hydrophone, three dimensions. Christopher O. Tiemann (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, tiemann@arlut.utexas.edu)

A passive acoustic localization method for tracking the movement of a clicking sperm whale in three dimensions using data from just one hydrophone is demonstrated, using data made available for the 3rd International

Workshop on Detection and Classification of Marine Mammals. One recording contains sperm whale clicks recorded on a bottom-mounted hydrophone on a steep slope of the Navy's AUTEC test range. When the direct-path acoustic ray arrivals from several clicks are time-aligned, persistent associated multipath arrivals of reflected ray paths can be identified for each click event and used for localization. Although the use of multipath arrival information is a standard procedure for range-depth tracking, a three-dimensional estimate of whale position can be obtained from the same multipath information with knowledge of an azimuthally-dependent environment, relative to the receiver. In this case, azimuthal distinction arises from varied bathymetry. Multipath arrival patterns are matched to unique range-, depth-, and azimuth-dependent modeled arrival patterns to make an estimate of whale location. A three-dimensional whale track in range, depth, and bearing from the fixed hydrophone will be presented.

9:20

3aAB6. Tracking baleen whales using the relative relocation method. Catherine L. Berchok, Gerald L. D'Spain, and John A. Hildebrand (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 93940-0701)

Passive acoustic tracking of baleen whales is typically accomplished using time-of-arrival difference measurements from widely-separated acoustic sensor arrays. Because these arrays are large, the propagation conditions along the paths from the source to each of the receivers can be quite different, causing the received signal characteristics to vary among arrivals from the same vocalization. These signal variations introduce uncertainty in the time-of-arrival difference measurements, which in turn creates scatter in the localizations, obscuring the tracks of the individual whales. The relative relocation method, developed in the seismic community, provides a means of reducing scatter in the localizations by using the differences in time-of-arrival differences to adjust the locations of signals relative to one another. The method, as used for earthquake localizations, will be described, then applied to a set of fin whale vocalizations recorded on an ocean bottom seismometer array deployed off southern California. The results provide better defined tracks with less scatter than traditional localization techniques.

9:35

3aAB7. Whale tracking underwater: High frequency acoustic pingers and the instrumented tag (DTAG). Val E. Schmidt, Thomas C. Weber, Colin Ware, Roland Arsenault (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824), David Wiley (Natl. Oceanograph. and Atmospheric Assoc., Situate, MA), Mark P. Johnson, Erik Dawe (Woods Hole Oceanograph. Inst., Woods Hole, MA), and Ari S. Friedlaender (Duke Univ., Beaufort, NC)

Since 2004, scientists have been tagging and tracking humpback whales in Stellwagen Bank National Marine Sanctuary to better understand their behavior. Stellwagen Bank is a shoal area east of Boston and north of Cape Cod, MA where many species of baleen whale feed during the summer months. Instrumented tags (DTAGs) are suction-cupped to the whales back from a RHIB. DTAGS, developed at WHOI, record whale pitch, roll, and heading, 3-D acceleration, depth, and sound for up to 20 h. A pseudotrack for the tagged whale can be generated using visual fixes at the surface and dead-reckoning while the whale is underwater. During extended dives, the solution is expected to exhibit substantial drift, placing limits on the ability to understand feeding behavior, mother-calf interactions, etc. In order to develop higher accuracy whale tracks, three GPS-positioned high-frequency (25–32 kHz) acoustic pingers were deployed around tagged animals in July 2007. The pingers produce time-encoded pulses from known positions, which are recorded along with whale vocalizations and ambient noise on the whale tag. Pulse arrival times from each pinger are converted into ranges from the known pinger locations to generate an underwater whale track. Results from this work will be presented.

9:50

3aAB8. Acoustic identification of beaked and sperm whales. George E. Ioup, Juliette W. Ioup, Lisa A. Pflug, Arslan M. Tashmukhambetov (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148), and Natalia A. Sidorovskaia (Univ. of Lousiana at Lafayette, Lafayette, LA 70504)

Identification of individual marine mammals acoustically was initially motivated in this research by spectrograms of the littoral acoustic demonstration center Northern Gulf of Mexico data containing sperm whale click codas, which showed that clicks in a coda have a spectral pattern that persists across all the clicks in that coda. The hypothesis is that each coda originates from a single whale, and all codas with similar properties come from the same whale. Self-organizing maps (SOMs) are used to compare and classify the time series, Fourier transforms, and wavelet transforms of each coda in order to determine how many whales are present. The results show that SOMs have promise for classifying underwater acoustic coda signals from sperm whales. A similarity measure has been applied to both coda and echolocation clicks and has shown some success in associating both types of clicks with individuals. Other forms of cluster analysis are also considered. Progress in the classification of sperm whale clicks has motivated the application of similar analysis to beaked whale echolocation clicks. Time-frequency plots show interesting details. Preliminary results for beaked whale click analysis are presented. [Research supported by ONR.]

10:05–10:35 Break

10:35

3aAB9. Spawning behavior and spatial distribution of Atlantic herring on Georges Bank revealed by ocean acoustics waveguide remote sensing. Zheng Gong, Daniel Cocuzzo, Mark Andrews, Purnima Ratilal (Northeastern Univ., 360 Huntington Ave., Boston, MA 02115), Srinivasan Jagannathan, Nicholas Makris (Massachussetts Inst. of Technol., Cambridge, MA 02139), Hector Pena, Ruben Patel (Inst. of Marine Res., Bergen, Norway), and J. Michael Jech (NOAA/Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543)

An ocean acoustics waveguide remote sensing (OAWRS) system was deployed in the Gulf of Maine, near Georges Bank to image Atlantic herring and other fish population from September–October 2006. OAWRS provides spatially unaliased imaging of herring over wide areas, spannning over 100 km diam. Migration and spawning behavior of Atlantic herring were observed using OAWRS over several diurnal periods, including massive movements on and off the bank to spawn. Measurements made simultaneously with a conventional fish-finding echosounder (CFFS) and a multibeam sonar provide the depth distribution and local 3-D morphology, respectively, of the herring schools in the water column. Concurrent trawl surveys provide identification of the fish species. Measurements made by OAWRS and CFFS systems are highly correlated. Examples will be provided of the co-registration between the two systems over a one-week period. Calibration of the OAWRS system using CFFS estimates of fish population densities, along with a full-field scattering model that takes into account both coherent and incoherent scattering from a fish group, is discussed. Resonance scattering behavior of herring is observed in the OAWRS system with significant changes in scattering amplitude over the 300 Hz to 1.5 kHz frequency range of the OAWRS system.

10:50

3aAB10. Multibeam and single-beam sonar observations of Atlantic herring in the Gulf of Maine. Thomas C. Weber (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824), Hector Pena (Inst. of Marine Res., Bergen, Norway), and J. Michael Jech (Northeast Fisheries Sci. Ctr., Woods Hole, MA)

In September 2006, several schools of Atlantic herring were simultaneously imaged with a Reson 7125 SeaBat multibeam sonar (400 kHz) and a Simrad EK60 Scientific echo sounder (38 kHz, 120 kHz, and 200 kHz) on Georges Bank in the Gulf of Maine. One school was imaged five

separate times over the span of one hour. These two sonar systems, along with pelagic trawl catch data collected on a separate ship, provided a synoptic view of the fish school as it changed depth and size, fragmented, and eventually dispersed. Of particular interest is the combination of data from the two sonars. The multifrequency, split-beam EK60 provided estimates of the fish school density directly under the ship, but was not capable of accurately extrapolating these measurements to the entire fish school, which was largely out of the EK60 field of view. The multibeam sonar was generally able to image the entire fish school and provided useful information on school volume, fragmentation, and general school morphology, but was more difficult to use for quantifiable measurements, due to the unknown angle of ensonfication of the fish in the outer beams.

11:05

3aAB11. Measurement of material properties of two gelatinous, coastal zooplankton. Joy N. Smith (Dept. of Marine Sci., Coastal Carolina Univ., Conway, SC 29528) and Joseph D. Warren (Stony Brook Univ., Southampton, NY 11968, joe.warren@stonybrook.edu)

Acoustic scattering models for zooplankton are often used to transform acoustic data (backscattered energy) into biological units (number or size of animals). In order for these models to be used successfully, accurate descriptions of the animals themselves are required including: Species, size, orientation, and their material properties, specifically the density and sound speed of the organism relative to their surroundings. Measurements were made of the density and sound speed of two coastal, gelatinous zooplankton (ctenophore *(Mnemiopsis leidyi)*, lion's mane jellyfish *(Cyanea capillata)*) found in the waters around Long Island, New York. Two different methods were used to determine the density of these zooplankton, while an acoustic-travel-time approach was used to measure the sound speed of these animals. Mean animal density was related to animal size for both species and varied between individuals. Values of density and sound speed are reported and their implications with regard to the use of acoustic surveys to assess gelatinous zooplankton are discussed. [Work supported by NSF.]

11:20

3aAB12. Effects of otolith geometry on steady streaming flows in the fish ear. Charlotte W. Kotas, Peter H. Rogers, and Minami Yoda (Georgia Inst. of Technol., Atlanta, GA 30332-0405, charlotte.kotas@gatech.edu)

The dense, bony otolith contained in the fish ear oscillates relative to its fluidic surroundings in the presence of a sound wave. How the otolith actually transduces this acoustically induced fluid motion into the hair cell displacements that the fish hears is not fully understood, however. Nevertheless, it is likely that the complicated geometry of the otolith, including the groove (sulcus) where most of the hair cells are found, contributes in some fashion to shaping the flow patterns that excite the hair cells. The effect of these grooves on the induced steady streaming flows was studied experimentally using both oscillating grooved spheroids and scaled models of actual otolith sulci to simulate acoustically induced motions of the otolith for oscillation orientation angles ranging from 0 deg to 90 deg. Particle-image velocimetry and flow visualization results obtained from images phase-locked to the oscillations are presented for normalized oscillation amplitudes, $\epsilon \equiv s/L = 0.05-0.2$ and Reynolds number, $\mathrm{Re} \equiv \omega L^2/\nu \approx 1-10^2$, where s is the oscillation amplitude, ω is the oscillation frequency, L is a typical otolith length scale, and ν is the fluid kinematic viscosity. The effect of phase on these data was also studied. [Work supported by ONR.]

THURSDAY MORNING, 29 NOVEMBER 2007 GRAND COUTEAU, 8:00 TO 10:20 A.M.

Session 3aAO

Acoustical Oceanography: Acoustic Effects of Internal Waves and Other Finescale Oceanography

James F. Lynch, Chair

Woods Hole Oceanographic Inst., Bigelow Bldg., Woods Hole, MA 02543

Chair's Introduction—8:00

Contributed Papers

8:05

3aAO1. A feed-back algorithm using ocean acoustic modeling to improve shallow water soliton simulations. Stanley A. Chin-Bing, Alex C. Warn-Varnas, and David B. King (Naval Res. Lab., Stennis Space Ctr., MS 39529, chinbing@nrlssc.navy.mil)

Ocean solitons can affect the acoustic signals that travel through them. The effects can range from slight to severe. Computer simulations can be used to predict the large-scale effects on the acoustic signal. The typical sequence of events requires that an ocean model be initialized by tidal velocity and used to estimate the changes in the environmental parameters due to soliton creation and propagation. Changes in the environment are used to calculate the related changes in the ocean sound speed field. The last step is to run an ocean acoustic computer model to predict the changes in the acoustic signal. Often, the tidal velocity is not precisely known and assumptions have to be made. Any variations in the tidal velocity require the time consuming sequence to be repeated. Recently, we have demonstrated a way of estimating the soliton structure that could affect the acoustic signal. This estimation is made before any ocean model simulation is made. This can greatly reduce the number of computer simulations, since only ocean model simulations are made for those conditions that might significantly affect the acoustic signal. Examples will be shown that illustrate this acoustic ''feedback'' method. [Work supported by the NRL Base program.]

8:20

3aAO2. Acoustic cost functions for autonomous adaptive sampling. Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039)

Recent advances in ocean modeling have provided the opportunity to forecast ocean conditions out to 48 h. Accuracy of these models is determined primarily by the input conditions (boundary, as well as internal state). An automated algorithm for adaptively sampling the ocean environment has been developed with the goal of providing sampling paths for regions where input drives the model solutions the most. In this paper, a set of acoustic cost functions is presented, which should indicate regions where acoustic propagation is most sensitive to model uncertainty and/or model-predicted temporal variability. The four cost functions explored will be: Matched mode coherence, mode coupling coefficients, transmission loss variability, and detection range coverage. This technique will be ap-

plied to the data-assimilated model results from the Shallow Water 2006 experiment, where oceanographic measurements consisted of a large array of thermistors, six gliders operating continuously, and a ship towed scanfish. Acoustic measurements included fixed source/receivers, AUV mounted sources, as well as sonobuoy receivers. Sampling guidance based upon the acoustic cost functions will be used in an observational systems simulation experiment (OSSE), in collaboration with Pierre Lermusiaux and his group at MIT. For the ocean modeling, data assimilation and sampling suggestions see http://modelseas.mit.edu.

8:35

3aAO3. Tidal time-scale variations of sound propagation in the Hudson River Estuary. Sreeram Radhakrishnan, Alexander Sutin, and Alan Blumberg (Ctr. for Maritime Systems, Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030)

The nowcast and forecast of sound propagation in an urban estuary is important for the development of acoustic surveillance systems. Sound propagation in an estuary is highly affected by the temporal and spatial variability of salinity and temperature due to tides, freshwater inflows, winds, etc. Estuarine processes are analyzed here focusing on the formation and breakdown of the salinity and temperature stratification and their influence on sound attenuation. Transmission loss (TL) variability was determined first using real-time data collected in the Hudson River near Hoboken, New Jersey. For TL forecasting, 24-h forecasts of salinity and temperature distributions from a high-resolution New York harbor observing and prediction system (NYHOPS) developed at Stevens were used (www.stevens.edu/maritimeforecast/). These forecasts provided the basis for calculating sound speed profiles. A parabolic equation based acoustic model was implemented to simulate the acoustic field structure. Real-time TL measurements were made by transmitting sweep frequency signals (1–100 kHz) to a distance of 175 m. The observed tidal time-scale TL variation of 10 dB was in good agreement with the model calculations. The results indicate that TL can be predicted within the context of the NYHOPS system. [Work supported by ONR Project No. N00014-05-1-0632: Navy Force Protection Technology Assessment Project.]

8:50

3aAO4. Applying data nullspace projection method on matched-field source localization in a random internal wave field. Ying-Tsong Lin, James F. Lynch, and Arthur E. Newhall (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

In a random internal wave field where temporal varying water-column soundspeed profiles are very difficult to measure, matched-field source localization methods can be greatly degraded due to water-column model mismatch. In this paper, the data nullspace projection is applied to this source localization technique to reduce the effect of unknown water-column soundspeed variations. The basis of this method is to project the acoustic signal onto a data nullspace that is insensitive to water-column soundspeed fluctuations. This method only requires the mean and the second-order statistics of temporal varying water-column soundspeed profiles to calculate soundspeed empirical-orthogonal-functions. It does not require measurements of the exact soundspeed field, i.e., each snapshot, for the matched-field processing. In a simulation test case, a linear wave model is used to generate a random internal wave field, an acoustic source continuously transmits a single frequency tone, and the matched-field processing is implemented with the signal received on both a VLA and a HLA, respectively, for localizing the source position. The simulation results show that applying the data nullspace projection method can dramatically improve the robustness and accuracy of the matched-field source localization, resulting in a random internal wave field.

9:05

3aAO5. Coherence function of a sound field in an oceanic waveguide with horizontally isotropic random inhomogeneities. Alexander G. Voronovich and Vladimir E. Ostashev (NOAA/Earth System Res. Lab., 325 Broadway, Boulder, CO 80303)

Closed equations for the coherence function of a monochromatic sound field propagating in a 3-D oceanic waveguide with random inhomogeneities were derived in many works, e.g., see [A. Voronovich and V. Ostashev, J. Acoust. Soc. Am. **119**, 1406–1419 (2006)] and references therein. However, due to high dimensions of matrices appearing in these equations, their numerical solution still remains problematic. In this paper, a closed equation for the coherence function due to a point omnidirectional source in an oceanic waveguide is derived for the case of random inhomogeneities, which are statically isotropic in a horizontal plane. Due to cylindrical symmetry of the problem, dimensions of the matrices appearing in the derived equation are significantly reduced, as compared to a general case. This makes the equation for the coherence function readily amenable for numerical calculations. Using this equation, the effects of internal waves with the Garret-Munk spectrum on the coherence function of a sound field propagating in an oceanic waveguide are studied numerically.

9:20

3aAO6. Effect of ocean internal waves on the interference component of the acoustic field in the Long-range Ocean Acoustic Propagation Experiment. Natalie S. Grigorieva, Gregory M. Fridman (Dept. of Appl. Math., St. Petersburg State Marine Tech. Univ., 3 Lotsmanskaya Str., St. Petersburg, 190008, Russia, nsgrig@natalie.spb.su), James Mercer, Rex Andrew, Bruce Howe, Michael Wolfson (Univ. of Washington, Seattle, WA 98105), and John Colosi (Naval Postgrad. School, Monterey, CA 93943)

The propagation of energy along the sound-channel axis cannot be described in terms of geometrical acoustics because of the presence of cusped caustics repeatedly along the axis. In neighborhoods of these cusped caustics a very complicated interference pattern is observed. Neighborhoods of interference grow with range and at long ranges they overlap. This results in the formation of a complex interference wave-the axial wave-that propagates along the sound-channel axis like a wave belonging to a crescendo of near-axial arrivals. In this paper, the axial wave is simulated for the LOAPEX CTD data measured at seven different ranges from the vertical line array. A signal with the center frequency of 75 Hz and 37.5-Hz bandwidth is used for computations. This signal well approximates one transmitted m-sequence in the LOAPEX experiment. The effect of environmental variability, induced by internal waves, on the axial wave is studied. The sound-speed fluctuations caused by ocean internal waves are obtained with the use of the buoyancy frequency profile measured in the LOAPEX. Calculations are based on 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

3aAO7. Horizontal coherence of low-frequency shallow water signals. Jon M. Collis (Boston Univ., Boston, MA, jcollis@whoi.edu), Timothy F. Duda, James F. Lynch, Arthur E. Newhall, Keith von der Heydt (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Harry DeFerrari, and Hien Ba Nguyen (Univ. of Miami, Miami, FL 33149)

Signals from moored sources at a range of approximately 20 km distant from a fixed L-shaped receiver array (joint horizontal and vertical line array) on the New Jersey area shelf have been analyzed to determine horizontal coherence properties of the signals. The signals, from 100 Hz to 800 Hz, do not arrive broadside to the horizontal array. Subsequently they show phase and amplitude fluctuations attributable to interference within a field having an azimuthally uniform normal mode structure, and azimuthal variations of the modal structure. Synthesis of an interference pattern using the acoustic field measured with the vertical array is used to distinguish between the two effects. High temporal variability of the inferred

azimuthal variation of modal content correlates with internal wave activity observed along the path. Horizontal coherence lengths are often of the order of ten acoustic wavelengths. [Work supported by ONR.]

9:50

3aAO8. Effects of offshore mesoscale eddies and fronts on inshore shallow water acoustic propagation. Harry Deferrari (Div. of Appl. Marine Phys., Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Shallow water shelf areas inside of western boundary currents have two distinctly different ocean acoustic environments determined by the type of front that separates deep offshore from shallow inshore. Prograde fronts allow for the ducting of offshore internal waves up onto the shelf as the main source the sound speed fluctuations over the internal wave band. Retrograde fronts block the propagation of offshore internal waves setting up stability conditions that allow for propagation of locally generated, large amplitude non-linear solitary waves as the major source of sound speed variability. Here, acoustic propagation is examined for both environments with data from two similar fixed system propagation experiments, one for the prograde environment off the coast of south Florida near the site off the Acoustic Observatory, and the second for the retrograde front environment of the Mid-Atlantic Bight the SWO6 site. Offshore mesoscale features of fronts and eddies are shown to determine mean sound speed profiles and the energy of the internal wave fields. In turn, intensity fluctuations and temporal coherencies of broadband acoustic signals over several octaves are observed to vary with variations of the sound speed. For some locations, observations of mesoscale features alone can predict sonar performance.

10:05

3aAO9. Observations of temporal coherence for broad band acoustic transmission in shallow water. Harry Deferrari (Univ. of Miami, 4600 Rickenbacker Cswy. Miami FL, 33149)

The temporal coherencies of broadband signals are computed with data from fixed system measurements at two experimental sites. The acoustic source for both measurements was the Miami Sound Machine (MSM) which transmits continuous signals at 6 carrier frequencies from 100 to 3200 Hz in octave steps. Each signal has a bandwidth of 25% of carrier. The bandwidth of the acoustic signal allows the identification of paths by arrival time, so that coherence can be estimated for individual paths and for multipath groupings. The relation between internal wave energy and acoustic signal coherence times is explored. Internal wave energy is estimated with moored T-D instruments along the path of propagation. The sound speed variability resulting from internal waves at the two sites is markedly different: one is driven by locally generated nonlinear solitary waves and, the other, by offshore internal waves that propagation into the shallow water. The low frequency signals are directly modulated in amplitude and phase at the period of the wave passing the site, whereas, the higher frequencies have saturated statistics without evidence of direct modulation. A puzzling observation is that coherence times do not change much while the IW energy increases by 2 orders of magnitude.

THURSDAY MORNING, 29 NOVEMBER 2007 MAUREPAS, 8:15 TO 11:45 A.M.

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration: Current Topics in Diagnostic and Therapeutic Ultrasound

Jeffrey A. Ketterling, Chair
Riverside Research Inst., 156 William St., New York, NY 10038-2609

Contributed Papers

8:15

3aBB1. Optimized translation of microbubbles driven by acoustic fields. Jean O. Toilliez and Andrew J. Szeri (Dept. of Mech. Eng., Univ. of California, 6112 Etcheverry Hall, Berkeley, CA 94720-1740, toilliez@me.berkeley.edu)

The problem of a single acoustically driven bubble translating unsteadily in a fluid is considered. The investigation of the translation equation allows for identifying the inverse Reynolds number as small perturbation parameter. The objective is to obtain a closed-form, leading order solution for the translation of the bubble, assuming nonlinear radial oscillations and a pressure field as the forcing term. The result is the ability to predict and explicitly understand the rapid and slow transients of bubble displacement, which is proportional to the average acoustic radiation force. The periodic attractor of the Raleigh-Plesset equation serves as basis for an optimal acoustic forcing designed to achieve maximized bubble translation over one dimensionless period. At moderate acoustic intensity, maximizing the radial variance, thereby enhancing bubble collapse, leads to displacement many times larger than the case of purely sinusoidal forcing. The survey covers a wide spectrum of driving ratios and bubble diameters physically relevant to biomedical applications. Shape stability issues are considered. Together, these results suggest new ways to predict some of the direct and indirect effects of the acoustic radiation force in biomedical applications: e.g., targeted drug delivery, selective bubble driving and accumulation. [Work supported by NASA Microgravity Fluid Physics Program.]

8:30

3aBB2. Growth and dissolution of a contrast microbubble: Effects of encapsulation. Kausik Sarkar and Pankaj Jain (Mech. Eng., Univ. of Delaware, Newark, DE 19716)

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 (4–10nm) of protein, lipids and other surface active materials, to prevent their premature dissolution in the blood. A model will be presented that describes the dissolution of a microbubble accounting 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.

8:45

3aBB3. Echogenic liposomes loaded with recombinant tissue-type plasminogen activator (rt-PA) for image-guided, ultrasound-triggered drug release. Denise A. B. Smith, Sampada Vaidya, Jonathan A. Kopechek, Kathryn E. Hitchcock (Dept. of Biomed. Eng, Univ. of Cincinnati, Medical Sci. Bldg, Rm. 6152, 231 Albert Sabin Way, Cincinnati, OH 45267-0586, smitdn@email.uc.edu), Shaoling L. Huang, David D. McPherson (Univ. of Texas Health Sci. Ctr., Houston, TX 77030), and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH 45267-0586)

A recently developed ultrasound contrast agent, rt-PA-loaded echogenic liposomes (TELIP), was assessed in vitro using a clinical diagnostic ultrasound scanner (Philips HDI 5000) equipped with a linear array (L12-5). The stability and echogenicity of static TELIP suspensions were determined using 4.5-MHz harmonic B-mode pulses ($P_r = 120$ kPa; MI $= 0.04$) in an anechoic chamber. An in vitro flow phantom with a flow rate of 5 ml/min at 37 °C was also used to assess TELIP for ultrasonically-triggered drug release. TELIP samples were exposed to: (1) Fundamental 6.9-MHz B-mode pulses ($P_r = 600$ kPa; MI $= 0.04$) where diffusion of gas out of the liposomes occurs over 60 min, or (2) 6.0-MHz color Doppler pulses (PD $= 3.33$ μs, PRF $= 1$ kHz) at two exposure levels, 0.8 MPa (MI $= 0.22$) for which acoustically driven diffusion was evident or 2.6 MPa (MI $= 0.7$), for which rapid fragmentation was confirmed. Exposure of TELIP to Triton-X, a nonionic detergent, served as a positive control for drug release. Release of rt-PA for each ultrasound exposure protocol was assayed spectrophotometrically (Shimadzu UV-1700). The thrombolytic drug remained associated with the lipid bilayer when exposed to B-mode pulses over time and was released when exposed to color Doppler pulses. [This work was supported by NIH 1RO1 NS047603 and NIH 1R01 HL074002.]

9:00

3aBB4. Ultrasound-mediated release of calcein from echogenic liposomes. Jonathan A. Kopechek, Stephen M. Chrzanowski, Denise A. B. Smith, Whitney B. Gaskins (Dept. of Biomed. Eng., Univ. of Cincinnati, Medical Sci. Bldg., Rm. 6152, 231 Albert Sabin Way, Cincinnati, OH 45267-0586, kopechja@uc.edu), Todd A. Abruzzo (Univ. of Cincinnati, Cincinnati, OH 45267), Shaoling L. Huang, David D. McPherson (Univ. of Texas Health Sci. Ctr., Houston, TX 77030), and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH 45267-0586)

Echogenic liposomes (ELIP) have been developed for drug encapsulation. The gas contents in ELIP present a potential mechanism for ultrasound-triggered release of drug contents. Calcein, a fluorescent dye, was loaded in ELIP as a drug substitute (C-ELIP) and ultrasound-induced release was quantified with fluorescence spectrophotometry. Pulsed 6.0-MHz color Doppler from a clinical diagnostic ultrasound scanner (CL15-7 transducer, Philips HDI 5000, MI of 1.3, 150 Hz pulse repetition frequency) was applied to samples of C-ELIP in an *in vitro* flow phantom (2.2 ml/min). For comparison, Triton X-100 was added to C-ELIP to release calcein. Control samples of C-ELIP were not treated with ultrasound or Triton X-100. The echogenicity of C-ELIP (expressed as mean digital intensity in a 0.5 cm^2 region of interest) decreased by 9.6 ± 2.1 dB after exposure to ultrasound. The observed calcein concentration (μg/ml) was 3.1 ± 0.1 for the untreated sample, 4.5 ± 0.1 after Triton X-100 treatment, and 4.2 ± 0.2 after ultrasound exposure. 65.6% of encapsulated calcein was released with ultrasound. These results demonstrate that ultrasound-mediated release of drugs from ELIP using a clinical diagnostic ultrasound scanner is feasible. [Work supported by an AIUM Education and Research Grant and NIH 1R01 HL074002, and NIH 1RO1 NS047603-01S1.]

9:15

3aBB5. Acoustic targeted drug delivery in neurological tissue. George Lewis, Jr., William Olbricht (Cornell Univ., 108 Olin Hall, Ithaca NY 14850, george@cornellbme.com), and George Lewis, Sr. (Transducer Eng. Inc, P.O. Box 4034, Andover MA 01810, thearrayman@transducerengineering.com)

The success of treating brain cancer such as neuroblastomas and neurofibromatosis has not been very effective, and is in fact the leading cause of cancer-related death in patients younger than age 35. In the last 10 years recent developments in drug delivery methods have allowed doctors to implant/inject time-release drugs into the tumor cavity that allows for continuous release of chemotherapy; however results from these studies have not been as successful as anticipated. It is believed that the non-treated cancerous cells are able to migrate from the original tumor site, and relocate beyond the diffusion range for effective drug treatment. In this study we utilize high frequency focused ultrasound to increase drug perfusion into phantoms that mimic brain tissue as a method to increase the rate of drug permeation into the tissue before vascular clearance and cell migration reduce its effectiveness. Using various acoustic pulse sequences we show a substantial increase in drug perfusion into brain mimicking tissue then can be achieved by diffusion alone. We are therefore able to reduce the time of delivery, the amount of drug delivered and the drugs local impact range. This could significantly increase the success of treatment and reduce systemic effects of chemotherapy.

9:30

3aBB6. Robotic high intensity focused ultrasound (HIFU) control for tumor treatment. Shivkumar Kambhampati and Vesna Zderic (The George Washington Univ., 801 22nd St. NW, 624E, Washington, DC 20052, shivak@gwu.edu)

Introduction: Robotic surgery, with more precise and defined movements compared to manual human surgery, is the wave of the future. The purpose of this project is to move the focus of the HIFU transducer using a 3-axis robotic work cell in order to treat the whole volume of a tumor in a precise manner. Methods: A 3-axis robotic work cell from Arricks Robotics was chosen based on price ($3300 USD), precision, range of movement, and payload capacity (up to 2Kg). A LabVIEW program serves as the entire surgery interface including movement options, treatment algorithms with up to 25 points within the tumor, and a guidance system via ultrasound imaging and the hyperechoic spot of the HIFU focus. Results: The robotic work cell could achieve 0.5mm per 10mm precision and speeds of 10mm/sec in real-time within a cell of 180mm by 180mm by 50mm with a HIFU transducer mounted. Testing with native MD-2xp software has resulted in distance testing with less than 2% error in all three axes (0.7% in X, 1.5% in Y, 1.3% in Z). Conclusion: The final outcome will provide a robust, computer-assisted, cost-effective way to perform robotic HIFU surgery.

9:45

3aBB7. Comparison of pathway in high-intensity focused ultrasound lesion production. Yufeng Zhou, Joo Ha Hwang (Div. of Gastroenterology, School of Medicine, Univ. of Washington, Seattle, WA 98195), Kwang Kim, and Joo Ha Hwang (Univ. of Washington, Seattle, WA 981095)

High intensity focused ultrasound (HIFU) is being evaluated for non-invasive treatment of solid tumors. The temperature at the HIFU focus can reach over 65 °C, denaturing cellular proteins and resulting in coagulative necrosis and lesion formation. One common method for delivering HIFU therapy clinically is using the spot accumulation method that delivers sequential individual treatment spots. Because of thermal diffusion from nearby treatment spots, the size of subsequent lesions will gradually become larger as the HIFU therapy progresses, which may cause insufficient treatment of the initial spots, and over-treatment of later spots unless parameters are changed during treatment. A new pathway for HIFU treatment is proposed and compared with the conventional sequential path. Modeling, in vitro phantom and ex vivo bovine liver experiments demonstrate that the new treatment path produces more uniform lesions than the

conventional treatment path ($p<0.05$). The relationship between lesion area/volume and delivered ultrasound energy and the dose-dependent discrepancies between scanning paths were also studied. In addition, the temperature changes in the ex vivo system were measured using a thermocouple array. Altogether, the new treatment path appears to be advantageous for producing more uniform lesions without modifying HIFU parameters during treatment or significantly increasing the scanning time.

10:00–10:15 Break

10:15

3aBB8. Optimization of lesion formation using high-intensity focused ultrasound at large tissue depths. Joshua Samuels and Vesna Zderic (Dept. of Elec. and Comput. Eng., George Washington Univ., 725 23rd St NW, Washington, DC 20052, joshasam@gwu.edu)

Background: The objective was to determine the parameters to making a sizable, controlled lesion deep within the treated tissue using high-intensity focused ultrasound (HIFU). High HIFU intensities allow for rapid temperature rises in tissue used for tumor ablation. Methods: Various experiments were conducted manipulating amplitude and treatment time to produce lesions at full focal length (5.2 cm) of a 3.35 MHz HIFU transducer. Beef thigh was used, as it closely resembles the nonuniformity of human tissue unlike phantom-gels or turkey breasts. Results/Discussion: Inconsistent lesion formation in our experiments showed that deep lesions in a nonuniform medium are not easily created. At 5.2 cm, in situ HIFU intensities dropped to 170–260 W/cm^2 (vs. 40,000–60,000 W/cm^2 free field). Reflections off fat, fascia, and bubbles formed at the HIFU focus (at higher intensities) often appear to result in prefocal lesion formation (average total depth of 3.8 cm). Lower intensities over longer treatment times (up to 120 s) yielded desirable lesion depth (average depth of 4.5 cm, highest depth of 6.8 cm), showing longer treatments at lower intensity could be the key to precise deep lesions. Lesion volumes ranged from 0.1 to 26.5 cc. Conclusion: The challenges posed by treatment of nonuniform tissues are similar to those to be encountered in clinical applications.

10:30

3aBB9. Enhanced high-intensity focused ultrasound lesion formation in tissue phantoms using embedded fibers. Cecille Labuda and Charles C. Church (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, cpembert@olemiss.edu)

High intensity focused ultrasound (HIFU) is currently being developed for hemorrhage control, since it provides rapid energy deposition in the focal region of the field. Near large vessels, the deposition rate is limited by loss of heat to the blood flow, making hemorrhage control difficult. In this study, nylon, copper, and stainless steel fibers were embedded in an albumen-containing tissue phantom to investigate whether these fibers would enhance the HIFU heating effect. The embedded fibers were sonicated at high power, and control sonications were performed away from the fibers. Visible regions of protein denaturation, or lesions, were produced. Control lesions were ellipsoidal and elongated along the acoustic axis. Lesions at the stainless steel and copper fibers were similar in shape, but smaller than the control lesions. Lesions produced at the nylon fibers were teardrop-shaped, elongated along the fiber, and larger than the controls. The roles of thermal conductivity and ultrasonic absorption in the enhanced heating effect were considered. It was concluded that a material of low thermal conductivity and high ultrasonic absorption, such as a polymer fiber, was most suitable for enhancing the HIFU heating effect. [Work supported by DAMD17-02-2-0014.]

10:45

3aBB10. Cost-effective radiation force balance for calibration of therapeutic ultrasound devices. Faezeh Razjouyan and Vesna Zderic (The George Washington Univ., 6882 Lafayette Park Dr., Annandale, VA 22003, faezeh@gwu.edu)

The objective was to create an inexpensive, portable, and accurate absorptive radiation force balance to measure acoustic powers of up to 100 watts generated by the high intensity focused ultrasound (HIFU) transducer. This paper describes the process of making an effective absorbing target with commercially available ingredients. Four different absorbing targets consisting of nickel powder, silicone elastomer, and microballoons were prepared and tested. Silicone Sylgard (Dow Corning, MI) was used for all samples. However, two different microballoons (acrylic and phenolic) and two different nickel powders (high density and spherical nickel powder) were used. The final results were compared with a commercially available reflection radiation force balance (RRFB). The results for the same 3.5 MHz HIFU transducer (Sonic Concepts, WA) revealed that a combination of spherical nickel powder (Alfa Aesar, MA) with acrylic microballoons (Douglas Sturgess, CA) offered an average efficiency of 89.8%, compared to that of RRFB, which was 80.7%. A combination of high density nickel powder (Inco Inc., Canada) with acrylic and phenolic microballoons were 82.5% and 84.2%, respectively (for the same HIFU transducer), while spherical nickel powder and phenolic microballoons (US Composites, FL) had efficiency of 64.4%, indicating incorrect measurements of HIFU transducer efficiency.

11:00

3aBB11. Characterization of low-profile fresnel lenses for annular high-intensity focused ultrasound radiators. David Woolworth, Jason Raymond, and Joel Mobley (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, jmobley@olemiss.edu)

As part of an effort to develop high intensity focused ultrasound (HIFU) applicators for remote hemostasis, we are investigating the use of low-profile stepwise Fresnel lenses for varying the focal depths of single-element annular HIFU transducers. The Fresnel lenses are interchangeable and represent a cost-effective approach to variable HIFU focusing in comparison to multi-element phased array systems. In this work, we report on the characterization of a Fresnel lens coupled with a 1.2 MHz annular transducer as the inner radius of the annulus is varied. We present data derived from hydrophone-based measurements of the pressure field of the transducer/lens system. In addition, we compare our findings with simulations derived using the angular spectrum technique.

11:15

3aBB12. Detection of brachytherapy seeds at varying angles with a singular-spectrum-analysis algorithm. Sarayu Ramachandran, Jonathan Mamou, and Ernest J. Feleppa (F. L. Lizzi Ctr. for Biomed. Eng., Riverside Res. Inst., 156 William St., New York, NY 10038)

A singular-spectrum-analysis (SSA) algorithm previously was shown to be successful in detecting brachytherapy seeds in B-mode images when seeds were orthogonal to the ultrasound beam. In this study, the algorithm was extended to detect seeds at angles to the beam. The SSA algorithm derives P-values from selected eigenvalue pairs of the autocorrelation matrix of seed echo signals. The P-value indicates the probability of presence of a seed. A seed inserted in a gel pad and another inserted in beef tissue were scanned with a 5-MHz transducer. The seed angle was varied from orthogonal to the beam to 23 deg from orthogonal, using 1-degree increments. Simulations of seed echoes based on empirical data were generated and used to test the algorithm. Success in seed detection was expressed by a score computed from P-values. Scores for simulations decreased from 0.8 to 0.45 as the angle increased from 0 to 12 deg. Scores for experimental data varied from 0.74 at 0 deg to 0.62 at 23 deg; the lowest value was 0.1 at 7 deg. Accordingly, the SSA algorithm was successful in detecting the seed with clinically relevant angles between seed and beam.

11:30

3aBB13. Flexible tools for time reversal acoustics focusing applications. Laurent Fillinger (Stevens Inst. of Technol., Castle Point on Hudson, Hoboken, NJ 07030-5991, laurent.fillinger@artannlabs.com), Viktors Kurtenoks, Sam Rosenblum, Alexander Sutin, and Armen Sarvazyan (Artann Labs., West Trenton, NJ 08618)

Time reversal focusing of acoustic waves has numerous biomedical applications including medical imaging, HIFU treatment, and targeted drug delivery. These various applications pose different requirements. Imaging requires a small spatial focused spot, which can be scanned in 2-D or 3-D; therapy requires high deposited energy and controlled-shape extended focus. Artann Laboratories is developing flexible tools including hardware, algorithms, and software with user-friendly interface as a universal platform for wide-range applications, based on the use of the time reversal focusing of acoustic waves. We have built a ten-channel electronic system with sampling frequency up to 33.3 MHz. It enables simultaneous broadcasting of independent arbitrary signals and recording of the response by a scanning receiver/beacon. A set of functions were developed to enable the control of this setup within Matlab, which offers a high level computing language for numerical computation. With a developed system, a diverse variety of experiments and data analysis can be conveniently conducted within the same platform. Illustrative results on numerous biomedical applications of the time reversal focusing obtained within this framework will be presented.

THURSDAY MORNING, 29 NOVEMBER 2007 NAPOLEON B1, 10:30 A.M. TO 12:00 NOON

Session 3aEA

Engineering Acoustics and Underwater Acoustics: Underwater Transduction

Elizabeth A. McLaughlin, Chair
Naval Undersea Warfare Center, 1176 Howell St., Newport, RI 02841

Contributed Papers

10:30

3aEA1. An investigation of the capabilities of a short hydrophone array towed by an ocean glider. Georges A. Dossot, James H. Miller, Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett Bay Campus, Narragansett, RI 02882, georges@oce.uri.edu), Edmund J. Sullivan (EJS Consultants, Portsmouth, RI), Jason D. Holmes (BBN Technologies, Cambridge, MA), James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA), and Scott Glenn (Rutgers Univ., New Brunswick, NJ)

During the Shallow Water Experiment (SW06) a Webb Slocum glider, deployed by Rutgers University, demonstrated that gliders are promising vehicles for towing short acoustic arrays. The gliders saw-tooth trajectory allows for sampling the water column at varying depths and ranges. Further, the glider provides a low-speed platform, allowing for a flow-noise free towed array, which is ideal for the processing of low level signals. One attractive application for glider-towed arrays is target tracking. By using the passive synthetic aperture effect, coupled with a near-field model for the signal, the coordinates of an acoustic source can be estimated using a Kalman filter, but without the necessity of the maneuver normally required by bearings-only tracking. This is possible since the large aperture traced out by the glider permits wavefront curvature to be exploited for range estimation. Using synthetic narrowband data, it is shown that the range and bearing of a low-level acoustic source can be estimated without changing the gliders course. The algorithm is based on an Unscented Kalman Filter. Also, an approach for the broadband problem is outlined. [Work sponsored by the Office of Naval Research.]

10:45

3aEA2. Impact of structural support on vector sensor acoustic performance. David Deveau (PSC 1012, Box 701, FPO, AA 34058, david.deveau@autec.navy.mil)

Vector sensors hold great promise for the exploitation of undersea acoustics by providing the ability to measure the acoustic intensity. Current sensor developments can measure the acoustic acceleration in three dimensions as well as the pressure, which when combined through a cross-correlation produces the acoustic intensity field surrounding the sensor. Mathematical studies indicate that a gain of 4.6 dB is achievable by beamforming the sensor's three cardioid outputs with the pressure sensor's omnidirectional response. In these studies, the mechanics of physically retaining the sensor in position is not considered, but is necessary to allow the sensor to freely sense particle velocity/acceleration and not shadow the sensor to the surrounding pressure field. A set of PVC frames was constructed to hold an individual sensor, which was then combined to form an array of sensors for use in a long-term shallow water acoustic field intensity study. These frames are modeled to determine the level of forward and backscatter caused by the flooded PVC pipes and the possible impact to the theoretical patterns. The physical structures are then calibrated in a controlled measurement facility to determine if the accelerometers move freely in all three dimensions to generate the cardioid patterns necessary to achieve the gains predicted by intensity models over similar pressure sensor configurations.

11:00

3aEA3. Coverage metric for track planning and location of stationary bottom targets. Steven M. Dennis (Code 7183, Naval Res. Lab., Stennis Space Ctr., MS 39529, steven.dennis@nrlssc.navy.mil)

With the increasing dependence on optimization algorithms in track generation in the presence of variable or uncertain environments, the desire for a more environmentally-sensitive metric of sensor performance has arisen. A measure of performance based upon acoustic coverage area has shown promise as a basis for sensor utilization and optimized track generation applications, and is presented here for the case of a bi-static, active sensor used in the search for stationary bottom targets (e.g., wrecks). Coverage can be defined as the areas throughout which a sensor has a sufficiently high signal-to-noise ratio or alternatively, probability, of making positive observations of objects on the seafloor. For the purposes of determining optimal sensor placement and track generation, the area of interest is divided into a sufficiently sampled grid of calculation points. Computing and compiling acoustic sensor coverage area information for grid points throughout the area of interest into an acoustic coverage map gives immediate visual feedback on locations of optimal performance for a specific sensor in the current ocean environment. [The authors appreciate and acknowledge the funding support from the Naval Research Laboratory Base Program.]

11:15

3aEA4. Enhancing underwater acoustic vector sensor measurement performance by point source analysis. Joseph A. Clark (NSWCCD, Code 7340, 9500 MacArthur Blvd., West Bethesda, MD 20817-5000)

The beam pattern response to an actual point-like source that is measured with a vector sensor has been found to differ noticeably from the computed response to an idealized point source even under very low ambient noise conditions. This difference can be exploited to enhance underwater acoustic vector sensor measurements by a repeated process of subtracting an equivalent point source response from the measured response and analyzing the remainder beam. Details of a method for carrying out this type of analysis will be presented. The method will be illustrated with data from an at-sea experiment conducted at a U. S. Navy radiated noise facility near Ketchikan, Alaska (SEAFAC). Features of the equivalent point sources determined by the method will be shown to correspond to known characteristics of the ambient noise field at the measurement site. An evaluation of the improvement in detection and localization performance which can be achieved by the point source analysis method will be reported.

11:30

3aEA5. Acrylic plate acoustic transmission experiment and theory. David R. Dowling and Natasha A. Chang (Dept. of Mech. Eng., Univ. of Mich., 2350 Hayward St., 2212 G. G. Brown Lab., Ann Arbor, MI 48109)

An approximate model of the transmission of sound waves in water through a plate of PMMA was developed for a point source. The model is based on spherical wave propagation and plane-wave transmission through a solid layer, and it is calibrated by identifying the properties of the plastic. By minimizing the error between the modeled and the experimentally measured acoustic wave, the mechanical properties of PMMA can be estimated for PMMA plates of varying thickness d for the d/fluid range of 0.04 to 2.5, and sound-incident angles of 0 deg to 35 deg. This calibration requires only the compression wave ray-path information and was shown to achieve 90% correlation between experimental and predicted waveforms for synthetic cavitation pulses with a nominal bandwidth from 40 kHz to 200 kHz. At larger angles of incidence as measured based on the compression wave, it was necessary to track the various waves that occur in the solid, i.e., compression, shear, and evanescent. The acoustic pressure waves that they generate have to be appropriately added at the receiver location to then recreate the experimentally-measured acoustic signal. The achieved model-experiment correlation at these greater angles of incidence was 80%. [Work supported by ONR.]

11:45

3aEA6. Development of a flextensional transducer for high resolution seismic imaging. Bertrand Dubus, Gerard Haw, Christian Granger, Pascal Mosbah (IEMN dpt ISEN, UMR CNRS 8520, Lille, France), Arkadiusz Kosecki, Bogdan Piwakowski (IEMN, Villeneuve, d'Ascq, France), and Patrick Meynier (Institut Francais du Petrole, Rueil Malmaison, France)

Piezoelectric transducers are considered as potential sources for geophysical applications such as 4D monitoring or shallow subsurface survey, which require high repeatability and precise control of the emitted signal. However, the use of piezoelectric sources is limited by their low power and narrow bandwidth. Flextensional transducers are compact, high power, low frequency, wide-band projectors used in underwater acoustics. The design, fabrication and test of a flextensional transducer for high resolution seismic imaging are presented in this paper. Circuit modeling is used to discuss the general properties of piezoelectric sources radiating in the ground. Finite element simulations (ATILA code) are conducted to design a wide band 500–1500 Hz flextensional transducer. On-site tests are performed for a typical configuration of shallow subsurface survey. The prototype transducer is compared to a weight drop which constitutes a well-known seismic source. By taking advantage of the source repeatability and using specific signal processing, it is found that the flextensional transducer is equivalent to weight drop ranging from 300 to 1300 J. Bandwidth is increased by 200 to 300%, leading to a significant improvement of image resolution.

THURSDAY MORNING, 29 NOVEMBER 2007 NAPOLEON A3, 8:00 TO 11:30 A.M.

Session 3aNS

Noise: Soundscape Developments: Case Studies and Best Practices

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Fluid Mechanics and Engineering, Secr TA 7, 10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Chair's Introduction—8:00

Invited Papers

8:05

3aNS1. Better soundscapes for all—Report on the Workshop in Standardization for Soundscape Techniques held in Salt Lake City, 5 June 2007. Brigitte Schulte-Fortkamp (TU-Berlin ISTA, Einsteinufer 25, D-10587 Berlin, Germany, brigitte.schulte-fortkamp@tu-berlin.de) and Bennett M. Brooks (Brooks Acoust. Corp., Vernon, CT 06066)

The goal of the Soundscape Workshop in Salt Lake City (153rd ASA meeting) was to explore standardization as a possible future advancement in the evolving field of Soundscape measurement, analysis, and design. The perception of the soundscape can provide comfort, tranquility, and needed information to the person concerned, or may be a source of annoyance. The combination of physical acoustical measurements with scientific evaluation of perceptual responses to environmental sound, known as soundscaping, is an essential method for the assessment and actualization of positive outdoor environments. Engineering and aesthetic soundscape design may only proceed based on a complete characterization of the acoustical environment, including the nature of the sound sources and

the reactions of the perceivers. Soundscaping provides for the measurement, analysis, and design of environmental sound by applying the knowledge of both science and community experts. Much fundamental and practical research has been conducted to establish the bases for the soundscape field. The next step is for researchers and practitioners to standardize the available soundscape techniques to allow for more comparison of test and survey results and wider application in design. This presentation outlines the proposed methods and means developed at the workshop for follow-up and further action.

8:30

3aNS2. Soundscapes without decibels. Leslie D. Blomberg (Noise Pollution Clearinghouse, Box 1137, Montpelier, VT 05601)

This paper will examine the advantages and limitations of describing soundscapes in terms of the effects of on specific populations.

8:55

3aNS3. Identification of distinctive patterns and features in soundscapes. Klaus Genuit and Andre Fiebig (HEAD acoustics GmbH, Eberstrasse 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

The environmental noise of several cities located all over the world was recorded, compared, and analyzed. The analyses were focused on the determination of distinctive features and significant patterns of the respective soundscapes, which led to the unique sound of the urban places and that are important because of their individuality, numerousness, or domination. By means of this study and the applied analyses, several acoustical properties, particularities, and patterns were identified, which could be the starting point for an acoustical-orientated soundscape classification. It was found that the environmental noise differs, for example, in character, spectral content, time structure, degree of variation of certain parameters, background to foreground relationship caused by the location-specific noise sources, and their spatial as well as temporal occurrence. It is important to emphasize that the acoustical analyses, on the one hand, have to capture the global, overall impression of the soundscape and, on the other hand, must recognize and adequately interpret single noise events, which also cause strong reactions and emotions. The presentation will show and discuss the case study results.

9:20

3aNS4. From global and semantic evaluations to physical measurements and modeling. Daniele S. Dubois (LCPE/LAM/IJLRDA, 11 rue de Lourmel, 75015 Paris, France, ddubois@ccr.jussieu.fr)

Soundscape research has now largely pointed out the limits of indicators solely (directly) grounded in pressure level measurements and other simple static acoustic parameters to account for nuisance as well as sound quality, as perceived by human listeners. Advances in cognitive research has allowed to identify semantic categories of soundscapes contrasting pleasant areas versus annoying ones with source identification and evaluation as one major criterion of categorization. The present challenge, both theoretical and pragmatic, is therefore to connect such global semantic (psychological) evaluations to analytical (physical) instrumental measurements. We suggest that physical modeling of source signatures and artificial intelligence models of source identification processes could now provide formal representations of typical environmental categories, that could be the missing link to fill the gap between these two types of sound descriptions and interpretations. We will present a state of art concerning our present day knowledge of the structural properties of categorical structures we identified in psychological and semantic investigations, to specify the conditions under which cognitive modeling could account for the structural complexity of source contribution to soundscapes evaluation.

9:45

3aNS5. "The Grand Canyon" vs. "Soundscape of Nowhere." Dickson J Hingson (SIERRA CLUB—Natl. Parks and Monuments Committee, 275 River Run Rd., #3, Flagstaff, AZ 86001)

Twenty years have elapsed since the National Parks Overflights Act of 1987 targeted the once quiet but aviation-imperiled Grand Canyon National Park soundscape as worthy of an expeditious, potentially difficult ''substantial restoration.'' Within several more months, this protracted acoustic effort reaches its long-standing presidential deadline (April, 2008) for completion, as according to specifications/standards set by the National Park Service. A preliminary assessment (first presented by this author at ASA's June biennial meeting) will be refined/updated in view of subsequent developments. Primary and emerging supplemental noise indicators and long-established (or possibly revised) Park Service restoration standards-based mainly on audibility and ''restored'' acreage—will be reviewed. Pertinent words and dimensions of soundscape assessment from this and other soundscape studies will be reviewed and compared, particularly re concepts of ''wilderness character'' and ''beauty.'' Effectiveness of anticipated management actions in the face of ongoing political/legal controversy will be examined, pitting restoration of the authentic Grand Canyon wilderness soundscape against a relatively unsavory option, ''The Soundscape of Nowhere.'' The Grand Canyon soundscape situation situation will be compared with preservation/restoration needs facing other similarly imperiled, iconic national parks in the West, which are at continuing risk for long-term aviation noise impairment.

10:10–10:25 Break

Contributed Papers

10:25

3aNS6. Development of soundscape assessment and design methods for large institutional projects. Gary W. Siebein, Robert M. Lilkendey, and Hyun Paek (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607)

Soundscape assessment and design processes were used to develop design criteria for noise mitigation strategies for the expansion of an institutional complex in a medium-sized city. A series of long-term acoustical measurements of average A-weighted sound levels in the community as well as more detailed octave band measurements and calibrated aural recordings of specific acoustic events that comprised the ambient sounds were mapped for the neighborhood. Sound walks were conducted at various times of day to understand the dynamics of the acoustical environment and to identify issues. Focus group discussions among stakeholders and team members developed the long-term plans for the community and appropriate architectural and acoustical design criteria for the project. Analysis of auralizations of various design options by stakeholders and team members using sophisticated multi-channel playback systems and full-scale aural demonstrations on site were used as part of the evaluation process.

10:40

3aNS7. Toward a less chaotic sound environment for nurses: A study investigating the relationship between layout design, aural connectivity, and user activities. Sele Okcu and Craig Zimring (College of Architecture, Georgia Inst. of Technol., Atlanta, GA 30332-0155)

An intensive care unit has a particularly demanding aural environment, where nurses must be able to hear and respond to numerous alarms, hear critical orders, and respond to patients and family members. This study explores a network measure called aural connectivity that reflects the overall pattern of where users can hear and respond to all key sounds. Aural connectivity is explored in a 20-bed neuro-ICU in a major teaching hospital by mapping locations where each alarm and key sound from the patient rooms can be heard, using both self-report and objective measures of sound. This study explores how aural connectivity relates to tasks, unit layout and design, and the nature of noise sources. Aural connectivity has the promise of being a useful overall design and evaluation tool because it allows movement and tasks to be mapped onto a network of where effective auditory monitoring can occur.

10:55–11:30
Panel Discussion

THURSDAY MORNING, 29 NOVEMBER 2007 BORGNE, 8:25 TO 10:00 A.M.

Session 3aPAa

Physical Acoustics: Ultrasound Measurements and Methods for Condensed Matter

Veerle M. Keppens, Chair
Univ. of Tennessee, Materials Science and Engineering, Knoxville, TN 37996

Chair's Introduction—8:25

Invited Papers

8:30

3aPAa1. Slow dynamics and the Larsen effect at millisecond time scales. Richard Weaver and Oleg Lobkis (Dept. of Phys., Univ. of Illinois, 1110 W Green St., Urbana, IL 61801)

At sufficient gain, an ultrasonic feedback circuit rings with a ''Larsen'' tone that depends on the nonlinear electronics, and also on the acoustic properties of the solid body to which it is attached. Because the spectrum of this tone is extraordinarily narrow and stable, it may be measured with high precision. With a goal of using this to monitor small rapid changes in materials, here the stability of the signal to perturbations is quantified, and the results applied in the monitoring of the evolution of the effective modulus of mesoscopically elastic bodies (cements and stones) after brief transient loads. In accord with other studies, it is found that the modulus drops after the load, but then recovers in a characteristic manner, like $\log(t)$. The present technique, using as it does frequencies of order MHz and loads with durations of order 10 μs, extends these studies of slow dynamics to early times. It is found that $\log(t)$ behavior can be sustained over the full range investigated, from a couple of μs to hundreds of seconds. [Work supported by NSF CMS 05-28096.]

8:55

3aPAa2. Elastic constant measurements in heterogeneous systems: Thin films on a substrate. Joseph Gladden III (Dept. of Phys. and Astron., Univ. of Mississippi, 108 Lewis Hall, University, MS 38677, jgladden@phy.olemiss.edu)

Resonant ultrasound spectroscopy (RUS) has shown itself to be an efficient and accurate method for determining the full elastic tensor of a single crystal sample. Elastic constants are a measure of the curvature of the interatomic potentials, and are thus sensitive to a variety of phase transitions in materials. In this talk, an extension of RUS will be presented that allows for the determination of the elastic constants of a crystalline thin film deposited on a substrate with thicknesses on the order of 100 nm or greater. Both

experimental and numerical analysis issues for thin film measurements will be addressed. Some ''proof of principle'' results will be presented as well as experimental data showing magnetic phase transitions in colossal magnetoresistance films between 200 nm and 400 nm thick. Thin film RUS should prove to be a useful tool in the effort to better understand phase transitions in reduced dimensionality and effects of lattice mismatch between substrate and film.

9:20

3aPAa3. Making resonant ultrasound spectroscopy measurements from milliKelvin temperatures to well above room temperature. J. B. Betts, A. Migliori, and I. Stroe (NHMFL, Los Alamos Natl. Lab. Los Alamos, NM 87545, jbbetts@lanl.gov)

Measuring elastic moduli using resonant ultrasound spectroscopy (RUS) can be very challenging at the best of times. When the measurement environment needs to be controlled from milliKelvin to many hundreds of Kelvin, this challenge becomes extremely acute. We will present methods and results from such experiments showing that with care, elastic moduli can be very accurately determined over these extreme temperature ranges. [Work supported by the National Nuclear Security Administration, U.S. Dept. of Energy, State of Florida, and the National Science Foundation.]

Contributed Paper

9:45

3aPAa4. Common acoustic properties of solid aggregate at micro- and nano-scales. Hasson M. Tavossi (Dept. of Phys., Astron., and Geosciences, Valdosta State Univ., 1500 N. Patterson St., Valdosta, GA 31698)

Acoustic properties of solid aggregate, when measured at macroscopic scale, have been found experimentally to have remarkable similarities with the same wave properties observed at the atomic and nano-scales. It can be shown that the elastic moduli and other wave properties such as, wave tunneling, attenuation, cutoff-frequency, and dispersion, depend on the same structural factors at macro- and nano-scales. The model constructed for the acoustic properties of solid aggregate, expressed in terms of wave-number (ka), can be applicable in the wide range of frequencies and length scales. The findings on acoustic properties of solid aggregate at macroscopic scale could lead to a better understanding of the wave properties of solids at nanoscales. These readily analyzable models for acoustic properties at macroscopic scales are convenient tools for verification of theoretical models for acoustic behavior of complex solids at atomic level. Experimental data and numerical results are compared, showing acoustic wave responses of the solid aggregate for a wide range of frequencies, from audible to ultrasound. The relevance of these results to nano systems will be discussed to show that parallel acoustic properties can exist in the range of characteristic lengths from macro- to nano-scales.

THURSDAY MORNING, 29 NOVEMBER 2007 BORGNE, 10:15 A.M. TO 12:00 NOON

Session 3aPAb

Physical Acoustics: Thermoacoustics

William V. Slaton, Chair

Univ. of Central Arkansas, Dept. of Physics and Astronomy, 201 Donaghey Ave., Conway, AR 72035-0001

Contributed Papers

10:15

3aPAb1. Shape factor characterization of fibrous media with a temperature gradient. William V. Slaton (Dept. of Phys. & Astron., The Univ. of Central Arkansas, Conway, AR 72035, wvslaton@uca.edu)

Recent theoretical work generalizes thermoacoustic theory to random porous media [H. S. Roh et. al., J. Acoust. Soc. Am. **121**(3), 1413–1422 (2007)]. Characteristics of the porous media, such as the tortuosity and dynamic shape factors for viscous and thermal effects, are introduced into the thermoacoustic wave equation and may be determined by suitable impedance measurements at zero temperature gradient. This theoretical approach may also be used to model fibrous media, such as fiberglass or steel wool. A new technique to determine the scaling factors for fibrous media utilizing a single-step finite difference inversion of the thermoacoustic wave equation [R. Raspet et. al., J. Acoust. Soc. Am. **103**(5), 2395–2402 (1998)] with zero temperature gradient will be presented. Roh et. al. predict that acoustic gain with a nonzero temperature gradient may be written in terms of these scaled cylindrical dissipation functions. The acoustic gain term may also be determined by suitable application of the single-step finite difference technique. [This material is based upon work supported by the U. S. Army Space and Missile Defense Command under Contract No. W9113M-06-C-0029.]

10:30

3aPAb2. Simulating thermoacoustics in random stack materials. Carl Jensen (Dept. of Phys. and Astron. and the Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

A lattice Boltzmann based computational fluid solver is being developed to investigate the thermoacoustic properties of stack materials with irregular pore geometries. The goal of the investigation is to perform a direct simulation of the time-varying thermohydrodynamic flow within a porous material and determine the material's thermoacoustic properties from its base characteristics. Accurately simulating a porous material's pore scale flow requires methods for handling the irregular boundary conditions presented by the material and the use of high performance computing to handle the large number of elements required to represent the material's bulk behavior. Progress in both of these areas will be presented as results in applying the use of commodity graphics programming units for high performance computing and a local grid refinement scheme for increased resolution at the wall boundaries. [Work supported by U.S. Army Space & Missile Defense Command.]

10:45

3aPAb3. Design environment for low-amplitude thermoacoustic energy conversion (DeltaEC). John P. Clark, William C. Ward, and Gregory W. Swift (M.S. K764, Los Alamos Natl. Lab., Los Alamos, NM 87545, swift@lanl.gov)

The Los Alamos thermoacoustics code, available at www.lanl.gov/thermoacoustics/, has undergone extensive revision this year. New calculation features have been added to the original Fortran computational core, and a Python-based graphical user interface wrapped around that core provides improved usability. A plotter routinely displays thermoacoustic wave properties as a function of x or tracks results when a user-specified input variable, such as frequency or amplitude, is varied. The Windows-like user interface provides mouse-based control, scrolling, and simultaneous displays of plots and of several categories of numerical values, in which color indicates important features. Thermoacoustic phenomena can be calculated with superimposed steady flow, and time-averaged pressure gradients are calculated. In thermoacoustic systems with toroidal topology, this allows modeling of steady flow caused by gas diodes (with or without time-averaged heat transfer) and Gedeon streaming. Thermoacoustic mixture separation is included, also with superimposed steady flow. The volume integral of the complex gas momentum is available, so vibrations of thermoacoustic systems can be analyzed.

11:00

3aPAb4. Thermoacoustic quality factor measurement in a Helmholtz resonator. Holly Smith and William Slaton (Dept. of Phys. & Astron., The Univ. of Central Arkansas, Conway, AR 72035)

A Helmholtz resonator consists of a hollow neck attached to an empty chamber. This resonator can be modeled as a spring-mass system in which the air moving inside the neck acts as the mass and the gas inside the chamber acts as the spring. Every Helmholtz resonator has a characteristic quality factor that is dependent upon the total mechanical resistance present. A system with low resistance will have a narrow peak on its amplitude versus frequency graph and a high quality factor, whereas a system with high resistance will have a broader peak and a low quality factor. In this experiment, a porous ceramic substrate is inserted into the neck of a Helmholtz resonator. Introducing this substrate into the neck of the Helmholtz resonator alters the resonance frequency and the quality factor of the resonator. The quality factor will be shown to be increased with increased temperature difference until spontaneous onset of sound generation. The effect of convection will be studied by exploring the dependence of the quality factor on the physical orientation of the resonator with respect to the vertical. [This work was supported by the Arkansas Space Grant Consortium and the University of Central Arkansas University Research Council.]

11:15

3aPAb5. Effect of neck geometry on aeroacoustic excitation of a Helmholtz resonator. Stephanie Lanier and William Slaton (Dept. of Phys. & Astron., The Univ. of Central Arkansas, Conway, AR 72035)

The aeroacoustic response of a Helmholtz resonator, when attached to a wind tunnel via junctions with different neck geometries, has been examined. The wind tunnel consists of a 2-inch ID glass tube with variable mean flow at atmospheric pressure connected to a cross-junction with two coaxial 5-liter Helmholtz resonators. During the experiment, the acoustic pressure and corresponding frequency is measured in one of the Helmholtz resonators at different wind tunnel flow velocities. By knowing these quantities, the acoustic velocity of the air at the neck of the resonator can be determined. The ratio of the acoustic velocity to the flow velocity as a function of the Strouhal number can be studied for differing neck geometries. Helmholtz resonator neck geometries studied include 90-deg bends and straight sections in various combinations. [This work was supported by the Arkansas Space Grant Consortium and the University of Central Arkansas University Research Council.]

11:30

3aPAb6. An analytical model for the design of optimized heat exchangers for thermoacoustic systems. Arganthaël Berson, Philippe Blanc-Benon (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 36 avenue Guy de Collongue, 69134 Ecully Cedex, France), and Vitalyi Gusev (Université du Maine, 72085 Le Mans Cedex 9, France)

A 1-D nonlinear model for the heat transport from the edge of a thermoacoustic stack to a heat exchanger is developed. This model is an extension of the previous study by [Gusev et al., J. of Sound and Vibration **235**, (2000)] to the case of a finite dimension heat exchanger with a temperature difference between the stack and the heat exchanger plates. The model is based on a relaxation-time approximation for transverse heat transfer that allows for the calculation of the heat flux through an adiabatic gap, separating the stack and the heat exchangers. It shows the generation of temperature harmonics close to the edge of the plates that leads to nonlinear heat transport through the gap. The heat flux extracted by the hot side heat exchanger is calculated, taking into account viscous dissipation along the plates and reverse heat conduction toward the stack. An optimal set of geometrical and relaxation parameters for the heat exchanger is obtained. Results are compared with experimental measurements of temperature fluctuations behind the stack plates, and with the results of numerical simulations from the literature. [This work is supported by ANR (project MicroThermoAc $NT05_{1}2101$).]

11:45

3aPAb7. Thermoacoustics and thermal dissociation of water. Gregory W. Swift and Drew A. Geller (M.S. K764, Los Alamos Natl. Lab., Los Alamos, NM 87545, swift@lanl.gov)

Near 2600 K, 10% of water molecules are thermally dissociated at atmospheric pressure, with a reaction time constant below 1 ms. Such temperatures can be reached with focused sunlight. To use this endothermic reaction for the production of hydrogen, the hydrogen must be separated from the oxygen at high temperature, because they would quickly recombine to form water again, if the unseparated mixture were simply returned to lower temperatures. We have considered thermoacoustic mixture separation for this purpose. Our calculations show that the thermal-diffusion ratios are high enough to yield steadily flowing streams of hydrogen-enriched steam and oxygen-enriched steam in a separation channel less than a wavelength long. However, the thermoacoustic power density in 1-bar steam is low, so the required apparatus would be large, needing alot of expensive and fragile high-temperature material, such as calcia-stabilized zirconia. Our estimates show that this approach to solar hydrogen production would be approximately 30 times more expensive than solar-Stirling electricity generation driving traditional water electrolysis. [Work supported by DOE Office of Science.]

Session 3aSCa

Speech Communication: Speech Intelligibility and the Vowel Space

Sarah Hargus Ferguson, Cochair
Univ. of Kansas, Dept. Speech Language Hearing Science and Disorders, Lawrence, KS 66045

Gary G. Weismer, Cochair
Univ. of Wisconsin, Waisman Ctr., Madison, WI 53705-2280

Chair's Introduction—8:00

Invited Papers

8:05

3aSCa1. Vowel space parameters. D. H. Whalen (Natl. Sci. Foundation, 4201 Wilson Blvd., Ste. 995, Arlington, VA 22230, and Haskins Labs.)

Vowels are universal in human languages and have long been categorized in feature systems. However, the feature systems are based on introspection and include a notion of ''height'' or ''closeness,'' that is not anatomically straightforward. From x-ray and other imaging data, the important aspects seem to be palatal, velar, and pharyngeal closures, with secondary effects of lip rounding. Describing vowels in terms of location of constriction makes some assimilation patterns more direct, as with palatal vowels (e.g., /i/) conditioning palatal fricatives. Vowels can also be described as intrinsically slower than consonants, which allows for a more straightforward description of their timing relationships, and a possible explanation for the crosslinguistic preference for CV syllables. The use of fricatives for syllabic nuclei is common in languages and derives easily from a narrowing of the constriction to the point at which frication is generated. More and less vowel-like patterns in consonants have also been noted, such as in coarticulation of laterals and rhotics. Gesture-based systems have difficulty describing some phonological patterns (e.g., vowel lowering). The status of schwa, which seems to lack a constriction, is unclear. With two competing systems having different advantages, work directly comparing the approaches is needed within speech science.

8:35

3aSCa2. Vowels and speech intelligibility. Diane Kewley-Port (Speech Psychophysics Lab., Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Recent research on the contribution of vowel versus consonant segments to sentence intelligibility has demonstrated a 2:1 advantage for vowels in a noise replacement paradigm [Kewley-Port et al., J. Acoust. Soc. Am. (in press)]. The vowel advantage is sustained when boundaries are slightly modified in sentences, but different results have been reported for isolated words. Our follow up research with typical elderly hearing-impaired listeners has demonstrated that the vowel segmental information is especially important in sentence intelligibility, even though it is commonly believed that consonants are more important than vowels for speech recognition. Additional research in our lab has examined the relation between listeners' ability to discriminate acoustic detail in vowel formants at a more peripheral level of the auditory system, and to identify vowel categories at more cognitive/linguistic levels. Three populations of listeners have participated in these experiments: Normal-hearing Americans, hearing-impaired Americans, and normal-hearing second language learners of English. Moderate correlations between formant discrimination and vowel categorization were observed for all three groups. Implications of our results will be discussed in terms of theories of speech perception and of second language speech learning. [Work supported by NIH.]

9:05

3aSCa3. The acoustic vowel space in speech disorders: Data and interpretation. Gary Weismer (Dept. of Communicative Disord. and Waisman Ctr., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706)

The acoustic vowel space, usually defined by the area enclosed by the corner vowels in an F1-F2 plot, may have some use as a general index of speech severity in various types of speech disorders. In this presentation, vowel space data will be reviewed for a number of different speech disorders, and the issue of the role of vowel space measurements in understanding both the underlying speech production deficit, as well as deficits in speech intelligibility, will be considered. Vowel space estimates, it will be argued, are in large part a gross estimate of speech motor control deficit. This is in contrast to a view of vowel space estimates as direct predictors of, say, specific aspects of segmental articulation difficulties or componential aspects of speech intelligibility deficits (for example, the extent to which vowel problems predict a speech intelligibility deficit). Some suggestions for future work, including extensions of the vowel space concept to include planar estimates of formant motion, will be presented. [Work supported by NIH.]

9:35

3aSCa4. Speech intelligibility and the vowel space: Normal speech. Amy T. Neel (Dept. of Speech and Hearing Sci., Univ. of New Mexico, MSC01 1195, Albuquerque, NM 87131-0001)

Vowel formant frequency values and related vowel space measures have been widely used in the study of speech intelligibility for normal and disordered talkers. The strength of vowel space measures in predicting speech intelligibility for normal talkers, however, is relatively low: [Bradlow et al. (1996)] found that vowel space dispersion predicted 19% of variance in sentence intelligibility, and [Hazan and Markham (2004)] found that the F2 difference between /i/ and /u/ predicted 16% of variance in word intelligibility scores. Using data from the 45 men and 48 women speakers in the Hillenbrand et al. (1995) database, acoustic characteristics of ten vowels were used to predict identification accuracy. Multiple regressions revealed that global measures (mean f0, F1 and F2, duration, and amount of formant movement) and fine-grained measures (vowel space area, mean distance among vowels, ranges for f0, F1, and F2, duration ratio between long and short vowels, and dynamic ratio between dynamic and static vowels) accounted for less than one-quarter of the variance in identification scores across talkers. Focusing on confusions among spectrally similar vowels may provide better information on intelligibility differences among normal talkers. Goodness ratings may provide a wider range of scores for statistical analysis than identification accuracy.

10:05–10:30 Break

10:30

3aSCa5. Clear speech effect on vowel production across languages. Rajka Smiljanic and Ann Bradlow (Linguist., Northwesten Univeristy, 2016 Sheridan Rd., Evanston, IL 60208 rajka@northwestern.edu)

In this talk, we discuss vowel production in clear speech, a distinct mode of speech production intended to enhance intelligibility, across languages with different phonological properties (Croatian and Spanish with small vowel inventories vs. English with a large vowel inventory). A comparison of Croatian and English revealed the equivalent clear speech vowel space expansion in both languages. In addition, listeners recognized words in noise more accurately in clear than in plain speech in their native language, establishing that the plain-to-clear speech articulatory modifications increased intelligibility. A comparison of Spanish and English also showed similar amounts of vowel space expansion as well as the maintenance of coarticulatory patterns in clear speech for both languages. The inventory-independent vowel space expansion in all three languages suggests that talkers hyperarticulate even when segments are unlikely to be perceptually confusable (few vowel categories that are fairly distinct). The maintenance of coarticulation further suggests that spreading of segment identity across neighboring sounds may be beneficial to the listener. In this talk, we also discuss various measures of vowel space expansion and different clear speech strategies across talkers. We conclude with some remarks about the link between the identified acoustic-phonetic features of clear speech and intelligibility.

11:00

3aSCa6. Listener influences on vowel intelligibility. Tessa Bent (Dept. of Psychol. and Brain Sci., Indiana Univ., 1101 E. 10th St., Bloomington, IN 47405, tbent@indiana.edu), Sarah Hargus Ferguson (Univ. of Kansas, Lawrence, KS 66045), and Diane Kewley-Port (Indiana Univ., Bloomington, IN 47405)

Do the acoustic-phonetic parameters that promote highly intelligible speech vary across different listener populations? The current study investigated whether inter-talker differences in American English (AE) vowel intelligibility were maintained across three listener groups: Normal hearing native AE-speaking adults (NH), normal hearing Korean-speaking adults learning AE as a second language (L2), and hearing impaired native AE-speaking adults (HI). These groups heard recordings of 10 AE vowels in /bVd/ context produced in conversational speech style by 12 talkers. The stimuli were mixed with noise and presented for identification in a 10-alternative forced choice task. Vowel intelligibility varied substantially among the talkers. There were significant correlations of talker scores for the NH and HI listeners, but not for NH and L2 listeners. An analysis of acoustic-phonetic vowel features indicated that vowel intelligibility was negatively correlated with vowel space expansion for the HI group, while for the L2 group, intelligibility was positively correlated with vowel duration. These data suggest that although both HI and L2 listeners have difficulty in noisy listening situations, their degraded performance has different underlying causes that lead to substantial differences in the perception of the intelligibility of different talkers. [Work supported by NIH-NIHDCD 02229 and T32-DC00012.]

Session 3aSCb

Speech Communication: Speech and Prosodic Issues (Poster Session)

Winifred Strange, Chair
CUNY Graduate School, Program in Speech and Hearing, 365 Fifth Ave., New York, NY 10016

Contributed Papers

All posters will be on display from 9:00 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 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

3aSCb1. How did a stimulus order in AX discrimination training influence improvement in the ability to perceive an American English /l-r/ and /gl-gr/ among Japanese listeners? Teruaki Tsushima (Ryutsu-Kagaku Univ., 3-1, Gakuen-Nishi-Machi, Nishi-Ku, Kobe 651-2188, Japan, Teruaki_Tsushima@red.umds.ac.jp)

Previous research has shown that discriminability of /l-r/ in AX discrimination may be significantly lower when /l/ is presented as the first stimulus than otherwise, among Japanese listeners under certain experimental conditions. The present study examined whether the stimulus order in AX discrimination training had significant effects on improvements of the ability to discriminate and identify English /l-r/ and /gl-gr/. One group of Japanese listeners ($N=12$) received twelve sessions of AX training in which stimuli (i.e., lexical items) were presented in the order of /l/ to /r/ (e.g., glock-rockh), while the stimulus order was reversed (e.g., grock-lockh) in the other ($N=12$). A comparison of pre-test and post-test results found the former group showed a significant improvement of the ability to discriminate /l-r/ ($p=0.037$), to identify /l/ in /l-r/ ($p=0.023$), and a significantly better-defined category boundary of synthetic /la-ra/ stimuli than the other group in the post-test ($p=0.05$), while none of the improvements were significant in the other group. The results indicated, unlike what was predicted from the previous findings, the ability to perceive /l-r/ may be better improved when /l/ was presented first than otherwise during AX training. The results will be discussed in terms of models of L2 speech learning.

3aSCb2. Nonnative phonemes are open to native interpretation: A perceptual learning study. Matthias J. Sjerps and James M. McQueen (MPI for Psycholinguist., Postbus 310, 6500 AH, Nijmegen, The Netherlands, matthias.sjerps@mpi.nl)

Four experiments examined whether Dutch listeners can learn to interpret a nonnative phoneme (English [θ]) as an instance of a native category (Dutch [f] or [s]). During exposure in Experiment 1, two listener groups made lexical decisions to words and nonwords. Listeners heard [θ] replacing [f] in 20 [f]-final words (Group 1), or [s] in 20 [s]-final words (Group 2). At test, participants heard e.g., [doθ], based on the minimal pair doof/doos (deaf/box), and made visual lexical decisions to e.g., doof or doos. Group 1 were faster on doof decisions after [doθ] than after an unrelated prime; Group 2 were faster on doos decisions. The groups had thus learned that [θ] was, respectively, [f] or [s]. This learning was thorough: effects were just as large when the exposure sound was an ambiguous [fs]-mixture (Experiment 2) and when the test primes contained unambiguous [f] or [s] (Experiment 3). In Experiment 4, signal-correlated noise was used as the exposure sound. Listeners learned that the noise was an [f], irrespective of [f]- or [s]-biased exposure, showing that learning is determined by the new sounds spectral characteristics. Perceptual learning in a native language is thorough, and can override years of second-language phonetic learning.

3aSCb3. Effectiveness of a robust computer assisted pronunciation training tool. Kwansun Cho, John G. Harris (Dept. of Elec. and Comput. Eng., Univ. of Florida, Gainesville, FL 32611, kscho@cnel.ufl.edu), and Ratree Wayland (Univ. of Florida, Gainesville, FL 32611)

A reliable ASR-based pronunciation training tool named STAR (self-training accent reduction) is implemented for native speakers of Korean learning American English. STAR is designed to focus on the most frequent phonemic errors made by Korean adult learners and to provide instantaneous feedback. In order to investigate the effectiveness of STAR, ten Korean participants are recruited for this pilot study. The study consists of three phases: Pre-test, training, and post-test. During the pre-test, the participants read a pre-designed word list containing accent sensitive phonemes into a microphone and participate in one session of training using the STAR system. The participants return for two more sessions of training on the following day. The post-test is administered on the third day, after one additional training session. During the post-test, the participants read the same wordlist as the one administered during the pre-test, and an additional wordlist that they were not trained on. Two trained phoneticians listen and transcribe all recordings to examine whether the participants' productions are more accurate after the STAR training. The results indicate that most of the participants who practice pronunciation with STAR show improvement in their pronunciation.

3aSCb4. The adaptability of laboratory phonemic perception training protocols to common second language instruction situations. Thomas R. Sawallis and Michael W. Townley (English Dept., Univ. of Alabama, Box 870244, Tuscaloosa, AL 35487, tsawalli@bama.ua.edu)

For the past two decades, second language pedagogy has de-emphasized phoneme-level pronunciation training. During the same period, laboratory experiments have demonstrated a variety of benefits from training non-natives in the perception of difficult target language contrasts. Experiments have shown that learners' perceptual performance improves (Jamieson & Morosan, 1986; Flege, 1989 & 1995), that the improvements generalize to new talkers and words (Lively, Logan, & Pisoni, 1993), that perceptual training triggers production improvements (i.e., without specific pronunciation training, Bradlow et al., 1997), and that both perceptual improvements (Flege, 1995; Lively et al., 1994) and production improvements (Bradlow et al., 1999) continue over several months. Protocols for these training experiments are generally intensive, and may include sessions of 40 min, three sessions per day, or sessions for 15 consecutive business days. Such onerous protocols are acceptable for basic research with compensated subjects. They are unacceptable for normal L2 instruction, with paying learners and with limited time for pronunciation training. The present study analyzes the training protocols from selected research articles, and documents the types and amounts of training. It then suggests

acceptable alternative regimens for achieving comparable training in typical learning situations, including 16-week college semesters and shorter language institute terms.

3aSCb5. The categorical nature of tones and consonants: Evidence from second language perception and production. Yen-Chen Hao and Kenneth de Jong (Dept. of Linguist., Indiana Univ., 322 Memorial Hall, 1021 E. 3rd Str., Bloomington, IN 47405, yehao@indiana.edu)

Tones are commonly considered to be psychologically equivalent to phonetic segments like consonants and vowels [Ladd, Intonational Phonology (1996)]. However, comparing two current studies reveals intrinsic differences between tones and consonants. One study examines ten American learners acquiring Chinese tones. The other examines 40 Korean learners acquiring English obstruents. Both studies include three tasks: Identification—subjects identified the tones or consonants of L2 nonsense words. This task requires auditory perception and associating the sound with a linguistic category. Mimicry—subjects listened and repeated the stimuli. This task requires perception and production. Reading—subjects read a list of nonsense words. This task requires associating linguistic labels with speech production. These two studies yield different patterns. For tones, accuracy rates for Mimicry are the highest, suggesting learners have more problems with linguistic association, the component shared by Identification and Reading. For consonants, accuracy rates for Mimicry are the lowest, signifying difficulty with perception and production, but not with linguistic association. Taken together, these findings suggest a pervasive difference in categorical nature of tones and consonants. These results also highlight the necessity of multiple tasks in assessing how linguistic contrasts function in a phonetic system. [Work supported by the NSF.]

3aSCb6. Neutralization in the perception and production of English coda obstruents by Korean learners of English. Hanyong Park, Yen-chen Hao, and Kenneth J. de Jong (Dept. of Linguist., Indiana Univ., 332 Memorial Hall, Bloomington, IN 47405, hanypark@indiana.edu)

Phonological neutralization rules require the suspension of differences between segments in perception and production. This paper examines the role of neutralization in the production and perception of Korean learners of English. In Korean, laryngeal and manner contrasts found in initial obstruents are systematically neutralized in coda position into stops that sound like voiceless stops in English. The current paper pursues two questions: (1) Does neutralization have the same effect on perceptual and production abilities? (2) are neutralization effects found with all English segments, or are they restricted to stops, which are transferred from Korean? /p, b, t, d, f, v, θ, δ/ were placed after /a/. Two tasks are compared: (A) Identification. Forty Korean learners identified coda consonants produced by four native speakers of English, using English labels. (B) Reading. Ten native English listeners identified coda consonants produced by four Korean learners. Identification errors were largely unidirectional; for example, many segments were labeled /f/, while /f/ identification was highly accurate. Reading errors were largely bidirectional with a systematic tendency for more errors, resulting in voiceless consonants and stops, conforming to the Korean neutralization pattern. Such neutralization effects were as prevalent in fricatives as stops. [Work supported by NSF.]

3aSCb7. Perceptual learning about word boundaries with familiar and unfamiliar voices and accents. Rachel Smith (Dept. of English Lang., Univ. of Glasgow, 12 University Gardens, Glasgow G12 8QQ, U.K., R.Smith@englang.arts.gla.ac.uk)

An experiment investigated perception of speaker-specific variation in phonetic detail around word boundaries. Two male speakers of standard Southern British English (SSBE) read 24 phonemically-identical sentence pairs that differed in the location of a word boundary, e.g., *So he diced them* versus *So he'd iced them*. In pre- and post-tests, 80 SSBE subjects heard the sentences in cafeteria noise (SNR 2 dB) and typed the words they heard. Between these tests, subjects received 40 minutes' training: They heard different tokens of the same sentences without noise in a disambiguating context, and answered questions about their meaning. A counterbalanced between-groups design manipulated the voices heard in tests (speaker 1 or 2; always the same in pre- and post-test) and in training (Same as in tests, or Different). Training with the Same voice caused significantly more improvement in identification of words and syllable constituents at word boundaries than training with a Different voice. Preliminary data from 13 subjects with a different accent (Glaswegian Scottish) indicate, in contrast, more improvement after training with a Different voice. Apparently, when an accent is familiar, perceptual learning about word boundaries is voice-specific; when the accent is unfamiliar, learning generalizes to other speakers of the accent.

3aSCb8. Experimental paradigm influence subject's perception of attitudes. Caroline Menezes, Donna Erickson, Kikuo Maekawa, and Hideki Kawahara (NIJL, 3591-2 Midori-cho, Tachikawa-shi, Tokyo 190-8561, Japan; Showa Music Univ., Kawasaki City, Kanagawa ken, 215-8558, Japan; Wakayama Univ., Japan)

Japanese utterances spoken with different attitudes (admiration, suspicion, and disappointment) were independently manipulated for pitch-contour and voice quality using STRAIGHT {Kawahara}. Close copy stylization of the prototypical pitch contour for the three attitudes was imposed on the naturally spoken utterances, producing stimuli with all combinations of voice quality types and pitch-contour shapes. The utterances were submitted to two separate experiments, wherein subjects were asked to judge the attitude of the morphed utterances. The first, a forced choice experiment, asked subjects to choose if the utterances were admiration, suspicion, or disappointment, and the second, a free choice experiment, where subjects could freely choose the attitude they perceived. The results from the forced choice test indicated that subjects used pitch contour cues to choose speaker attitude. However, the results from the free choice paradigm indicated that subjects used both voice quality and intonation cues, and were able to perceive more affect types than the original three. Whereas the forced choice paradigm suggested that intonation was the primary cue for listeners to determine speaker attitudes, the free choice paradigm suggested an independence of voice quality and intonation in determining speaker attitudes, which resulted in a larger number of perceived attitudes by listeners.

3aSCb9. The role of prosodic boundaries in comprehension of Korean pseudo-cleft sentences. Jaehoon Jeong (Dept. of Linguist., Univ. of Hawaii, Honolulu, HI 96816, jhjeong@hawaii.edu)

A large prosodic boundary triggers, in general, a high attachment of an ambiguously attached phrase (e.g., Susie learned that Bill called // on Monday). The informative boundary hypothesis (IBH), however, claims that such an effect of a local prosodic boundary is neutralized by the presence of another boundary (e.g., Susie learned // that Bill called // on Monday). According to the IBH, it is not the absolute size of the later boundary (here, pre-PP boundary) but its size relative to any earlier boundary before a constituent containing only the lower attachment site. Since the IBH implies nothing about the interpretation of an ambiguous phrase whose attachment sites appear later, the current study, as an extension of the IBH, investigates whether the effectiveness of a prosodic boundary can be determined by its size relative to a relevant subsequent boundary. The results showed that the interpretation of the sentence-initial subject NP was affected by the absolute size of the post-NP boundary, but the later boundary type had nothing to do with the interpretation. This provides further evidence that prosodic information is processed in an incremental fashion (left-to-right) and perceivers tend to impose prosodic structures as fast as possible.

3aSCb10. On voicing activity under the control of emotion and loudness. Samuel Kim, Sungbok Lee, and Shrikanth Narayanan (Speech Anal. & Interpretation Lab. (SAIL), Univ. of Southern California, 3740 McClintock Ave., Los Angeles, CA 90089, kimsamue@usc.edu)

For a parametric study of interplay between loudness control and emotional modulation in voicing activity, electroglottography (EGG) data collected from two male and two female subjects are examined. The subjects read emotionally neutral sentences in a self-controlled manner with four different emotional states (neutral, angry, sad, happy) and three levels of loudness (soft, normal, loud). The analysis focused on the timing and shape-related EGG waveform parameters of the vowel /a/ in the data as a function of emotional state and loudness level. Specifically, open quotient (OQ), speed quotient (SQ), and noise-to-harmonic ratio (NHR) of EGG waveform are investigated. Pitch, root-mean-square (RMS) energy, and duration of speech waveform are also analyzed. Despite inter-subject differences, some general tendencies of the EGG parameters can be observed. It is found that angry emotion shows the lowest OQ in a given loudness level, and happy emotion exhibits significantly higher OQ than angry emotion. SQ mainly varies along the loudness dimension, smaller SQ in louder voice. NHR is the highest in soft voice and shows no clear emotion-dependent pattern. The results suggest that OQ is the main EGG parameter that is controlled by speakers for emotion expression, in addition to its role in loudness control.

3aSCb11. Intraglottal pressures in a static physical model of the uniform glottis: Entrance loss coefficients and viscous effects. Lewis Fulcher and Ronald Scherer (Phys. and Astron. and Commun. Disord., Bowling Green State Univ., Bowling Green, OH 43403, fulcher@bgsu.edu)

Pressure distributions for the uniform glottis were taken with a static physical model (M5) for the diameters $d=0.005$, 0.0075, 0.01, 0.02, 0.04, 0.08, 0.16, and 0.32 cm for a number of transglottal pressures of interest for phonation. At each pressure and diameter, entrance loss and exit coefficients are calculated. The pressure dependence and the diameter dependence of these coefficients are catalogued and compared with some standard values from the earlier literature. The accuracy with which tabulations of these coefficients reproduce the M5 pressure distributions is examined. To an excellent approximation, the intraglottal pressures at smaller diameters decrease linearly with the axial distance, and remnants of this behavior are seen at $d=0.08$ cm and 0.16 cm. It is shown that the intraglottal pressure gradients are linear functions of the glottal flow rates. Thus, dividing the pressure gradients by the flow rates allows one to isolate the geometric dependence of viscous effects. It is shown that an inverse five-halves power law for the diameter dependence of the viscous glottal resistance is a better approximation than the inverse cube law proposed by van den Berg, Zantema, and Doornenbal. [Work supported by NIH R01DC03577.]

3aSCb12. The effects of linguistic experience on the perception of pathologically-disordered phonation. Christina Esposito (Macalester College, 1600 Grand Ave., St Paul, MN 55105)

Ladefoged [Vocal Fold Physiology: Contemporary Research and Clinical Issues, College Hill Press, San Diego, 1983] wrote, what is a pathological voice quality in one language may be phonologically contrastive in another If one persons pathology is another persons phonemic phonation, then speakers of languages with phonation contrasts should treat pathologically-disordered and linguistically-relevant phonations in the same way. To test this, Gujarati (breathy versus modal contrast), Spanish (no breathiness) and English (allophonic breathiness) listeners participated in similarity-rating tasks. One used Mazatec (breathy, modal, and creaky phonation) stimuli, the other pathologically-disordered stimuli. Gujaratis treated pathologically-disordered stimuli and Mazatec stimuli in the same way. With both types of stimuli, Gujaratis based their judgments on H1 $-$H2 (amplitude of the first harmonic minus the second) and mapped the stimuli into clusters with similar H1$-$H2 values across the two experiments. In both experiments, English listeners relied on cepstral peak prominence and H1$-$H2, and Spanish listeners relied on H1$-$H2 and H1$-$A1 (amplitude of first formant peak). English and Spanish listeners grouped both types of stimuli into clusters with similar values for the acoustic measures. However, English and Spanish listeners differed in how they weighed the importance of the different dimensions across the experiments. To English and Spanish listeners, pathologically-disordered phonations are subtly different from linguistically-relevant ones.

3aSCb13. Analysis of voice perturbations using an asymmetric model of the vocal folds. Marco Nardone, Lewis Fulcher, and Ronald Scherer (Phys. and Astron. and Commun. Disord., Bowling Green State Univ., Bowling Green, OH 43403, mnardone1@verizon.net)

A mathematical model was developed to investigate possible causes of jitter and shimmer. The model builds on the classic, lumped element model of Ishizaka and Flanagan and allows for asymmetric motions of the vocal folds and aerodynamic imbalances. The intraglottal pressures were derived from empirical pressure data obtained from a static physical model of the larynx (M5). The mathematical model is based on ten, second-order, nonlinear, coupled, ordinary differential equations that were solved simultaneously using the software Mathematica. The solutions were analyzed graphically and numerically to identify perturbations in the fundamental frequency and amplitude of the glottal airflow. Jitter and shimmer were quantified using the jitter factor and the amplitude variability index. The results indicate that only time-dependent variations in biomechanical and aerodynamic parameters result in jitter and shimmer. The magnitudes of jitter and shimmer tend to be less than those observed in the natural sounding voice, even when the asymmetries are large. Although time-independent asymmetries may cause the vocal folds to oscillate out of phase or with different amplitudes, they tend to entrain and vibrate at a common frequency. [Work supported by NIH R01DC03577.]

3aSCb14. Phonetic correlates of phonological register in Takhian Thong Chong. Christian DiCanio (Dept. of Linguist., Univ. of California, Berkeley, 1202 Dwinelle Hall, Berkeley, CA 94720-2650)

The author's phonetic fieldwork from the Takhian Thong Chong language demonstrates that the four phonological registers are distinguished phonetically with distinct voice quality (phonation) contours, pitch contours, and durational properties. Both EGG and acoustic data were gathered from seven native speakers. A dynamic open quotient contour extracted from electroglottographic recordings demonstrates register-specific phonation type correlates. The four distinct registers are characterized by: Modal phonation throughout with a mid-level pitch, a contour from modal to tense phonation with a high rising and falling pitch contour, a contour from breathy to modal phonation with a mid-falling pitch contour, and a contour from breathy to tense phonation with a midrising and falling pitch contour. Investigations into the correlation between OQ (open quotient) values and pitch indicate a positive global correlation between the relative length of the glottal cycle and the open period, but a lack of correlation between dynamic OQ and pitch curves. Aside from providing a phonetic description of a typologically rare phonological contrast, these findings suggest that there is not a strong relationship between absolute changes in degree of glottal abduction and changes in pitch, contrary to previous claims (Silverman 1997). [Work supported by UC Berkeley Graduate Division.]

3aSCb15. Response bias, type and token frequency, and prosodic context in segment identification. Noah Silbert and Kenneth de Jong (Dept. of Linguist., Indiana Univ., Memorial Hall, Rm. 322, 1021 E. Third St., Bloomington, IN 47401)

Frequency effects are well documented in a variety of linguistic domains. However, the relationship between response bias and frequency in segment identification has received little attention. We hypothesize that

listeners are biased toward higher frequency segments. In order to evaluate this hypothesis, maximum likelihood response bias parameters from the similarity choice model (Shepard, 1957; Luce, 1963) were obtained for a number of previously published segment identification data sets and rank-order correlated with a number of type and token frequency measures calculated from the Hoosier Mental Lexicon. Results indicate that bias tends to correlate more highly with type than with token frequency. Furthermore, in native and non-native individual listener data [Cutler et al. (2004)], correlations are slightly higher between bias and prosodically conditioned frequency than between bias and position-independent frequency, and correlations between bias and coda frequency are large and positive, whereas bias is not correlated with onset frequency. In addition, coda and position-independent measures of (type) frequency correlate more highly with one another than either do with onset measures. Accuracy variation and range of frequency distributions do not account for these effects. No other explanations for the observed prosodic differences are apparent. [Work supported by the NSF.]

3aSCb16. Model-based quantification of pathological voice production. Raphael Schwarz, Dimitar Deliyski (Dept. of Commun. Sci. and Disord., Univ. of South Carolina, 1621 Greene St., Columbia, SC 29208), Joerg Lohscheller, and Michael Doellinger (Univ. Hospital Erlangen, 91054 Erlangen, Germany)

Hoarseness, the primary symptom of voice disorders, results from irregular vocal fold vibrations. The oncological therapy of laryngeal cancer may even result in a total loss of voice if an excision of the larynx, and thus, the vocal folds, is necessary. State-of-the-art voice rehabilitation technique in this case is the utilization of scarred tissue in the upper part of the esophagus for substitute voice production. The quality of laryngeal voice, as well as the substitute voice, primarily depends on the anatomy and the vibration patterns of the voice-producing element. Using endoscopic high-speed recordings, the voice generators are observed during voice production. In this work, a model-based approach feasible for the analysis and objective quantification of vocal fold vibrations, as well as the PE dynamics, is presented. By means of an automatic parameter optimization, the dynamic of a biomechanical model of the considered voice-producing element is fitted to the recorded vibration patterns. Thereby spatial and temporal properties of the vibrations are incorporated. The resulting values of the optimization parameters represent an objective quantification of the vibration patterns. In addition, the model parameters enable an approximation of physiological tissue parameters as stiffness and mass distribution.

3aSCb17. Neural representation of pitch-shifted voice feedback at N100 and P200 components: An event-related potentials (ERP) study. Roozbeh Behroozmand, Hanjun Liu, and Charles R. Larson (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208)

Numerous studies have shown that perturbations in voice pitch feedback lead to compensatory responses in voice fundamental frequency (F0), but few have reported on neural responses to the acoustical stimulation during the pitch-shifted voice feedback. In this study, randomized-onset, upward pitch-shift stimuli (60 ms duration) were presented to ten subjects during sustained vowel phonation while the event-related potentials (ERPs) were recorded through Cz-linked earlobe surface electrodes. ERPs were also recorded as subjects passively listened to the same feedback signal that was recorded during vocalization. The experimental task was repeated for 100, 200, and 500 cents conditions, respectively, and the results were compared in vocalization and listening conditions. Across three stimulus magnitudes, the N100 responses during vocalization were suppressed in comparison with those recorded in the passive-listening condition. On the other hand, the P200 responses in the passive-listening task were graded in amplitude according to the magnitude of the stimulus, with the smallest responses for 100 cents and largest for 500 cents stimuli. However, during vocalization, the P200 responses were found to be graded in the opposite way, where the largest responses were associated with 100 cents and the smallest with 500 cents stimuli.

3aSCb18. Detecting rhythmical prominence in speech by an optimized convolution kernel. Christina Orphanidou and Greg Kochanski (Phonet. Lab., Univ. of Oxford, 41 Wellington Sq., Oxford OX1 2JF, UK)

We present an approach for detecting rhythmical prominence in read speech. A production experiment was conducted during which subjects repetitively read out speech to a metronome, trying to match stressed syllables to its beat. In the analysis, we compute a function from the speech waveform, related to acoustic properties of speech such as specific loudness, pitch, voicing, and spectral slope. The function is then convolved with a Mexican Hat convolution kernel. Taking large maxima in the function to be predictions of the metronome ticks, we adjust the parameters of the signal to maximize the accuracy of the predictions. The parameters are adjusted by minimizing the phase variation between metronome ticks and ticks predicted from the audio, over a specified time interval. We confirm the results by Bootstrap resampling. We find that the most important factor is the contrast in specific loudness between a syllable and its neighbors. The prominence can be deduced from the specific loudness in an (approximately) 360 ms window centered on the syllable in question relative to an (approximately) 800 ms-wide symmetric window.

3aSCb19. The influence of position-in-utterance on the scale of articulatory gestures. Christina Kuo and Gary Weismer (Dept. of Communicative Disord. and Waisman Ctr., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706)

The influence of prosodic structures of utterances on articulation has been of interest in speech production. Data on fundamental frequency, peak oral airflow, movement, and speech timing has been reported. Nonetheless, little is available in the literature regarding the magnitude of articulatory behavior as a prosodic phenomenon. This study is motivated by the phenomenon of articulatory declination, which is a hypothesized gestural weakening over an utterance, analogous to the better known F0 declination. An attempt is made to describe and examine articulation at the start and the end of an utterance under varied stress during sentence reading of normal speakers aged 20 to 30. Target words were ''row,'' ''sew,'' and ''sigh.'' A control word ''bee'' was included given the diphthongal nature of the three targets. Each target word was surrounded by the neutral vowel ə and embedded in a grammatical sentence under five position-in-utterance conditions: initial, initial-stressed, final, final-stressed, and final without a following ə. Formant transitions were measured from the first glottal pulse after the consonant through the following ə. Transition extent and duration were derived for analysis. Pilot data with two participants (one male, one female) and work in progress are discussed within the framework of articulatory declination.

3aSCb20. Gradiency and categoricity in prosodic boundary production and perception. Jelena Krivokapic (Dept. of Linguist., Univ. of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693) and Sankaranarayanan Ananthakrishnan (Univ. of Southern California, Los Angeles, CA 90089)

A study investigating whether the production and perception of prosodic boundaries is categorical or gradient is presented. Most theories assume a small set of prosodic categories which are marked by categorically different prosodic boundaries (e.g., Beckman & Pierrehumbert 1986 Phonology Yearbook, 3, 255–309, Nespor & Vogel 1986 Prosodic Phonology), while an alternative view suggests the possibility of gradiently varying prosodic boundaries (e.g., Byrd & Saltzman, 2003 J. of Phonetics, 31, 149–180). The first part of this study is an articulatory study investigating the production of twenty-four prosodic junctures ranging from no

boundary to very strong boundary. In the second part of the study listeners evaluate the perception of these same boundaries. Two evaluations (histograms and fitting mixture distributions) were conducted on both the production and on the perception data. The production of prosodic boundaries was found to be categorical, showing a large (IP) and a predominantly small prosodic boundary. The perception of prosodic boundaries showed that listeners perceive two distinct categories, but also that the data is better explained if more prosodic categories, up to 8, are assumed. [Supported by NIH.]

3aSCb21. Discrimination of lexical tone contrasts in disyllabic nonsense pairs by adult speakers of Mandarin and of English. Shari Berkowitz and Winifred Strange (CUNY Grad. Ctr., Speech Acoust. and Percept. Lab, 365 Fifth Ave., New York, NY 10016, sberkowitz1@gc.cuny.edu)

Most previous research on cross-language perception of lexical tone has used monosyllabic stimuli. In running speech, coarticulation affects the surface forms of the tones, creating contours that may be much more difficult to discriminate. Two tokens per type of disyllabic Mandarin nonsense words served as stimuli in a categorial same/different task. Pairings were chosen to be challenging, with some easy control pairs. Native Mandarin listeners ($N=5$) and American English listeners ($N=15$) with no background in tone languages completed 224 trials with feedback on 14 contrasts, then completed two additional blocks of 224 trials without feedback. Mandarin speakers were at ceiling on all contrasts; for Americans, accuracy did not improve across training and test blocks. Overall accuracy for different pairs averaged 78% (range 54–99%), and overall accuracy for (categorial) same pairs averaged 84% (range 75–97%). The mean A' score for Americans was 0.88 (range 0.75-0.96). Performance across the 14 contrasts varied from $A'=0.74$ to 0.96, and showed significant context effects. Error analysis showed a pattern of errors across tone pairs that supported the hypothesis that American speakers relied on the overall pitch contour across the two syllables, rather than a syllable-by-syllable analysis.

3aSCb22. An exemplar-based approach to automatically detect burst in word-initial voiceless stops in spontaneous speech. Yao Yao (Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720, yaoyao@berkeley.edu)

This paper discusses a methodological question: How to automatically extract phonetically important but untranscribed information (such as VOT and closure duration) from a large spontaneous speech corpus? In this study, we made use of the exemplar-based similarity scores, first developed in Johnson 2006, to find the point of burst in voiceless stops in the Buckeye speech corpus. A similarity score is in essence a measure of how similar one piece of acoustic information is to another on the spectrogram. When used together with phonal spectral templates of the speaker, it can be used to measure how similar a piece of acoustic data is to, say, the typical [aa], or the typical [f] of the same speaker. We find that a two-dimensional score vector (<silence> score plus <sh> score) is already adequate to recognize burst in stops and the pattern is robust enough across a wide range of context in uncontrolled spontaneous speech. It works confidently on average around 90% of the time; the mean error is estimated to be around 3–5 ms (the optimal error in theory is 2.5 ms due to the 5 ms step size in similarity scores). Its robustness, accuracy, and efficiency is examined in detail in the paper.

3aSCb23. Listener sensitivity to spectral slope attributes. Jody Kreiman and Bruce R. Gerratt (Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794)

Researchers have long known that the shape of the vocal source spectrum is an important determinant of vocal quality, but details regarding the perceptual importance of individual spectral features remain unclear. This study provides preliminary evidence about the perceptual importance of four acoustic source features: H1-H2 and the spectral slopes from 1.52 kHz, from 24 kHz, and above 4 kHz. Vowel stimuli were synthesized by varying each spectral parameter in steps. Because the perceptual salience of source parameters depends on F0 and on the inharmonic source spectrum, different series were created for a male and a female speaker, and for a source with little versus a large amount of inharmonic energy. Listeners heard all possible pairs of voices within each series and were asked whether stimuli were the same or different. We hypothesize 1) that listeners sensitivity to H1-H2 will greatly exceed their sensitivity to other parameters; and 2) that listeners sensitivity to H1-H2 and the slope of the spectrum from 1.52 kHz will be independent of noise, but that sensitivity to changes in the spectral shape above 2 kHz will depend on the presence of noise excitation in the voice.

3aSCb24. Robust and accurate F0 estimation for reverberant speech by utilizing complex cepstrum analysis. Masashi Unoki and Toshihiro Hosorogiya (School of Information Sci., JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan)

This paper presents comparative evaluations of 12 typical methods of estimating the fundamental frequency (F0) over huge speech-sound datasets in reverberant environments. They involve several classical algorithms such as cepstrum, AMDF, LPC, and autocorrelation methods. Other methods involve a few modern algorithms, i.e., instantaneous amplitude and/or frequency-based algorithms, such as TEMPO, IFHC, and PHIA. The comparative results revealed that the percentage of correct rates and SNRs of the estimated F0s were reduced drastically as reverberation time increased. This paper, thus, proposes a method of robustly and accurately estimating F0 in reverberant environments by utilizing the MTF concept and the source-filter model in complex cepstrum analysis. The MTF concept is used in this method to eliminate dominant reverberant characteristics from observed reverberant speech. The source-filter model is used to extract source information from the processed cepstrum. Finally, F0s are estimated from them by using the comb-filtering method. Additive-comparative evaluation was carried out on the proposed method and other typical methods. The results demonstrated that it was better than the previously reported methods in terms of robustness and in providing accurate F0 estimates in reverberant environments. [Work supported by a Grant-in-Aid for Science Research from the Japanese Ministry of Education No. 18680017.]

Session 3aUW

Underwater Acoustics: Time Reversal and Matched Field Processing

Claire Debever, Chair
Scripps Inst. of Oceanography, Marine Physics Lab., La Jolla, CA 92093-0238

Chair's Introduction—8:00

Contributed Papers

8:05

3aUW1. Broadband, coherent inter-array processing of horizontal arrays. Claire Debever and W. A. Kuperman (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238, cdebever@ucsd.edu)

We study source localization in a shallow water environment using broadband data recorded on two horizontal arrays lying on the sea floor, each of length 250 m. They each had about 25 elements and were approximately 3.5 km apart [N. O. Booth et al., IEEE, JOE, **25**(3), (2000)]. The data on the arrays were produced by a source tow covering ranges of about 1 km to 10 km emitting tones from 50 Hz to 400 Hz. The data were processed using various conventional and adaptive algorithms for all the cases covering incoherent frequency and incoherent inter-array processing to coherent frequency and coherent inter-array processing, as a function of source position relative to the arrays. The results are compared to simulations. [Work supported by ONR.]

8:20

3aUW2. Bistatic target detection using a time-reversal operator. Geoffrey F. Edelmann, David M. Fromm, and Charles F. Gaumond (Acoust. Div., Naval Res. Lab., 4555 Overlook Ave., Code 7145, Washington, DC 20375)

Previous work has shown that time-reversal operator (TRO) methods can detect ocean targets using a time-reversal mirror [C. Gaumond, J. Acoust. Soc. Am **119**, 976–990 (2006)]. To increase the signal-to-noise ratio, the TRO was measured by transmitting a set of orthogonal beams. The method is extended to a bistatic case where the receiver and transmitter are no longer collocated. Simulated deep ocean convergence zone and up-slope detection at 200 Hz will be discussed. The consequences of array geometry (horizontal and vertical) to beam orthogonality will be addressed. Specifically, under what conditions will a horizontal-receive array have enough aperture to measure a TRO of sufficient rank? It will be shown that the system must be able to both transmit and receive beams that are orthogonal with respect to the waveguide (and not simple free space arguments). [Work supported by the Office of Naval Research.]

8:35

3aUW3. Extraction of a backscattered target signature in a shallow water waveguide with decomposition of the time reversal operator method. Franck Philippe, Claire Prada, Julien de Rosny, Dominique Clorennec, and Mathias Fink (LOA, ESPCI, 10 rue Vauquelin, 75005 Paris, France)

In a shallow water waveguide, detection and characterization of a target is an active field of research. Retrieving the target signature is extremely difficult due to the complex mode coupling between the incident wave and the backscattered wave. One solution consists of deconvolving the target response from the waveguide impulse response. However this method needs an extensive a priori knowledge of the waveguide geometry [Mignerey *et al.*, JASA 1992 and Yang *et al.*, JASA 1994]. To overcome this difficulty, we propose to apply the DORT method (French acronym for decomposition of the time reversal operator). This method is based on time-reversal invariance. Time reversal is commonly known to be an effective means to focus on a target without any a priori knowledge of the propagation medium. In the same way, here we show that the eigenvalues of the time reversal operator are directly proportional to the backscattered target signatures, and the waveguide geometry only acts on the proportionality factor. A theory based on modal waveguide analysis is presented. Ultrasonic scale experiments prove the effectiveness of this method.

8:50

3aUW4. Effect on surface ship interference on adaptive beamforming for the shallow water array performance (SWAP) project. Richard L. Campbell, Jr. and Lisa M. Zurk (NEAR-LAB, ECE Dept., Portland State Univ., PO Box 751, Portland, OR 97207-0751, rlcamp@pdx.edu)

The shallow water array performance (SWAP) project is designed to explore the limits of large-aperture passive sonar array processing capability in a shallow-water environment with moving surface ship interference. The array, off the eastern coast of Florida near Ft. Lauderdale, has 500 elements with a total length of approximately 900 m. One challenge for the processing is the length of observation time needed by traditional adaptive beamforming formulations—with an array of this size and element quantity, the resolution of range and bearing cells is such that a ship may move across many cells during the snapshot time, spreading the resulting eigenvector structure and decreasing effective signal gain. A central question is the trade-off between array gain and this eigenvector spreading loss. To explore this question, tracks from actual ships in the vicinity of the array site, combined with sound speed and bathymetry data from the site, are used in an adiabatic normal mode simulation to predict the acoustic response across the array. The resulting simulated snapshots are used in adaptive and non-adaptive formulations to predict target detection performance as a function of the interference environment and processing parameters, for both full and sub-aperture processing schemes.

9:05

3aUW5. Source localization and interference suppression using mode space estimation. Kyungseop Kim, Woojae Seong (Dept. of Ocean Eng., Seoul Natl. Univ., Seoul 151-742, Korea), and Seongil Kim (Agency for Defense Development, Jinhae, Korea)

Weak target detection and localization in the presence of loud surface ship noise is a critical problem for matched field processing in shallow water. For stationary sources, each signal component of a received signal can be separated, and interference can be suppressed using eigen-space analysis schemes. However, source motion, in realistic cases, causes spreading of signal energies in their subspace. In this case, eigenvalues of target and interference signal components are mixed and hard to be separated with usual phone-space eigenvector decomposition (EVD) approaches. Our technique is based on mode space and utilizes the difference

in their physical characteristics of surface and submerged sources. Performing EVD for modal cross-spectral density matrix, interference components in the mode amplitude subspace can be classified and eliminated. This technique is demonstrated with data obtained from an L-shaped (vertical and horizontal) line array during the MAPLE IV experiment conducted in the East Sea, Korea, and results will be discussed.

9:20

3aUW6. Depth dependence of planewave beamformed data with a horizontal array in a waveguide. Kevin Cockrell and Henrik Schmidt (Dept. of Mech. Eng., Massachusetts Inst. of Tech., Rm. 5-204, 77 Massachusetts Ave., Cambridge, MA 02139, cockrell@mit.edu)

Despite extensive effort by the ocean acoustics community to design signal processing techniques that take advantage of the waveguide structure of ocean environments (e.g., matched field processing), planewave beamforming remains the only technique robust enough to be used with autonomous vehicles processing data in real-time. However, the waveguide affects the beamformed data in ways that can degrade the performance of bearing estimation. Precisely how the waveguide affects the beamformed data depends on several parameters: The acoustic source's range, the acoustic source's depth, the environmental parameters, and the horizontal array's depth. All of these parameters are ''out of our control'' except the depth of the towed array. Thus, numerical simulation is used to determine if there is an optimal depth to tow the horizontal array, which will minimize degradation to bearing estimation. [Work sponsored by ONR.]

9:35

3aUW7. Target scattering waves emitted from a time reversal array in shallow water. Yoshiaki Tsurugaya (NEC Corp., 1-10, Nissin-cho, Fuchu, Tokyo 183-8501, Japan, y-tsurugaya@bp.jp.nec.com), Toshiaki Kikuchi (NDA, Yokosuka, Kanagawa 238-0024, Japan), and Koich Mizutani (Univ. of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan)

When a target exists between a sound source and a TRA in shallow water, the TRA receives the sound waves from the sound source and the waves scattering by the target. If the time reversal processing to them is carried out and they are re-emitted from the TRA, it will be thought that they are converged on the position of the sound source and the target. However, since the waves converging on the sound source have a high level, the waves converging on the target position are usually masked by the high level sounds. We cannot observe the waves converged on the target. We then eliminate only the high level sounds from the sound fields. In each array element, the signals in the case of nontarget are subtracted from the signals including the target. As a result of subtraction, the components of the scattering wave by the target are left on the array elements. The time reversal fields of the scattering wave are constructed by radiating the components of the scattered wave from each element again.

9:50

3aUW8. Identification of a resonant target buried in sediment using iterative, single channel, time reversal. Zachary J. Waters, R. Glynn Holt, Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, ronroy@bu.edu), and Benjamin R. Dzikowicz (NSWC-Panama City, Panama City, FL 32407)

The presence of noise and clutter makes identification of targets buried beneath the seafloor a challenging problem. Iterative time reversal using a single-channel transducer is shown to enhance the signal-to-noise ratio of backscattered echo returns from a buried resonant target. Each iteration consists of: (1) Insonifying the target with a broadband pulse, (2) windowing a portion of the backscattered echo return, (3) reversing it in time, and (4) using this waveform as the source signal for the subsequent interrogation. Scaled laboratory experiments are performed with a broadband ($rmQ \sim 2$) transducer, operating between 500 kHz and 2 MHz, and a 6.35 mm diam hollow aluminum spherical shell target buried beneath a layer of simulated sediment. When the target is located within the time reversal window, there is a rapid convergence to a narrowband signal characteristic of a dominant mode of the target's scattering response. Images are generated by scanning the transducer laterally in two dimensions above the buried target. The images reveal enhancement of different resonant modes of the target, depending on the transducer's position. Results will also be reported from larger-scale experiments performed in a test pond at lower frequencies. [Work supported by Office of Naval Research Award No. N000140610044.]

10:05

3aUW9. Pressure sensitivity kernels applied to time-reversal acoustics. Kaustubha Raghukumar, Bruce D. Cornuelle, William S. Hodgkiss, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238)

Time-reversal has been found to be more robust to sound speed perturbations than a one-way transmission. Various explanations have been advanced to account for the lower sensitivity of time-reversal. In this contribution, the robustness of time-reversal is quantitatively examined using sensitivity kernels. A first-order Born approximation is used to obtain the pressure sensitivity of the received signal to small changes in medium sound speed. The pressure perturbation to the received signal caused by medium sound speed changes is expressed as a linear combination of single-frequency sensitivity kernels weighted by the transmit signal in the frequency domain. This formulation can be used to predict the response of a source transmission to sound speed perturbations. The stability of time-reversal is studied and compared to that of a one-way transmission using sensitivity kernels. In the absence of multipath, a reduction in pressure sensitivity using time-reversal is only obtained with multiple sources. This can be attributed both to the presence of independent paths and to cancellations that occur due to the overlap of sensitivity kernels for different source-receiver paths. [Work supported by ONR.]

Session 3pAA

Architectural Acoustics: Acoustics of Rehearsal Facilities

Damian J. Doria, Chair
Artec Consultants, 114 W. 26th St., 12th Floor, New York, NY 10001

Invited Papers

1:00

3pAA1. Acoustics of university rehearsal spaces: Not too live, not too dead. Russell Cooper (Jaffe Holden Acoust., 114A Washington St., Norwalk, CT 06854, rcooper@jaffeholden.com)

The design of music rehearsal spaces for university schools of music must balance many factors. Most important of these is loudness and decay time. Rooms must have sufficient floor area and volume so that the rooms are not overly loud for the teachers and students. The size and volume of the room depend on the type of ensemble rehearsing in the room (marching band, concert band, wind ensemble, choral, orchestra, percussion, jazz/rock, etc.). Decay times are often desired to match that of the performance stage. We have found this not to be ideal, that in fact rooms should be less reverberant than the stage platform, to provide students and teachers the ability to hear their instrument's articulation and syncopation. A certain amount of adjustable acoustics are necessary to account for multi-use of the room for different ensembles, and finishes should be selected for the proper diffusion of sound and to achieve the desired decay time. This paper will discuss many of our firm's designs and some of the successful common design themes we use.

1:25

3pAA2. Straying outside the rehearsal room box. David Conant and William Chu (McKay Conant Brook Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362)

Acoustical considerations are seldom the singular driving factor in rehearsal room design. The desire in academia, to maximize use of precious construction footprint by accommodating an increasing variety of programs, while simultaneously adjusting to ever-decreasing budgets, has resulted in the emergence of hybrid rehearsal spaces that defy conventional classification. The described spaces by MCB serve musical groups with divergent acoustical demands and are variously, yet successfully, pressed into service for lectures, recitals, recording studios, and even experimental theater.

1:50

3pAA3. School music rehearsal spaces—How dry is dry enough? Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066)

Handbook values for optimum reverberation time (RT60) in a music studio, such as a school band or chorus rehearsal space, fall into the 0.8 to 1.0 second range, depending on volume. One music room furniture manufacturer specifies RT60 values as high as 1.91 seconds. We submit that these RT60 values are excessive. The Classroom Acoustical Standard (ANSI S12.60-2002) specifies a maximum allowable RT60 of 0.7 seconds to achieve good speech intelligibility. Obviously, good speech communication is necessary for effective education of any kind. In special function rooms designed for music instruction, experience has shown that the design target RT60 should be as low as 0.2 seconds, for frequencies of 500 Hz and above. The requirement for such low reverberation is dictated by the need to reduce the extremely high sound levels that can be generated by a large group; and to give the instructor a dry environment that provides the ability to isolate, identify, and evaluate the musical performance of an individual student while rehearsing in a larger group. This allows the instructor greater opportunity to reinforce good student musicianship and to correct individual student mistakes, ultimately resulting in a higher level of concert performance. Sonic examples are given.

Contributed Papers

2:15

3pAA4. Remedial acoustical design of a college band instrument rehearsal room modified because of asbestos treatment. Steven D. Pettyjohn (The Acoust. & Vib. Group, 5700 Broadway, Sacramento, CA 95820-1852, spettyjohn@acousticsandvibration.com)

Removal of asbestos material from an instrument rehearsal room at CSU, Sacramento, including acoustical treatment led to a new design. The fix dramatically altered the characteristics of the room and resulted in unacceptable conditions. Temporary installation of heavy velour curtains made the room tolerable, but a permanent fix was sought. The room has nonparallel surfaces, large and long, barrel-shaped diffuser panels, hard walls and floor, and a mostly hard ceiling. Field measurements were made under three conditions to quantify the room's acoustical conditions, including reverberation time and background sound levels. One-third octave band measurements were made from 12.5 to 10,000 Hz to cover all instruments and background sources. The 1-inch acoustical treatment first added to the room was not adequate to attenuate sound below 500 Hz, and insufficient to handle the high acoustical energy produced by the band. The echo between the floor and ceiling was very distinct, as neither surface had any treatment. Calculations using two different methods showed that substantial low-frequency sound absorption was required to reduce the sound at all frequencies to the design goal. The room is currently undergoing the modification, and results may be available at the time of the presentation.

2:30

3pAA5. Improvement of reverberation and sound isolation in a high school music rehearsal facility. Clemeth L. Abercrombie (Daly-Standlee and Assoc., 4900 SW Griffith Dr., Ste. 216, Beaverton, OR 97005, cabercrombie@acoustechgroup.com)

Music rehearsal facilities are used by some of the most acoustically demanding individuals; extremely discerning even at an early age in their musical path. This paper examines acoustical issues in a high school marching and concert band rehearsal suite that is home to approximately 70 students. Even though the two-year-old building met acoustical design goals provided by an acoustical consultant, the music program director using the facility was concerned enough with acoustical performance that he successfully lobbied the school district for a renovation. Outstanding issues included clarity and low-frequency response in the main rehearsal room, sound isolation and loudness in individual practice rooms directly adjacent to the main room, and noise experienced in the nearby choir rehearsal suite. This presentation will review the importance of acoustical parameters specific to rehearsal, as well as the implementation of custom designed low-frequency absorption and cost-effective sound-isolation improvements.

2:45

3pAA6. Concert hall canopy height comparison and computer model resolution efficacy. Michael Ermann and Indhava Kunjara Na Ayudhya (Virginia Tech. School of Architecture + Design, 201 Cowgill Hall (0205), Blacksburg, VA 24061, mermann@vt.edu)

This line of inquiry is twofold: (1) It uses computer modeling to compare the height of an adjustable concert hall ceiling canopy with established room acoustics metrics, and (2) It does so simultaneously with two computer models, one at a rough resolution, and one at a resolved resolution. The height comparison looks to establish best practices in operation of such a canopy in a hall; how high should it be set? The model resolution comparison looks to establish best practices in room acoustics computer modeling; how detailed must the model be?

Session 3pBB

Biomedical Ultrasound/Bioresponse to Vibration: Biomedical Applications of Acoustic Radiation Force

Mostafa Fatemi, Chair
Mayo Clinic, Dept. of Physiology and Biophysics, Rochester, MN 55905

Chair's Introduction—1:10

Invited Papers

1:15

3pBB1. Modulated ultrasound and multifrequency radiation force. Matthew W. Urban, Mostafa Fatemi, and James F. Greenleaf (Dept. of Physiol. and Biomed. Eng., Mayo Clinic College of Medicine, 200 First St., SW, Rochester, MN 55905)

Modulating ultrasound can create dynamic ultrasound radiation force to induce local tissue vibration. Using different types of modulating signals produces radiation force with weighted multifrequency components. Modulation signals with frequency f_r produce force with components at multiples of f_r. Both amplitude modulation (AM) and double sideband suppressed carrier (DSB-SC) amplitude modulation were explored. Different waveforms were investigated such as sine, square, triangle, and sawtooth for modulating continuous wave ultrasound. A 3.0 MHz transducer produced ultrasound which was measured with a needle hydrophone and used to derive and analyze the spectrum of the radiation force function. The modulated ultrasound was used to vibrate a steel sphere in a gelatin phantom, and the motion was measured using a laser vibrometer. The measured spectra were used to find the frequency response of the sphere. Different modulating signals were used to produce multifrequency radiation force. The sphere frequency response was recovered using different modulating signals. Weighted multifrequency radiation force was produced by modulating ultrasound with different signals. Application of multifrequency radiation force can be used to measure the frequency response of objects or tissue. [This work was supported in part by grants EB002167, EB002640, and CA091956 from NIH.]

1:40

3pBB2. Negative radiation forces on spheres illuminated by Bessel beams: Modeling using finite elements. David B. Thiessen and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, thiessen@wsu.edu)

An analytical solution for the scattering of an acoustic Bessel beam by a sphere centered on the beam [P. L. Marston, J. Acoust. Soc. Am. **121**, 753–758 (2007)] has made it possible to explore the way the acoustic radiation force on elastic and fluid spheres depends on beam and material parameters. Situations have been previously noted where, even in the absence of absorption, the radiation force of the beam on the sphere is opposite the direction of beam propagation [P. L. Marston, J. Acoust. Soc. Am. **120**, 3518–3524 (2006); J. Acoust. Soc. Am. **121**, 3109 (2007)]. In the case of solid spheres, the interpretation is simplified using an analysis of the scattering [P. L. Marston, J. Acoust. Soc. Am. **122**, 247–252 (2007)]. In the present research, the finite-element method (FEM) is used to evaluate the total acoustic field in the region near the sphere. This makes it possible to evaluate the radiation force from numerical integration of an appropriate projection of the Brillouin radiation stress tensor. The result agrees with analytical results for plane wave and Bessel beam illumination. The FEM result predicts negative radiation forces for appropriate beam, material, and frequency parameters. [Supported in part by NASA.]

2:05

3pBB3. Vibro-acoustography of thyroid. Azra Allizad, Farid G. Mitri, Randall R. Kinnick, James F. Greenleaf, and Mostafa Fatemi (Mayo Clinic College of Medicine, Rochester, MN 55905, aza@mayo.edu)

Thyroid imaging with conventional ultrasound scanners does not always lead to conclusive results. Furthermore, current ultrasound imaging technology has difficulty in detecting important microcalcifications. Ambiguities in ultrasound and other thyroid imaging methods lead to a large number of unnecessary biopsies. For these reasons, alternative imaging methods capable of detecting both microcalcifications and lesions in thyroid are of great interest. This paper describes recent results on imaging human thyroids with vibro-acoustography. Experiments were conducted on eight excised human thyroids from autopsy. The specimens were each imbedded in a block of gel to simulate the surrounding tissue and to facilitate scanning. Three types of images were acquired from each specimen: X-ray, B-mode ultrasound, and vibro-acoustography. Anatomical details seen in the images were correlated and image quality was compared. X-ray images displayed calcifications with high contrast, but did not show anatomical details of soft tissue. B-mode displayed some soft tissue structures; however, the speckle masked the details and reduced the contrast. Vibro-acoustography displayed calcifications, anatomical details, and some nodules where they existed. Furthermore, vibro-acoustography images displayed tissue structures with high contrast and free from speckle. It is concluded that vibro-acoustography may be a suitable technique for thyroid imaging.

Contributed Papers

2:30

3pBB4. Acoustic radiation force biorheology. Marko Orescanin (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801), David Mahr, Sureshkumar Kalyanam, and Michael Insana (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Acoustic radiation force techniques are being developed to image rheological properties of engineered tissues and cell cultures. An embedded sphere couples an acoustic source field to the medium, while induced deformations are imaged dynamically via Doppler methods. We conducted three experiments. First, we modeled the response to a step force as a second-order differential equation. Fitting Doppler data to the model, we estimated shear modulus (G) and coefficient of viscosity (ν) for a broad range of tissue-like hydrogels. Measuring a 2% hydrogel over 8d, we estimated G=450–800 Pa, ν=0.26–0.4 Pa s, in close agreement with measurements using a cone-plate viscometer. In the second experiment, we applied harmonic acoustic radiation force stimuli and measured the complex modulus of the gels over a very broad bandwidth. The advantage obtained is significantly higher SNR. Finally, spatial variations in gel rheological properties were mapped from shear waves radiating from the harmonically-driven embedded sphere. The wavelength of the shear wave indicates the shear modulus. These three methods describe spatiotemporal variations in scaffolds designed for tissue engineering and cancer cell cultures, all without physically touching the samples. Acoustic radiation force rheometry is being developed as a tool for basic biological research.

2:45

3pBB5. Estimation of tissue stiffness with the surface wave generated by ultrasound. Xiaoming Zhang, Matthew W. Urban, Randall R. Kinnick, and James F. Greenleaf (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, Zhang.xiaoming@mayo.edu)

The mechanical response of tissues to external forces has gained considerable interest in medical diagnosis. Typically, a force is needed to produce displacement in the tissue for estimation of the tissue's elasticity. One approach is to apply an external force on the body's surface. Another emerging technique is to generate a localized force inside the tissue with ultrasound. Shear waves can be used in bulky tissues for estimating the tissue's elasticity. The bending wave can be generated and measured in vessels for noninvasive estimation of a vessel's elasticity. We have found that the Rayleigh surface wave can be generated on the tissue's surface, [X. Zhang et al., Proc. International Congress on Ultrasonics, Vienna, April 9–13, 2007], [X. Zhang et al., IEEE Trans. Medical Imaging, **26**, 843–852 (2007)]. In our method, a localized radiation force of ultrasound is remotely and noninvasively applied inside the tissue. This force can generate the shear wave as well as compression wave inside the tissue. However, only the surface wave can exist on the body' surface. The surface wave can be used to estimate and image the stiffness of tissue. Some results of using the surface wave are presented.

THURSDAY AFTERNOON, 29 NOVEMBER 2007 GRAND BALLROOM E, 2:20 TO 3:15 P.M.

Session 3pED

Education in Acoustics: Acoustics Education Prize Lecture

Mark F. Hamilton, Chair

Univ. of Texas at Austin, Dept. of Mechanical Engineering, Austin, TX 78712-0292

Chair's Introduction—2:20

Invited Paper

2:25

3pED1. Songs my students sang to me. David T. Blackstock (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, and M.E. Dept., UT Austin, Austin, TX 78712-0292)

Does the professor teach his/her students? Or do they teach the professor? While the answer to both questions is probably a qualified yes, in looking back, I see that what I know now is largely what they taught me. After a review of the research areas in which my students and I have worked, a few examples are highlighted that show that what I had expected is not how things turned out.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

James P. Chambers, Chair
National Center for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—1:05

Invited Papers

1:10

3pID1. Speech-in-noise perception and recognition. Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, abradlow@northwestern.edu) and Carol Espy-Wilson (Univ. of Maryland, College Park, MD 20742)

Recent speech research has established that, for humans, speech-in-noise and speech-in-quiet perception differ along numerous dimensions that span levels of signal encoding and linguistic representation. Listeners place more or less weight on specific acoustic cues, draw more or less on signal-independent, contextual information, and are more or less distracted by lexical neighbors depending on masker type and level. Moreover, noise has different effects at different levels for different listener populations. Machine recognition of noisy speech is well below (at least by an order of magnitude) human recognition of noisy speech. This difference is true for both additive noise and convolutive noise. Researchers have focused on the problem of machine recognition of noisy speech in several ways: Developing robust features, training systems on noisy speech, or developing speech enhancement algorithms to clean up the noisy speech signal before performing recognition. A remaining challenge for understanding how humans do and how computers should handle speech-in-noise is to develop a conceptual framework that goes beyond the division of masking effects into peripheral, energetic versus central, informational.

1:30

3pID2. Speech privacy in healthcare facilities. Gregory C. Tocci and David Sykes (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776, gtocci@cavtocci.com)

In response to the growing recognition in the U.S. that personal privacy requires more protection, two federal regulations were enacted. These are the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Gramm-Leach-Bliley Act of 1999 (GLBA). The first includes provisions for protecting patient healthcare information and the second for protecting personal financial records information. These focus on data privacy and only obliquely refer to speech privacy; however, the enforcement guidelines for HIPAA specifically require acceptable speech privacy consistent with the intention of HIPAA. The definition of speech privacy, how to evaluate it, and what constitutes acceptable speech privacy are not addressed in these Acts nor in enforcement guidelines. To provide much needed guidance on this, ANSI S12 Working Group 44 Speech Privacy was formed in 2006 to develop standards for speech privacy. The efforts of the working group are to be discussed. In networking with the healthcare building design profession, it became evident that the American Institute of Architects (AIA) Hospital Design Manual was in great need of general acoustical, as well as speech privacy guidelines. This presentation will discuss this and the many on-going efforts in the area of healthcare acoustics.

1:50

3pID3. Noise diagnosis using nearfield acoustical holography. Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This presentation focuses on diagnosis of air- and structure-borne sounds by using nearfield acoustical holography (NAH) with an emphasis on applications of Helmholtz equation least-squares method for various industrial projects. The presentation is structured with strong consideration towards the end result, which is the visualization and quantification of the acoustic field and ranking of relative contributions from individual noise sources. It is emphasized that, in general, this approach does not deal with actual noise control but rather with identifying the areas where such effort may yield the most cost-effective results. Comparison of NAH-based diagnosis with other techniques such as transfer path analysis and an intensity probe is presented, and their relative advantages and disadvantages are discussed. Specific examples of identifications of sound transmission paths into a passenger vehicle compartment, fuselage of an aircraft, and disk brake squeals are presented. The reconstructed acoustic quantities, such as the acoustic pressures and normal surface velocities, are compared with the benchmark results.

Session 3pSC

Speech Communication: Vowels, Vowel Tract Modeling, and Applications (Poster Session)

Catherine L. Rogers, Chair
Univ. of South Florida, Communication Science and Disorders, Tampa, FL 33620-8150

Contributed Papers

All posters will be on display from 1:00 p.m. to 3:20 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 2:10 p.m. and contributors of even-numbered papers will be at their posters from 2:10 p.m. to 3:20 p.m.

3pSC1. Speech-to-music ratio estimation using wavelets and hidden Markov models. Brett Smolenski (Res. Assoc. for Defense Conversion, 2433 Forest Ln., Marcy, NY 13403, Brett.Smolenski.ctr@rl.af.mil)

In this paper, a system capable of estimating the long-term, on the order of an utterance, speech-to-music energy ratio (SMR) is developed. The approach uses frame-based Teager energy values at the output of a seven-band wavelet decomposition as features. The features are then modeled using a two-state hidden Markov model (HMM), with each state producing observations having a 64-component Gaussian mixture. Using this approach, estimation of the long-term SMR with a standard error of less then 5% was obtained. In addition, accurate classification of music and speech with these models was still possible when both music and speech were simultaneously present, assuming one had a higher energy than the other.

3pSC2. Effect of level of presentation on scaled speech intelligibility of speakers with dysarthria. Yunjung Kim, Gary Weismer, Raymond D. Kent (Waisman Ctr., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706), and Joseph R. Duffy (Mayo Clinic, Rochester, MN)

A considerable amount of effort has been made to seek the acoustic correlates of reduced speech intelligibility in dysarthria, although some acoustic variables are under debate: For example, articulation rate, F2 slope, and vowel space have been frequently discussed in previous studies as acoustic predictors of speech intelligibility of dysarthria [Kim (2007)]. One variable that is not well understood, with respect to its effect on speech intelligibility in dysarthria, is level of presentation. This includes level generated by a speaker as well as level of presentation after the speaker has recorded utterances. Clearly, variations in level of presentation of recorded utterances will have some effect when an intelligibility test requires some segmental analysis (as in a minimal pairs test), but the potential effect when speech intelligibility is scaled is not so clear. This is an important question because of the frequent use of scaled intelligibility in dysarthria research. In this presentation, we will report data on how scaled speech intelligibility is affected by level of presentation of speech samples produced by speakers with various types and severity of dysarthria. [Work supported by NIH DC00319.]

3pSC3. Effects of clear speech on duration and fundamental frequency of vowels produced by monolingual and bilingual talkers. Catherine L. Rogers, Michelle Bianchi, Stefan A. Frisch, and Jean C. Krause (Dept. of Commun. Sci. & Dis., Univ. of South Florida, 4202 E. Fowler Ave., PCD 1017, Tampa, FL 33620)

Prosodic characteristics of vowels produced by monolingual and bilingual talkers were investigated. Ten monolingual, 15 early Spanish-English bilingual (age of onset of immersion of age 12 or earlier), and ten late Spanish-English bilingual (age of onset of immersion of age 15 or later) talkers produced the target words ''bead, bid, bayed, bed, bad,'' and ''bod'' in conversational and clear speech styles. Vowel duration was computed, and F0 measurements were made at 20%, 50%, and 80% of the vowel duration. A significant group by style by vowel interaction showed that monolingual and early bilingual talkers enhanced inherent duration differences between target vowels by lengthening long vowels significantly more than short vowels in clear speech. The vowels of the late bilingual talkers, by contrast, became more alike in duration in clear than in conversational speech. The monolingual talkers showed a falling F0 pattern from 20% to 80% of the vowel duration in both styles; the late bilingual talkers showed a flat or rising F0 pattern in both styles; and the early bilingual talkers showed a flat or rising pattern in conversational speech, but a falling pattern in clear speech. [Work supported by NIH-NIDCD #5R03DC005561.]

3pSC4. Tonal coarticulation in Yoruba: Locus equation analysis. Augustine Agwuele (Texas State Univ., San Marcos, 601 Univ. Dr., San Marcos, TX 78666)

This study examines the acoustic coarticulatory effects of tone on VCV sequences of Yoruba, when each vowel of the VCV sequences bears a different tone. The first aim is to understand how the presence of tone affects CV coarticulation in the traditional locus equations parlance as used by Krull (1987). This marks the first application of the locus equation metric to a tonal language in the study of CV coarticulation. The second aim is to determine how the stop consonant in VCV sequences is affected by prosodic overlay, independently of the vowels altered location in articulatory/acoustic space. In order to achieve the second objective, the modified locus equation regression metric that was used by [Lindblom et al. J. Acoust. Soc. Am. (2007)] to dissociate vowel context effects from rate-induced effects on consonantal F2 onsets are applied. Similar to the findings of Lindblom *et al.* (2007), the analyses document separate effect for F2 locus relative to F2 nucleus.

3pSC5. Speeded discrimination of American vowels by experienced Russian learners of English. Yana D. Gilichinskaya, Franzo Law II, and Winifred Strange (Speech Acoust. and Percept. Lab., City Univ. of New York—Grad. Ctr., 365 Fifth Ave., New York, NY 10016-4309)

This study is part of a project examining how L2 learners whose native languages have small vowel inventories (Japanese, Russian, Spanish) perceive American English (AE) vowels. Experienced Russian (RU) L2 learners' discrimination was evaluated in a speeded ABX task that included nil Experimental contrasts among adjacent height pairs and front/back pairs; four nonadjacent height pairs served as Controls. An earlier study of Japanese L2 learners indicated that AE contrasts with spectrum

and duration differences (S+D) [i:/ɪ, æ:/[g\], u:/ʊ, ɑ:/ʌ] were discriminated more rapidly, relative to Control contrasts [i:/[g\], ɪ/æ, u:/ʌ, ʊ/ɑ:], than were contrasts that differed only in spectral structure (S-Only) [ɪ/[g\], ʊ/ʌ, [g\]/ʌ, æ:/ɑ:]. Since vowel duration is not contrastive in Russian, it was hypothesized that discrimination of S-Only and S+D contrasts might be equally slow, relative to Controls, if RU listeners did not attend to duration differences. Preliminary findings confirm that reaction time (RT) difference scores (relative to Controls) were not different for S-Only and S+D contrasts. Relative RTs on both contrast types were slower than for AE listeners, whose RTs were about the same for S-Only and S+D pairs. These findings suggest that speeded discrimination is a sensitive measure of continuing perceptual difficulties of L2 learners. [Work supported by NSF.]

3pSC6. Perceptual similarity of American English and Japanese vowels for native speakers of American English and Japanese. Takeshi Nozawa (Program in Lang. Education, Ritsumeikan Univ., 1-1-1 Nojihigashi, Kusatsu Shiga 525-8577, Japan) and Elaina M. Frieda (Auburn Univ., Auburn, AL 36849)

Native speakers of American English and Japanese chose American English vowels that best represented Japanese vowels uttered in two different consonantal contexts. The two groups of subjects' responses differed noticeably, showing the effects of linguistic experience. The Japanese subjects tended to match Japanese long two-mora vowels with long, tense vowels and short one-mora vowels with short lax vowels, but the English subjects were relatively unaffected by durational differences. The Japanese subjects matched /a/ and /e/ to /æ/ and /[g\]/, respectively, but the English subjects matched /a/ with /ɑ/ or /ʌ/, and /e/ with /ɪ/ as well as /[g\]/. The Japanese subjects matched /o/ with /ɑ/, while the American subjects matched it with /oʊ/. The Japanese subjects took part in two other experiments. They equated American English vowels to Japanese vowel categories and they also identified American English vowels. They equated English /ɑ/ to Japanese /a/, and /ɑ/ was often misidentified as /æ/. The results of the three experiments show that there are discrepancies between how an American English vowel is mapped into Japanese vowel categories and how the Japanese subjects expect it to sound, which may result in the poor identification of American English vowels.

3pSC7. Vowel intrinsic structure involved in continuous speech in articulatory space. Xugang Lu, Jianwu Dang, and Satoru Fujita (Japan Adv. Inst. of Science and Technol., 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan, xugang@jaist.ac.jp)

The relation of the speech production and perception was described using the similarity of acoustic (F1-F2) space and articulatory (vowel-triangle) space for isolated vowels. However, such a relation has not been found in continuous speech. This study attempts to find out the intrinsic structure of vowels involved in continuous speech by focusing on the whole vowel structure involved in the articulation, by keeping the neighboring topological relationship of the data set. A nonlinear dimension reduction method, Laplacian eigenmap, is adopted to realize this mission. Japanese vowels were extracted from an articulatory data set that was obtained using the electromagnetic articulography system for continuous utterances. When the articulatory data were compressed from 14 dimensions to three dimensions, a nonhomogeneous structure emerged from the vowels. The vowels were distinctively categorized based on their similarity relationship revealed by their distribution on the structure along the three dimensions. For the three dimensions, the first dimension corresponds to the height of the tongue, and the second one relates to the lip rounding. The third dimension is concerned with the articulatory place along the vocal tract. The same method was also used to extract the intrinsic structure of the corresponding acoustic data set.

3pSC8. Acoustic articulatory evidence for quantal vowel categories: The feature [low]. Youngsook Jung (Speech and Hearing Bioscience and Technol., Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139) and Kenneth N. Stevens (MIT, Cambridge, MA 02139)

The goal of this study is to determine whether acoustic coupling between the first subglottal resonance F1sub (about 600 Hz) and the F1 frequency for vowels creates a region near 600 Hz in which the F1 prominence shows an irregularity. Such a finding would provide evidence for a defining quantal articulatory-acoustic relation for the distinctive feature [low]. The time course of F1 in relation to F1sub was examined for certain diphthongs and several monophthongs produced by a number of speakers of English using Chi's data [X. Chi and M. Sonderegger, J. Acoust. Soc. Am. **115**, 2540–2550 (2004)]. For the diphthongs, a discontinuity in F1 or a dip in amplitude of the F1 prominence was observed as it passed through F1sub, while for the monophthongs, F1 was usually above F1sub for [+low] vowels and below F1sub for [-low] vowels. A preliminary further study of data from the literature on F1 for vowels from various languages showed that the boundary between F1 values of [+low] vowels and those of [-low] vowels agrees with the average value of F1sub obtained from the laboratory study with English. [Supported by NIH Grant No. DC00075.]

3pSC9. Incorporating vocalic segment memories in automatic dialect identification. David M. Rojas (Linguist. Dept., Indiana Univ., MM 322, 1021 E. Third St., Bloomington, IN 47405, drojas@indiana.edu)

Systematic pronunciation differences among speakers of regional varieties of U.S. English are recognizable to other native speakers to varying degrees. This has often been demonstrated through experiments wherein listeners were asked to match a talker to his or her dialect region. Machines have also been able to identify the regional origin of a speaker to some degree, although attempts to this end have typically not been as successful as efforts to automatically identify the language of a speaker. In order to refine the dialect discrimination ability of a machine, this paper draws methodological inspiration from the area of musical artist classification, and from linguistic notions that vowels contribute more heavily than consonants to regional differences. Using this insight, an automatic dialect identification system is developed that first recognizes the more vowel-like slices of the signal, and then updates a vocalic segment memory component with Mel-frequency cepstral coefficient, formant, and pitch information from the current frame. Besides providing a means to analyze MFCC and formant trajectories, the segment memory enriches the representation of vocalic events by allowing the system to explicitly model prosodic aspects such as duration and tilt.

3pSC10. The application of a psychophysical difference metric to perceptual similarity judgments in vowels. James Harnsberger, Rahul Shrivastav (Inst. for Adv. Study of the Commun. Proc., Univ. of Florida, Gainesville, FL 32611), and Mark Skowronski (Univ. of Western Ontario, London, ON N6A 5B7, Canada)

Models of cross-language speech perception have had limited success in predicting the discriminability or perceptual similarity of non-native contrasts. These failures may be attributed partly to an inability to quantify the phonetic differences between non-native speech sounds. This study attempted to quantify such gross psychophysical differences between speech sounds, specifically by utilizing dynamic time warping (DTW) on human factor cepstral coefficients to compare the spectrum of the entire length of the speech sounds in question. This technique has been successfully applied to account for the discriminability of different non-native consonant contrasts [Harnsberger, J. D., Shrivastav, R., and Skowronski, M.; J. Acous. Soc. Am. 117, 2460, 2005]. This study extends this work to perceptual similarity judgments of vowels. Specifically, twenty native speakers of English were presented with all possible pairings of ten vowels produced by two speakers of English. Subjects were asked to rate their similarity on a seven point scale. The resulting similarity scores were then compared with the output matrix of the DTW psychophysical difference

metric for the same stimulus materials. The results showed a significant correlation (r=.60**) between the two measures, demonstrating the efficacy of the metric with a greater range of stimulus types and tasks.

3pSC11. Effects of several consonant environments on vowel formants. Michael Kiefte (School of Human Commun. Disord., Dalhousie Univ., Halifax, NS B3H 1R2 Canada)

A large body of evidence has shown that relative change in spectral pattern possesses some invariant properties for vowels across speakers in /hVd/ environment. While formant frequency plots from vowel steady-states result in categories with large overlap, better separation of vowel categories is obtained using onset and offset formant frequencies [e.g., Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. To a large extent, this invariance was also shown in other /CVC/ contexts using all combinations of /h,b,d,g,p,t,k/ [Hillenbrand *et al.*, J. Acoust. Soc. Am. **109**, 748–763 (2001)]. The present study explored spectral change in vowels produced in environments where larger differences might be expected across contexts. Nineteen men and 39 women were asked to produce fourteen Canadian English vowels in /hV/, /hVd/, /hVt/, /hVl/, /hVr/, /hVnd/, /hVg/, and /dVd/. Subjects read standard English orthographic representations of the target words which were embedded both in a sentence which indicated the rhyme of the target word—e.g., ''Swooned rhymes with hoond'' as well as in a sentence which indicated the pronunciation of onset and vowel—e.g., ''Hood sounds like hoog.'' Formant frequencies were measured and tracked for each token and differences across consonant contexts were analyzed. [Work supported by SSHRC.]

3pSC12. Speaker normalization using cortical strip maps: A neural model for steady state vowel identification. Heather Ames and Stephen Grossberg (Dept. of Cognit. and Neural Systems and CELEST, Boston Univ., 677 Beacon St, Boston, MA 02215)

Auditory signals of speech are speaker-dependent, but representations of language meaning are much more speaker-independent. Such a speaker normalization transformation enables speech to be learned and understood from different speakers. A neural model is presented that performs speaker normalization to generate a pitch-independent representation of speech sounds, while also preserving information about speaker identity. This speaker-invariant representation is categorized into unitized speech items, which input to sequential working memories whose distributed patterns can be categorized, or chunked, into syllable and word representations. The proposed model circuits fit into an emerging theory of auditory streaming and speech categorization in which auditory streaming and speaker normalization both use similar neural designs; namely, multiple cortical strip map representations of auditory signals. This design homology may clarify how speaker normalization circuits evolved from more primitive streaming mechanisms. Simulations with synthesized steady-state vowels from the Peterson and Barney (1952) vowel database achieve accuracy rates similar to those achieved by human listeners. These results are compared to behavioral data and other speaker normalization models. [Work supported in part by NSF and ONR.]

3pSC13. Vowel space and formant dynamics of vowels produced by monolingual and bilingual talkers in conversational and clear speech styles. Michelle Bianchi, Catherine L. Rogers, Stefan A. Frisch, and Jean C. Krause (Dept. of Commun. Sci. & Dis., Univ. of South Florida, 4202 E. Fowler Ave., PCD 1017, Tampa, FL 33620)

Formant frequency characteristics of vowels produced by monolingual and bilingual talkers were compared. Ten monolingual, 15 early Spanish-English bilingual (age of onset of immersion is 12 or earlier), and ten late Spanish-English bilingual (age of onset of immersion of age 15 or later) talkers produced the target words ''bead, bid, bayed, bed, bad,'' and ''bod'' in conversational and clear speech styles. Measurements of F1 and F2 were made at 20%, 50%, and 80% of vowel duration. Significant group effects for F1 and F2 at 50% of vowel duration showed similar locations and distances between the vowels for the monolingual and early bilingual talkers, except that F2 values were significantly higher for the early bilingual than for the monolingual talkers for the vowels in the target words ''bead, bid,'' and ''bed.'' Smaller between-vowel distances were found for both F1 and F2 for the late bilingual talkers, especially for the vowels in the target words ''bead, bid, bayed,'' and ''bed.'' Changes observed in clear speech were relatively modest for all three groups. Comparisons between formant dynamic vectors and angles across groups and style and between talkers showing large versus small degrees of clear speech intelligibility benefit will be presented. [Work supported by NIH-NIDCD No. 5R03DC005561.]

3pSC14. Robust unsupervised extraction of vocal tract variables from midsagittal real-time magnetic resonance image sequences using region segmentation. Erik Bresch and Shrikanth Narayanan (Univ. of Southern California, 3740 McClintock Ave., Rm. EEB400, Los Angeles, CA 90089)

The tracking of deformable objects in image sequences has been a topic of intensive research for many years, and many application-specific solutions have been proposed. In this work, we describe a method that was developed to robustly track the tissue structures of the human vocal tract in midsagittal, real-time magnetic resonance (MR) images. The goal of the algorithm is to fully automatically extract the vocal tract outline, the position of the articulators, and the tract variables to facilitate the study of the shaping of the vocal tract during speech production. The algorithm is unsupervised and requires only a one-time initialization step for a particular subject. Importantly, the tracking algorithm operates on the spatial frequency domain representation of the underlying images, and it is hence specifically fit to the data produced in the MR imaging process. The proposed method carries out a multiregion segmentation of the individual MR images using an anatomically informed model of the vocal tract whose fit to the observed image data is hierarchically optimized using an anatomically informed gradient descent procedure. The mathematical key components of the algorithm are the closed-form solution of the two-dimensional Fourier transform of a polygonal shape function and the design of alternative gradient descent flows for the iterative solution of an overdetermined nonlinear least squares optimization problem. Various examples of segmented real-time MR images and a summary of open challenges will be presented. [Work supported by NIH.]

3pSC15. Volumetric MRI acquisition and processing. Inês Carbone (Dept. Electrons, Telecommunications and Informatics/IEETA, Univ. of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal), Paula Martins, Augusto Silva, and António Teixeira (Univ. of Aveiro, 3810-193 Aveiro, Portugal)

To provide fast and potentially more accurate three-dimensional anatomic information on European Portuguese (EP) sounds, a direct acquisition of volumetric MRI information was attempted, through a 3-D spoiled fast gradient echo sequence. The viability of the used MRI acquisition protocol relies on the possibility of obtaining consistent information. For that, two different image processing approaches were attempted: First, a so-called 2.5-D method consisting in a first segmentation on the sagittal plane, multiplanar reslicing perpendicularly to the sound propagation in the vocal tract, and new segmentation in the 45 planes obtained; second, a direct 3-D level set segmentation. A method was applied to a comprehensive single speaker corpus contemplating: Oral and nasal vowels, nasal consonants, unvoiced fricatives, and laterals. Method advantages are: Reconstruction of slices with good detail in any direction, useful to obtain slices perfectly orthogonal to the vocal tract centerline; higher SNR when compared with 2-D imaging; easier 3-D visualization; and acquisition two times faster, with positive impact in the amount of material acquired in one session with less subject effort. 3-D information was especially important for the study of nasal [P. Martins et al., InterSpeech 2007, ac-

cepted] and lateral sounds, some of them with specific characteristics for EP. [Work supported by FCT, Portuguese Research Agency, by Project HERON POSC/PLP/57680/2004.]

3pSC16. MRI study of coarticulation in European Portuguese. Paula Martins (Escola Superior de Sade da Universidade de Aveiro, Univ. of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal), Inês Carbone, Augusto Silva, and António Teixeira (Univ. of Aveiro, 3801 193 Aveiro, Portugal)

For this study, a small part of a recently acquired MRI database for European Portuguese (EP) was used. The sounds studied were the 12 EP voiced and unvoiced fricatives and stop consonants, produced in a symmetric VCV context. The vowels chosen were the cardinal vowels [i, a, u]. The speaker sustained the fricative sounds and maintained stop articulation during the 5.6 s of the acquisition time (T1 TSE sequence). The contours were obtained using the seeded region growing method. The influence of the seed placement was evaluated through the generation of 100 contours, with random seeds, for each sound. It was possible to verify, for the EP, some facts already reported by other authors relative to coarticulation in other languages. In general, EP stops are less resistant to coarticulatory effects than fricatives. The EP sounds presenting the highest resistance to coarticulation are the unvoiced postalveolar fricative [ʃ] and corresponding voiced fricative. As an example of other relevant results, for stops [t], [d] and fricative [s] there is no significant effect of the vowel in the region tongue blade region, being the influence evident in the production of the stops [k] and [g]. [Work supported by FCT, Portuguese Research Agency, by Project HERON POSC/PLP/57680/2004.]

3pSC17. Anatomic development of the vocal tract during the first two decades of life: Evidence on prepubertal sexual dimorphism from MRI and CT studies. Houri K. Vorperian, E. Michael Schimek (Waisman Ctr., Univ. of Wisconsin-Madison, 1500 Highland Ave. # 430, Madison, WI 53705), Shubing Wang, Moo K. Chung, Ray D. Kent, Andrew J. Ziegert, and Lindell R. Gentry (Univ. of Wisconsin-Madison, Madison, WI)

The growth of the vocal tract (VT) is known to be nonuniform insofar as there are regional variations in anatomic maturation. This study presents quantitative anatomic data on the developing VT from 604 imaging studies (277 male; 327 female) between birth and 20 years. Data analyses include detailed assessment on the growth of the oral (anterior or horizontal) and pharyngeal (posterior or vertical) regions of the VT for both sexes. The oral region of the VT was segmented into lip-thickness, anterior-cavity-length, and orophayrngeal-width; and the pharyngeal region of the VT into posterior-cavity-length, and nasopharyngeal-length. Findings from all variables/segments indicate differences in growth trend, rate, and type (somatic versus neural) between males and females. However, it appears that prepubertal sex differences at specific age ranges are masked by overall growth rate differences between males and females. Comparing males versus females, using a limited age range of 60 months, unveiled prepubertal sexual dimorphism in the oral/horizontal region where differences appear more pronounced in the oropharyngeal-width segment. Such novel anatomic findings indicate one possible source for noted sex differences in formant frequencies before age 10, where there is no sexual dimorphism in VT length. [Work supported by grants NIH-NIDCD R03-DC4362, R01-DC006282, and NIH-NICHHDP30-HK03352.]

3pSC18. Production facilitates lexical acquisition in young children. Tania Zamuner, Paula Fikkert (Ctr. for Lang. Studies, Radboud Univ. Nijmegen, 6500 HD Nijmegen, The Netherlands, tzamuner@psych.ubc.ca), and Bryan Gick (UBC, Vancouver BC, V6T 1Z1, Canada)

Recent research in language acquisition argues that production plays a unique role in the acquisition of phonological and lexical representations [Fikkert. In press, LabPhonX]. In this research, we investigate the effect of production in early lexical acquisition. Twenty Dutch-learning children participated in a word-learning task, where children were taught and asked to imitate non-words. Children were then tested on their recall of the non-words, and their receptive knowledge of the non-words. Children's performance on the two tasks was compared to whether they imitated the non-words during training, the number of times they imitated the non-words, and the accuracy of their imitations. While no effect of production during training was found on children's receptive knowledge of the non-words, production during training was associated with children's ability to recall the non-words. Specifically, children who produced the non-words during training were more likely to recall the non-words at test. No relationship was found between the number of times children produced the non-words during training, or the accuracy of their imitations. These results indicate that production enhances early lexical acquisition, and suggests a unique role for articulation in the formation of phonological and lexical representations.

3pSC19. How acoustic cues in infant-directed speech facilitate 19-month-olds' word recognition: Evidence from time course analysis. Jae Yung Song (Dept. of Cognit. and Linguistic Sci., Brown Univ., Box 1978, Providence, RI 02912, Jae_Yung_Song@brown.edu)

This study investigated how the acoustic characteristics of infant-directed speech (IDS) facilitate infants' recognition of familiar words. The individual roles of three typical characteristics of IDS were examined: Exaggerated pitch range, slow speech rate, and hyperarticulated vowels. Using the intermodal preferential looking procedure, 34 19-month-olds were presented with 12 stimulus sentences (Where is the [target]?); six were spoken in typical IDS style and the other six were digitally altered to remove one of the three IDS characteristics. Infants' looks to target and distracter were coded frame by frame to examine the time course of word recognition. Results showed that infants' ability to recognize words was affected by speech rate. Total looking time to the target was greater ($t(11)=3.337$, $p=0.007$), latency of the first look was shorter ($t(71)=-1.919$, $p=0.059$), and the accuracy score was higher overall when listening to typical IDS as compared to accelerated IDS. In contrast, infants showed no difference in performance when pitch range or the degree of hyperarticulation was modified. This suggests that slow speech rate enhances infants' ability to recognize words more efficiently and reliably during the stage of rapid vocabulary development. [Work supported by NSF BCS-0544127, NIH R01MH60922.]

3pSC20. Investigation of the effect of low frequency components of sound to speech intelligibility. Mokhtar Harun, Mohamad Ngasri Dimon, Siti Zaleha Abdul Hamid, and Hamim Nasoha (Acoust. Res. Lab., Faculty of Elec. Eng., Universiti Teknologi Malaysia Skudai, 81310 Skudai Johor, Malaysia)

Sound attenuation depends on source-to-listener distance, the finishes, and the shape of the room. The amount of sound attenuation at different frequencies influences the overall speech intelligibility in a room. The sound at high frequencies, especially at 1 kHz and 2 kHz, contribute the most for the articulation of consonants, and thus affects speech intelligibility. However, it has been found previously that sound attenuation is lower and harder to achieve at lower frequencies. This would lead us to assume that reverberant sound is mostly of low frequency components. Even if the contribution of sound at low frequency is low for speech intelligibility, this paper attempts to gauge the effect of sound pollution at low frequencies to the overall speech intelligibility in the room. With fixed source-to-listener distances, the measurements were conducted in classrooms with different sizes and finishes.

3pSC21. Children's perception of American English consonants in noise. Kanae Nishi, Dawna E. Lewis, Brenda M. Hoover, Sangsook Choi, and Patricia G. Stelmachowicz (Boys Town National Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Following the seminal study [Miller and Nicely, J. Acoust. Soc. Am. **27**(2), 338–352 (1955)], numerous studies have explored consonant confusion patterns for adult listeners. Although children are known to be less skilled than adults in many aspects of auditory abilities, little is known about how these immature abilities affect consonant perception. In the current study, 72 children (4-, 5-, 6-, 7-, 8-, and 9-year-olds) and 12 adults identified 15 American English consonants in /aCa/ nonsense syllables presented in speech-shaped noise at 0, +5, and +10 dB signal-to-noise ratios (SNRs). Overall percent-correct identification of 15 target consonants improved with age and SNR, but no SNR age interaction was observed. Further analysis of confusion patterns using indexes for transmitted information (TI) for linguistic features (manner, place, and voicing) indicated the following: (1) TI for all features increased with improvement in SNR for all age groups; (2) TI for voicing did not vary with age; (3) TI for place was significantly lower for younger children than adults. These results will be discussed in relation to the development of auditory skills. [Work supported by NIH DC004300.]

3pSC22. The contributions of global spectral and amplitude structure to speech perception by English-speaking adults and children and Mandarin-speaking adults. Susan Nittrouer and Joanna H. Lowenstein (Speech and Hearing Sci., Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, nittrouer.1@osu.edu)

Two studies examined the recognition of signals preserving only the global spectral or amplitude structure of speech by three groups of listeners: Native English-speaking adults, native English-speaking 7-year-olds, and native Mandarin-speaking adults who learned English as a second language. Stimuli consisted of sine wave (SW) replicas and amplitude-modulated (AM) noise bands derived from semantically anomalous four-word sentences (Experiment 1) and semantically rich five-word sentences (Experiment 2). In both experiments, children demonstrated recognition scores comparable to those of the English-speaking adults for the SW sentences, but scores comparable to those of the Mandarin-speaking adults for the AM sentences. In both experiments, native Mandarin-speaking adults were much worse at understanding either kind of stimulus than were the English-speaking adults. Use of linguistic constraints was similar across listener groups, diminishing the possibility that results could be attributed to differences in ''top-down'' effects. Results were used to support three main conclusions: (1) Global spectral and amplitude structure varies across languages; (2) native speakers use this information for speech perception; and (3) in learning a native language, children discover global spectral before global amplitude structure. [Work supported by NIDCD Grant No. DC-00633.]

3pSC23. Identification of consonant-nucleus-consonant words produced by a female child. Morgan Meredith (Dept. of Communicative Disord., UW-Madison, 1975 Willow Dr., Madison, WI 53706, mmeredith@wisc.edu)

The CNC word identification test is part of the speech test battery used to evaluate speech perception by persons fitted with cochlear implants [Luxford et al, Otolaryngology-Head and Neck Surgery, 126–127 (2001)]. This test uses ten lists of 50 relatively common, phonetically balanced English words per list [Peterson and Lehiste, J. Speech. Hear. Disorders, 27:62–70 (1962)]. The version of this test, currently standardized and used across the U.S., contains the words spoken by one male speaker. Research has shown, however, that persons with cochlear implants have more difficulty perceiving speech produced by women and children than speech produced by men [Loizou et al,. J. Acoust. Soc. Am., 103: 1141–1149 (1998)]. The ten CNC word lists were recorded as produced by an eleven-year old girl. Each word was isolated and the word ready spoken by the same girl was added at the beginning of each stimulus sound file. Ten adults with normal hearing identified these words with 98% accuracy. Identification results will be obtained from persons fitted with cochlear implants and compared to the results of the same persons when listening to the original, male voice, CNC word lists. [Work supported by a UW-Madison Hilldale Research Scholarship.]

3pSC24. Why do children attend to global structure in speech signals? Joanna H. Lowenstein and Susan Nittrouer (Speech and Hearing Sci. The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210. lowenstein.6@osu.edu)

Adults can use global structure such as that found in sine wave and vocoded signals for speech perception. Children appear to weight global spectral structure (represented at the syllabic level by formant transitions) even more than adults, prompting the suggestion that children first attend to the relatively slow vocal-tract movements that create global spectral structure rather than to the rapid and discrete gestures more closely associated with individual phonetic segments. But this interpretation hinges on an articulatory account of speech perception, which lacks general support. An alternative explanation is that children attend to formant transitions because they are found in voiced signal portions and so adhere to principles of auditory scene analysis, such as having a common fundamental. This work tested that hypothesis by using stimuli with no fundamental. Adults and children (3, 5, and 7 years) were asked to label fricative-vowel syllables with three kinds of vocalic portions: natural, sine wave, and whispered. Results showed that all listeners attended to formant transitions as much in the sine wave and whispered stimuli as in the natural, leading to rejection of the hypothesis that children prefer formant transitions because they are signal components that adhere to principles of auditory scene analysis. [Work supported by NIDCD Grant No. DC-00633.]

3pSC25. Voicing and devoicing in similar German and English word pairs by native speakers of German. Bruce L. Smith (Dept. of Commun. Sci. and Disord., Univ. of Utah, 1201 Behavioral Sci. Bldg., Salt Lake City, UT 84112), Michael Bruss (Saarland Univ., Saarbruecken, Germany), Rachel Hayes-Harb and Amy Hamilton (Univ. of Utah, Salt Lake City, UT 84112)

A number of languages manifest a pattern of word-final obstruent devoicing, i.e., words ending with underlying voiced obstruents are pronounced so as to be (largely) indistinguishable from words ending with final voiceless obstruents. One question regarding this tendency to devoice voiced obstruent targets is what occurs when native speakers of such a language learn a second language that has a word-final obstruent voicing contrast. For example, do German speakers neutralize voicing contrasts in English as they tend to do in German, or can they learn to produce voiced obstruents in Englis, despite devoicing them in German? To examine this issue, ten native speakers of German produced various phonologically-similar minimal pairs in German and English (e.g., English: Lied/light; German: Leid/leit). Acoustic measurements were made of their productions in both languages, viz., vowel duration preceding final consonants, final consonant duration, duration of voicing during final consonants, and final release burst duration; their English productions were also compared to those of native English speakers. The native German speakers tended to neutralize the voicing distinction to a greater extent when producing German words versus phonologically-similar English words, but they typically did not produce as much of a contrast in English as native English speakers.

3pSC26. Psychometric functions and psychometric characteristics of Mandarin monosyllables. Kuen-Shian Tsai, Sheunn-Tsong Young (Inst. of Biomed. Eng., Natl. Yang-Ming Univ., No. 155, Sec. 2, Linong St., Taipei City 112, Taiwan, young@bme.ym.edu.tw), Li-Hui Tseng, and Cheng-Jung Wu (Natl. Taipei College of Nursing, Taipei City 108, Taiwan)

The psychometric functions of the 700 most frequently occurring Mandarin monosyllables (whose cumulative usage was 98.38% in 1,125 distinct monosyllables) were evaulated. Twenty normal-hearing subjects were asked to hear and to repeat the 700 monosyllables which were randomly presented at the level from 0 to 55 dB HL in 5-dB step. The psychometric functions for each of the 700 monosyllables were fit with third-degree polynomials. The fitted curves were used to calculate five psychometric characteristics: (1) the threshold at 0% correct; (2) the threshold at 50% correct; (3) the instantaneous slope at 50% correct; (4) the linear slope from 20% to 80% correct; and (5) the intelligibility at the highest presentation level. The mean values of such five psychometric characteristics for 700 monosyllables are 1.1 dBHL, 12.1 dBHL, 4.5 %/dB, 4.1% %/dB, and 93.4%, respectively. With the psychometric characteristics, the 700 monosyllables can be used to create word recognition test lists for different purposes, such as constructing test lists with different homogeneity, different familiarity, or different difficulty; and the recognition performance of test lists will be predictable. Futhermore, the research results also indicated that there is no significant correlation between the psychometric characteristics and the occurring frequences of monosyllables.

THURSDAY AFTERNOON, 29 NOVEMBER 2007 GRAND BALLROOM D, 1:00 TO 3:05 P.M.

Session 3pUW

Underwater Acoustics: Target Scattering

David C. Calvo, Chair
Naval Research Laboratory, Acoustics Div., 4555 Overlook Ave., SW, Washington, DC 20375-5350

Chair's Introduction—1:00

Contributed Papers

1:05

3pUW1. Numerical solution of high-frequency acoustic scattering problems by integral equations preconditioned using pseudo-differential impedance operators. David Calvo (Acoust. Div., Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375), and Paul Calamia (Rensselaer Polytechnic Inst., Troy, NY)

Exact computations of high-frequency scattering by targets or rough surfaces can be challenging because of the large number of unknowns. A compromise can be obtained, however, using approximate Dirichlet-to-Neumann surface operators that are known in pseudo-differential form and efficiently implemented using rational approximations. These operators can be used as preconditioners to obtain well-conditioned iterative algorithms to solve the Helmholtz integral equation. Using these algorithms, we compute scattering by two- and three-dimensional targets, and rough surfaces with smooth concave or convex parts with incident fields near grazing angles of interest. This case is particularly important for underwater acoustics, where sounds field graze the free surface or bottom in shallow water.

1:20

3pUW2. At-sea measurements and modeling of acoustic scattering from solid-filled elastic targets on the sea-bed. Mario Zampolli, Alessandra Tesei, and Gaetano Canepa (NATO Undersea Res. Ctr., Vle. S. Bartolomeo 400, 19126 La Spezia, Italy)

Monostatic and bistatic measurements of acoustic scattering from solid-filled fiberglass objects (a sphere and a cylinder with hemispherical endcaps) deployed proud on a sandy seabed and insonified by a rail-mounted parametric source at low frequency (roughly ka=540), are presented. The targets are thin-walled, approximately isotropic random-fiberglass shells, filled partially with an isotropic epoxy resin and partially with sea water. The targets are simple enough in shape to be treated by currently available modeling techniques, but realistic enough to give a first insight into the physics of the elastic waves supported by such material combinations. Preliminary data analysis indicates that the scattering signatures are dominated by the solid filling. The experimentally measured target responses are compared to numerical simulations obtained with a finite-element tool based on the decomposition of the field variables in a series of azimuthal Fourier modes. The experimental data were acquired in October 2006 during the EVA-06 trial off the Island of Elba, Italy.

1:35

3pUW3. Tank measurements and modeling of elastic scattering by resin-filled fiberglass spherical shells. Alessandra Tesei, Mario Zampolli, and Piero Guerrini (NATO Undersea Res. Ctr., Vle. S. Bartolomeo 400, 19126 La Spezia, Italy)

Acoustic elastic scattering measurements were conducted in a tank on 6 cm-radius fiberglass spherical shells filled partially or completely with a low-shear-speed epoxy resin. Preliminary measurements were conducted also on the void shell before filling and on a solid sphere of the same material of the filler, in order to estimate the constituent material parameters via acoustic inversion. The objects were measured in the backscatter direction, suspended at midwater, and insonified by a broadband directional transducer. From the inspection of the response of the solid-filled shell, it was possible to detect and characterize inhomogeneities of the interior (air inter-layers), the presence of which were later confirmed by CT scan and ultrasound measurements. Elastic wave analysis and modeling tools supported the physical interpretation of the measured responses. A spherical shell partially filled with epoxy resin and water was found to exhibit a strong focusing effect in the backscattering direction, for near-perpendicular incidence with respect to the planar internal water-resin interface. Spheres of this kind have the potential of being employed as sonar calibration targets.

1:50

3pUW4. Internal field in an immersed, absorbing fluid sphere excited by a plane acoustic wave. Kenneth G. Foote (Woods Hole Oceanograph. Inst., 98 Water St., Woods Hole, MA 02543)

Modeling of sonar interactions with complicated bodies whose characteristic dimensions are of the order of or greater than the acoustic wavelength generally requires use of numerical methods. For validation of as-

sociated computer code, a particularly useful example is the internal field excited in an immersed, homogeneous, absorbing sphere by an external plane acoustic wave, since this can be expressed analytically. Here, numerical results are given for a fluid sphere with wave number-radius products spanning the range 1–10 and absorption coefficients spanning the range 0–10 dB per wavelength. The density and sound speed contrasts of the fluid sphere relative to the immersion medium are both varied over the range 0.5–2. [Work supported by NOPP through ONR award no. N000140710992.]

2:05

3pUW5. Scattering analysis accelerated by a 3-D multilevel non-uniform grid field evaluation algorithm. Yaniv Brick and Amir Boag (School of Elec. Eng., Tel Aviv Univ., Tel Aviv 69978, Israel, boag@eng.tau.ac.il)

A fast algorithm for computing the scattering cross section of arbitrary shaped large rigid bodies using an iterative method of moments solver has been presented. The main computational bottleneck of such iterative solvers stems from the need to perform at each iteration at least one matrix-vector product. If performed directly, matrix-vector multiplication, which is equivalent to field evaluation for a given source distribution, is characterized by $O(N^2)$ complexity (N being the number of unknowns). To that end, a multilevel non-uniform grid (MLNG) algorithm for 3-D fast field evaluation has been proposed, developed, and tested on representative examples of elongated, quasi-planar, and full 3-D scatterers. The algorithm relies on hierarchical domain decomposition, field evaluation on highly sparse non-uniform grids, and multilevel field aggregation through phase-compensated interpolations. Computational complexity and memory requirements of $O(N \log_{10} N)$ have been achieved by the MLNG without affecting the convergence of the iterative solver. Complexity of the MLNG is similar to that of the multilevel fast multipole algorithm [S. Ko and W. C. Chew, J. Acoust. Soc. Am. **103**, 721–734 (1998)]. The MLNG approach is inherently geometrically adaptive, provides seamless transition from the high frequency to quasi-static regime, and is quite easy to implement.

2:20

3pUW6. The Lloyd-Berry effective wave number in multiple scattering. Dalcio K. Dacol and Gregory J. Orris (Acoust. Div., Naval Res. Lab., Washington, DC 20375-5350)

The scattering of an acoustic field by an ensemble of many discrete objects is a phenomenon observed in many different contexts. In underwater acoustics, one finds examples in propagation through bubbly water, fish schools, non-homogeneous sediments (embedded shells for example), etc. The effective wave number associated with the propagation of the statistically averaged acoustic field scattered by an ensemble of many scatterers is a quantity of physical interest that is also experimentally accessible. Theoretically, this quantity can be evaluated only approximately. Typically, it is evaluated as an expansion in the average density of scatterers. In the literature, one finds two distinct expressions for the second-order contribution, one, the oldest, due to Waterman and Truell, and another by Lloyd and Berry. A new derivation of the Lloyd-Berry formula is provided using an approximation similar to the one used by Waterman and Truell and also by Foldy and Lax. This derivation should shed light on the origin of the different expressions and lend support to the contention that the Lloyd-Berry expression is the correct second-order term in the density expansion.

2:35

3pUW7. Scattering by a sphere illuminated by a Bessel beam: Finite-element modeling. David B. Thiessen and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

The recent analytical solution for the scattering of an acoustic Bessel beam in water by a sphere centered on the beam [P. L. Marston, J. Acoust. Soc. Am. **121**, 753–758 (2007)] gives insight into the processes associated with excitation of target resonances and provides a benchmark for numerical approaches. For example, it is possible to suppress the excitation of certain modes of an elastic sphere by appropriate selection of the beam parameters [P. L. Marston, J. Acoust. Soc. Am. **122**, 247–252 (2007)]. The present work demonstrates that computation of the scattering for steady state illumination by a Bessel beam using the finite-element method (FEM) reproduces the analytical result for the scattering. The dependence of the mode coupling on beam parameters is also present in the FEM results. Some aspects of FEM-based scattering calculations for acoustic beams will be noted. [Work supported by ONR.]

2:50

3pUW8. Enhanced decay rate for the depth dependence of the backscattering of sound by simulated buried targets with evanescent wave illumination. Philip L. Marston, Aubrey L. Espana, Curtis F. Osterhoudt, and David B. Thiessen (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

Sound incident, with a sufficiently small grazing angle on a smooth sediment, produces an acoustic evanescent wave in the sediment, which decays exponentially with increasing depth. A system consisting of an immiscible liquid in contact with a second liquid, has been used in the laboratory to explore the backscattering of sound by a cylinder illuminated by an evanescent wave [C. F. Osterhoudt, Ph.D. Thesis, Washington State University (2007)]. In that work the organ-pipe mode of an open-ended liquid filled metal cylinder was excited. The backscattering was found to decay with increasing depth at a rate approximately twice the spatial decay rate of the incident wave in the simulated sediment. Reciprocity predicts the decay rate is doubled in the case of localized acoustic coupling at the target. Recent supporting measurements of the backscattering, associated with a low-frequency circumferential mode of a plastic cylinder, are noted. Finite-element-based computations of the far-field low-frequency backscattering by a rigid cylinder in the simulated sediment show a decay rate approximately double that of the incident evanescent wave. [Supported by ONR.]

Plenary Session and Awards Ceremony

Gilles A. Daigle, Chair
President, Acoustical Society of America

Business Meeting of the Acoustical Society of America

Motion to approve changes to the Bylaws of the Acoustical Society of America

Presentation of Certificates to New Fellows

William A. Ahroon
Jeffrey E. Boisvert
Elizabeth A. Cohen
Dimitri M. Donskoy
Bruce E. Douglas
William T. Ellison
Ronald L. Eshelman
Steven I. Finette
Sarah Hawkins
Jean-Pierre Hermand
R. Glynn Holt
Anthony P. Lyons
Masayuki Morimoto
Valdimir E. Ostashev
Brandon D. Tinianov
Ronald A. Wagstaff

Presentation of Science Writing Awards

Science Writing Award in Acoustics for Journalists

John Geirland for ''The Sound of Silence,'' WIRED magazine, December 2006

Don Monroe for ''Why the Inner Ear is Snail Shaped,'' Physical Review Focus website, March 2006

Science Writing Award for Professionals in Acoustics

Gary S. Settles, ''High-speed imaging of shock waves, explosions, and gunshots,'' in American Scientist magazine, January/February 2006

Presentation of Acoustical Society Medal Awards

Rossing Prize in Acoustics Education to David T. Blackstock

Pioneers of Underwater Acoustics Medal to William M. Carey

Silver Medal in Engineering Acoustics to Allan J. Zuckerwar

Silver Medal in Speech Communication to Ingo R. Titze

Session 3eMU

Musical Acoustics: Science and Performance

Uwe J. Hansen, Cochair
Indiana State Univ., Physics Dept., Terre Haute, IN 47809

Tyrone M. Porter, Cochair
Boston Univ., Dept. of Aerospace and Mechanical Engineering, Boston, MA 02215

Invited Paper

7:00

3eMU1. Musical acoustics: Science and performance. Uwe J. Hansen (Indiana State Univ., Terre Haute, IN 47809) and Tyrone Porter (Boston Univ., Boston MA 02215)

ASA members, local residents and students are invited to this session featuring a local Jazz combo. The scientific basis of each instrument will be introduced and performance examples will illustrate each instrument family in turn. The evening will conclude with performances of a number of pieces by the combo.

ACOUSTICAL SOCIETY OF AMERICA

PIONEERS OF UNDERWATER ACOUSTICS MEDAL

William M. Carey

2007

The Pioneers of Underwater Acoustics Medal is presented to an individual irrespective of nationality, age, or society affiliation, who has made an outstanding contribution to the science of underwater acoustics, as evidenced by publication of research in professional journals or by other accomplishments in the field. The award was named in honor of five pioneers in the field: H. J. W. Fay, R. A. Fessenden, H. C. Hayes, G. W. Pierce, and P. Langevin.

PREVIOUS RECIPIENTS

Harvey C. Hayes	1959	Fred N. Spiess	1985
Albert B. Wood	1961	Robert J. Urick	1988
J. Warren Horton	1963	Ivan Tolstoy	1990
Frederick V. Hunt	1965	Homer P. Bucker	1993
Harold L. Saxton	1970	William A. Kuperman	1995
Carl Eckart	1973	Darrell R. Jackson	2000
Claude W. Horton, Sr.	1980	Frederick D. Tappert	2002
Arthur O. Williams	1982	Henrik Schmidt	2005

CITATION FOR WILLIAM M. CAREY

. . . for contributions to understanding ocean ambient noise and in defining the limits of acoustic array performance in the ocean

NEW ORLEANS, LOUISIANA • 29 NOVEMBER 2007

The Irish call it the Gift of Blarney and Americans call it the gift of gab, but whatever you call it, we all know that Bill Carey is not afraid to speak his mind, and on a wide variety of subjects. Whether it is Celtic history, the politics of the Middle East and Northern Ireland, or the nonlinear acoustic dispersion properties of sediments, Bill will talk loud and long and knowledgeably and enthusiastically. Those who have the privilege of working more closely with Bill soon realize that there is a wealth of wisdom and experience in his flood of words, and a lot of scientific and engineering originality as well. That Bill's passion, experience, knowledge, and insight first gets expressed verbally is a stylistic thing—what is more important is that Bill's words are usually the prelude to some vigorous action, be it experimental, theoretical, pedagogical, advisory, or editorial. Even at this senior stage of his career, Bill still actively goes to sea, works hands on with electronic and mechanical equipment, develops new mathematical theory, and shows the students how it is done. At the same time, Bill still walks the halls of high level management, drawing upon his many years of experience as a program manager for the Defense Advanced Research Projects Agency (DARPA) and the Naval Undersea Warfare Center (NUWC). No matter what stage of your career you are at, or whether an academic, administrator, or manager, Bill can (and will) engage you with ideas and enthusiasm.

William M. Carey was born in Boston in 1943, but spent most of his youth in Germany. His higher education was at the Catholic University of America, where he received a Bachelor of Science in Mechanical Engineering in 1965, an MS in Physics in 1968, and a Ph.D. in 1974. After his doctoral work, he worked at the Argonne National Laboratory from 1974 to 1979.

Bill's return to the Washington area, at the Naval Research Laboratory (NRL), in 1981 was the beginning of his sustained and productive involvement in underwater acoustics. Over the next three decades, he worked for a number of different Naval laboratories and agencies doing both ocean acoustics research and managerial work.

On the research side, Bill's earliest ocean work concentrated on large ocean-going arrays and ocean ambient noise. In addition to these, shallow water acoustics and more recently ocean vehicles were later added to the list of Bill's specialties. In more recent years, Bill has added the hat of an academic professor to his collection, staying two years at the Massachusetts Institute of Technology from 1996-98 and then moving to Boston University, where he is currently a Professor in the Department of Mechanical and Aerospace Engineering.

In array technology, Bill has contributed heavily to the areas of array calibration, at-sea array measurements, and the understanding of how the ocean and seabed environmental properties determine array performance. Measuring and understanding how the ocean environment determines achievable horizontal array aperture is one of Bill's most significant contributions to underwater acoustics. Another notable contribution to array processing is his work in synthetic aperture arrays, at both low (10-1000 Hz) and medium (1000-10 000 Hz) frequencies.

Having been heavily involved in the study of noise from breaking waves and the low frequency noise that towed (and other) arrays sense in the real ocean environment, Bill Carey rather naturally became a leading expert on ocean ambient noise, and is now summarizing his knowledge in a monograph that will be published in the next year or so.

Bill Carey's work in shallow water acoustics is another one of his major contributions to underwater acoustics. Bill has been a leader in the measurement and understanding of the nonlinear frequency dependence of acoustic attenuation in shallow water and also in using modal wavenumber inverse techniques in shallow water to infer the properties of the ocean's seabed. Bill is currently involved in using autonomous underwater vehicles to address shallow water acoustics questions.

Bill's special research topic, bubbles, has also continued to interest him throughout his career. For several years researchers dismissed the role of entrained bubble clouds because it was widely thought that sub-millimeter bubbles could not influence sound propa-

gation in the low and sub-kilohertz frequency range. Bill was one of the first to argue that, at low frequencies, a bubble cloud can be modeled as a uniform compressible medium with acoustic properties determined by the geometry of the cloud and the effective compressibility and density of the bubbly mixture. Bill went on to study how bubbly assemblages can be hydrodynamically excited to radiate noise, and how bubbles entrained in (and organized by) turbulence can enhance both noise generation and sound scattering from turbulent jets and wakes.

In addition to his technical contributions to underwater acoustics, Bill Carey has also been a community leader for both the Acoustical Society of America (ASA) and the underwater acoustics community in general. Regarding the ASA, Bill was elected Fellow in 1986, has been a member of the Underwater Acoustics and Signal Processing Technical Committees, and serves as an Associate Editor of the *Journal of the Acoustical Society of America*. Bill also has contributed greatly to underwater acoustics via the other technical society that he has been deeply involved in beyond ASA, the IEEE Oceanic Engineering Society of which he was elected Fellow in 1996. He was Editor in Chief of the *Journal of Oceanic Engineering* for seven years, has received its Distinguished Technical Achievement and Service Awards, and is currently a member of its Administrative Committee. Bill is also an active member of the Connecticut Academy of Science and Engineering. In short, Bill is a lively, active leader in the technical societies to which he belongs, and contributes greatly to their success and stature.

To conclude, Bill Carey's contributions to underwater acoustics demonstrate the highest standards of technical excellence, and provide an example of the excitement one can have and the fulfillment one can obtain from a career in science and engineering. We are very happy to see his work and career acknowledged through the receipt of the Pioneers Medal.

JAMES F. LYNCH
ALLAN D. PIERCE
RONALD A. ROY

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Engineering Acoustics

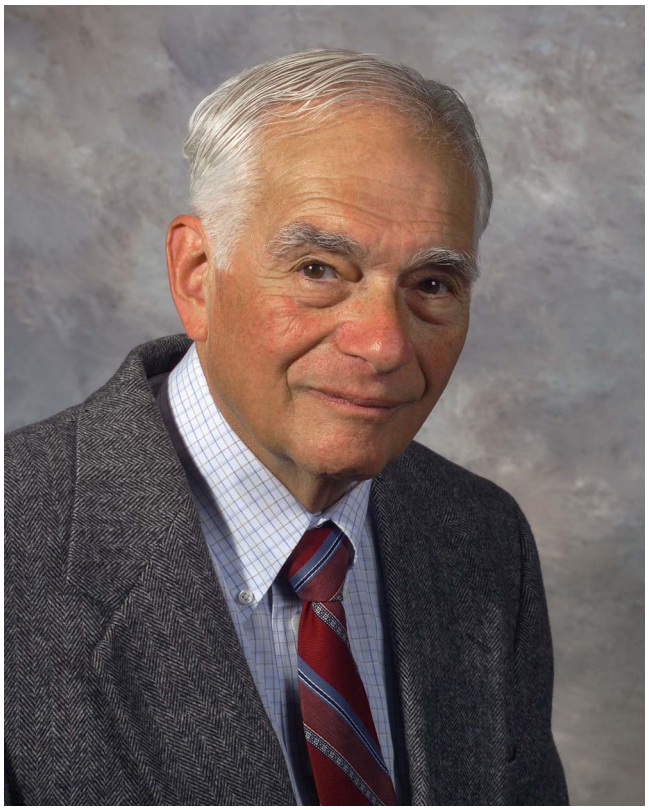
Allan J. Zuckerwar

2007

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

Harry F. Olson	1974	Joshua E. Greenspon	1989
Hugh S. Knowles	1976	Alan Powell	1992
Benjamin Bauer	1978	James E. West	1995
Per Vilhelm Brüel	1982	Richard H. Lyon	1998
Vincent Salmon	1984	Ilene J. Busch-Vishniac	2001
Albert G. Bodine	1986	John V. Bouyoucos	2004



CITATION FOR ALLAN J. ZUCKERWAR

. . . for contributions to the theory and practical development of transducers and their use in fundamental measurements

NEW ORLEANS, LOUISIANA • 29 NOVEMBER 2007

Allan Zuckerwar is a man who is always ready and willing to explain complex technical details, especially those related to transducers, to anyone who asks. He makes those explanations easy to understand because he is a born teacher. He loves figuring out ways to measure physical properties that have never been measured before, especially if others think it cannot be done or because it is tough to do.

Allan was born in 1934 in New York City. He received his undergraduate degree in 1956 and a MS in 1957 in electrical engineering both from Carnegie Institute of Technology. He obtained a Doctor of Science degree in physics in 1966 from Technische Hochschule Stuttgart (now the University of Stuttgart) under Professor Hans O. Kneser. Allan then did three years of postdoctoral research in crystallography under Professor Arthur S. Nowick at Columbia University. He was the Assistant Professor of electrical engineering at Youngstown State University in Ohio for three years before moving to Old Dominion University in Norfolk, Virginia in 1973 as Research Associate Professor of physics. In 1980 he joined NASA Langley Research Center in Hampton, Virginia as a research scientist. He retired in 2005 but does research part time now.

Allan's career in acoustics began in the early 1960s when he was a graduate student in Stuttgart working with Kneser. This introduction to the absorption of sound in gases would later have some practical application. When Allan was employed in the Instrument Research Division at NASA Langley Research Center he had the opportunity to revisit this problem. At that time a new standard was being developed, S1.26-1978, Method for the Absorption of Sound by the Atmosphere. One difficulty in the development of the standard was anomalous low-frequency absorption, which could not be explained by existing theory. Lou Sutherland and Joe Piercy independently postulated that the anomalous absorption was due to the hitherto neglected role of nitrogen. Preliminary results were encouraging, but the available data on nitrogen were too scattered to permit the evaluation of absorption parameters within reasonable limits of uncertainty. The main difficulty was that the peak absorption in nitrogen occurs at very low frequencies, requiring a resonance tube of impractical length (1700 m).

One day Allan noticed some shock tube sections in a storage yard at Langley Research Center. The specifications were suited for measuring the pure-nitrogen absorption peak. Four tube sections, having a total length of 17 m operating at 100 atmospheres proved ideal for the measurement. After receiving permission to use the tubes, and enlisting two graduate students, Allan and his group were able to measure the relaxation peak in nitrogen over wide ranges of temperature, pressure, and humidity. The development of this instrument extended the state-of-the-art by a factor of 40. The results were later incorporated into the 1995 revision of the above-mentioned American National Standards Institute (ANSI) standard.

While researching sound speed data in fluids for an article in the *Encyclopedia of Acoustics* (edited by Malcolm Crocker), Allan noted that there was a gap between prediction methods and experimental results. At that time the spherical resonator, as developed by Moldover and his group at the National Institute for Standards and Technology, proved to be an instrument of unprecedented accuracy, a measurement uncertainty of a few parts per million being attainable. Allan noticed that the applicable physical properties of a fluid (specific heat, virial coefficients, and relaxation parameters) were sometimes not adequately taken into account, especially over wide ranges of temperature and pressure. Allan was able to find the needed data for all the necessary gas properties of 52 gases, which ultimately served as the basis for the three volume *Handbook of the Speed of Sound in Real Gases* published in 2002. This is an outstanding contribution to the science of acoustics and a seven year labor of love.

A problem in the design of membrane-based microphones is to construct a membrane-backplate combination that will provide near-critical damping. Assigned the task of developing a high-temperature microphone for the National Aero-Space Plane (NASP), Allan developed a theory that takes into account the holes and slots in the backplate and

the role of the backchamber. In collaboration with the late Frank Cuomo, University of Rhode Island, the two succeeded in developing a microphone that operated at 538°C (1000°F). Subsequently they were in the preliminary stages of developing a 1093°C (2000°F) microphone, but the NASP program terminated before the work could be completed. Undoubtedly, Allan has made significant contributions to the theoretical design of condenser microphones particularly at high temperatures.

Allan has contributed to four books: *Encyclopedia of Physical Science and Technology* (1987), *Encyclopedia of Applied Physics* (1994), *Handbook of Condenser Microphones* (1995), and *Encyclopedia of Acoustics* (1997). He has more than 50 refereed technical publications, 19 patents, and is the recipient of several prestigious awards from NASA Langley. Allan served as a member of three ANSI standards working groups on atmospheric sound absorption, ground impedance, and microphone calibration, and was the chairman of standards working group S1/WG26. He was elected a Fellow of the Acoustical Society of America in 1985 and in 2002 was appointed the Associate Editor for Transduction, Acoustical Measurements, Instrumentation, and Applied Acoustics of the *Journal of the Acoustical Society of America*. For his continuous contributions to the theory and development of transducers and their use in fundamental measurements, his active participation in the development of acoustical standards, and his dedication to improvements of publications in acoustics, Allan Zuckerwar is undoubtedly a highly deserving recipient of the Acoustical Society of America Silver Medal in Engineering Acoustics.

FERI FERASSAT

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal

in

Speech Communication

Ingo R. Titze

2007

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

Franklin S. Cooper	1975	Arthur S. House	1991
Gunnar Fant	1980	Peter Ladefoged	1994
Kenneth N. Stevens	1983	Patricia K. Kuhl	1997
Dennis H. Klatt	1987	Katherine S. Harris	2005

CITATION FOR INGO R. TITZE

. . . for contributions to fundamental understanding of the physics and biomechanics of vocal fold vibration and for interdisciplinary work in voice studies

NEW ORLEANS, LOUISIANA • 29 NOVEMBER 2007

Artists and scientists have studied voice and singing for as long as there have been artists and scientists, but voice science did not really exist as a coherent discipline before Ingo Titze's work established a strong theoretical and empirical basis for studies of how voice is produced. His research has addressed virtually every aspect of laryngeal biomechanics, using a wide variety of methods drawn from many disciplines. Through his systematic and comprehensive efforts to understand how and why the vocal folds vibrate, he has helped establish the interdisciplinary field of voice studies, the blossoming of which is now so apparent in the pages of the Journal of this Society.

Ingo's scholarly achievements are truly impressive, in quality, quantity, and in the systematicity with which he has pursued an understanding of how sound is produced by vocal fold vibrations. His multi-mass, finite element, and continuum models of vocal fold vibration established standards for rigor, precision, and accuracy in a research area that had previously been characterized largely by observation and description. The impact of these papers has been such that they alone would justify this award.

However, Ingo's research goes far beyond mathematical modeling. His work formally examines virtually every aspect of the voice production process, including the role of tissue properties, airflow, and neuromuscular function, along with the manner in which each factor affects the sound produced. He has published important work describing the physiology of the vocal folds, and has examined the properties of the different types of tissue in the folds. He has studied the properties of the vocal ligament and laryngeal muscles, and has examined the function of laryngeal joints. Ingo has applied physiological models (*in vivo*, excised laryngeal, and hemilaryngeal) to validate and extend his mathematical models. He has pioneered the application of methods from nonlinear dynamics to the study of voice. He has used speech synthesis to study the perceptual importance of different laryngeal configurations, particularly in singing, and has given the world *Pavarobotti*, the singing robot. He has published a number of helpful methodological studies, evaluating and attempting to standardize recording and analysis techniques for clinical and scientific applications. In every case, Ingo has applied these varied methods to study a full range of vocal function, from pathological phonation to normal voice to the special case of trained singing voices.

The impact of this varied range of investigations extends beyond the core voice science community. Ingo's CV lists over 160 peer-reviewed scholarly publications in journals ranging across an array of disciplines, such as acoustics, biology, biomechanics, and biomechanical engineering, medicine, phonetics, and in journals devoted to the science and art of the singing voice, thanks to which a shared body of knowledge about voice now exists across disciplines. I know of no other scholar who has made such an effort to disseminate his findings so consistently to such a wide audience. Ingo also organized the international Vocal Fold Physiology conferences and the International Conference on Voice Physiology and Biomechanics series, at which researchers from many disciplines meet and share results. Without these efforts to integrate voice research across disciplines, voice science could not develop as a coherent discipline.

The importance of Ingo's work is further enhanced by his consistent attention to the practical implications of his basic research. Many of his published papers extend theoretical or experimental results regarding the physics and biomechanics of vocal fold vibration to clinical practice and voice training. For example, he has studied the viscoelastic properties of various biomaterials, to the end of improving implant techniques for clinical voice rehabilitation. Using experimentally-derived data on impact stresses, dehydration, and other causes of physical damage to the vocal cords, he has studied the prevalence and causes of occupational voice hazards, and has investigated many methods for reducing voice problems in professional voice users. He has published a variety of papers examining and comparing different experimental techniques (voice profile analysis, acoustic analyses, recording techniques), to the end of standardizing clinical and experimental practice.

As if that were not enough, Ingo has trained a generation of distinguished voice scientists, and has influenced many more scholars who have visited the University of Iowa or the Denver Center for the Performing Arts. Even students he has not taught personally feel his influence through his books *Principles of Voice Production* and *The Myoelastic Aerodynamic Theory of Phonation*, the first of which is the text of choice for introducing voice science to speech-language pathologists, engineers, linguists, vocalists, MDs, and physiologists.

Ingo received his PhD in Physics from Brigham Young University in 1972. He has spent most of his teaching career at the University of Iowa, where he founded the National Center for Voice and Speech (NCVS). He continues to be Executive Director of the NCVS which now has its headquarters at the Denver Center for the Performing Arts. He has been a Fellow of the ASA since 1983, and is active in the music, animal bioacoustics, and speech communication technical committees in the best interdisciplinary tradition of the Society.

Through his many groundbreaking contributions to basic research, his extensions of research findings to applications in a variety of fields, and his dedication to education, training, and outreach, Ingo Titze exemplifies the best traditions of scholarship. His concern for the application of research results and his contributions as a teacher have led to the advancement of clinical practice, biomedical engineering, and human welfare. He has mentored, influenced, or inspired everyone who works in the area of voice. He truly deserves this honor, and I know that his many colleagues join me in congratulating him.

JODY KREIMAN

Session 4aAA

Architectural Acoustics: Computational Room Acoustics and Room Acoustics Measurements

David S. Woolworth, Chair
Oxford Acoustics Inc., 356 CR 102, Oxford, MS 38655

Contributed Papers

8:30

4aAA1. *In situ* estimation of acoustic impedance on the surfaces of realistic interiors: An inverse approach. Gabriel Pablo Nava, Yoichi Sato, Shinichi Sakamoto (Inst. of Industrial Sci., The Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan), and Yosuke Yasuda (The Univ. of Tokyo, Kashiwashi, Chiba 277-8563, Japan)

In situ measurement of acoustic impedance is traditionally performed using pairs of microphones located close to the test surface. However, this method becomes troublesome if inaccessible complex-shaped surfaces, such as those in a real room, are considered. To overcome this problem, a method to estimate the normal acoustic impedance on the interior surfaces of a room is proposed. As input data, the algorithm takes: (1) The 3-D shape of the room; (2) the strength of the sound source; and (3) a set of sound pressures measured at random locations in the interior sound field. The estimation of the acoustic impedance at each surface is achieved via the solution of an inverse problem, which arises from the boundary-element method applied to the discretized interior boundaries of the room. Unfortunately, the solutions of this kind of problems are known to be unstable and sensitive to noise, due to a rank-deficient linear system. Dealing with such a system is avoided in the proposed method by formulating an iterative optimization approach, which is shown to be more robust to noise. Previous work has reported examples with numerical simulations. This paper goes further and presents results obtained with real data from experiments.

8:45

4aAA2. Acoustical renovation of rooms for music instruction in schools to improve teaching and listening conditions. Hyun Paek (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607) and Gary W. Siebein (Univ. of Florida, Gainesville, FL 32611-5702)

Several music classrooms were renovated because music instructors complained of excessive noise levels during classes. Instructors also could not hear students well and commented on harsh reflected sounds. The rooms were trapezoidal shaped with relatively low, angled gypsum board ceilings with few sound absorbing materials installed. Acoustical measurements of reverberation times, overall sound levels, and reflected sounds were used to identify the causes of the perceived problems. Renovation schemes included floating planes of sound-diffusing panels or prefabricated-diffusing panels suspended in a ceiling grid to allow students to hear each other; raising the ceiling height to reduce excessive sound levels; and strategically placing sound-absorbent materials to reduce reverberation and acoustic defects. Post-construction acoustical measurements, as well as discussions with music instructors, revealed that while both reverberation times and loudness levels decreased, the early sound energy reflected across the room increased and resulted in improved musical communication among students and instructors.

9:00

4aAA3. Perception and performance of hearing-impaired versus normal-hearing persons under noise. Lauren M. Ronsse and Lily M. Wang (Architectural Engr. Prog., Univ. of Nebraska Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, lronsse@mail.unomaha.edu)

Examining the effects of mechanical system noise on worker performance and perception for normal-hearing persons has been an area of previous research; this study extends this research area to include hearing-impaired subjects. The goal is to determine if persons with hearing impairments respond similarly to seven different mechanical system background noise conditions as persons with normal hearing. The seven noise signals vary in terms of level and spectral quality, but are all within the range of background noise conditions found in commercial offices. Performance is gauged on three types of tests (math, verbal and typing), while subjective perception is measured via a subjective questionnaire. The results of this investigation will show if significant differences are present between the two groups, and if so, they may lead to the development of different standards for noise criteria levels in spaces designed for the hearing-impaired. [Work supported by a Univ. of Nebraska Layman Award and an ASHRAE Graduate Student Grant-in-Aid.]

9:15

4aAA4. Residential wall construction transmission losses and home office productivity. Alicia J. Wagner and Lily M. Wang (Architectural Engr. Prog., Univ. of Nebraska Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, ajwagner@mail.unomaha.edu)

With advances in technology, the office worker finds herself in a changing work environment; there are now more opportunities to work from home. While recent research has addressed variations in worker productivity due to existing sound conditions within a space, home office performance has a greater dependence on isolation from exterior sound sources than conventional office work. This project investigates how sound levels produced by walls with different levels of transmission loss (TL) affect an individual's performance. Performance on math, verbal, and typing tests was assessed for sound signals played across four residential wall constructions ranging from a sound transmission class of 35 to 55. Correlations between the TLs and math, verbal, and typing productivity will be shown and used to determine if there is a minimum acceptable TL level.

9:30

4aAA5. Evaluation of accoustic comfort in classrooms of the Federal University of Parana. Andressa M. C. Ferreira and Paulo H. T. Zannin (Departamento de Engenharia Mecanica, Universidade Federal Do Parana, arqferreira@yahoo.com.br)

This research is a case study developed to evaluate and compare the classroom acoustics of two different buildings from the Federal University of Paran, built in different times: Building CP (from the 1960s) and building JB (built in 2000). Acoustical parameters have been measured in eight classrooms of the CP building, and seven classrooms of the JB building: Sound pressure level, reverberation time, and sound insulation. The methodology employed followed standard procedures for acoustical evaluation

of classrooms. Results showed that the classrooms of the CP building, from the 1960s, display satisfactory acoustical conditions, especially in their reverberation times. On the contrary, classrooms of the more recently built JB building yielded poor results with respect to reverberation times, either in empty classrooms, or in partially or even totally occupied classrooms, not meeting the values set by the standards. Thus, despite the evolution of the studies and research on room acoustics, older classrooms of the Federal University of Paran have better acoustics than more recently built classrooms.

10:00–10:30 Break

9:45

4aAA6. Use of 2-D and 3-D visualization and auralization tools to assess the acoustics of unusual shaped spaces. Joshua M. Cushner and Ryan B. Biziorek (Arup Acoust., 155 Ave. of the Americas, New York, NY 10013, joshua.cushner@arup.com)

The assessment of room acoustics commonly includes the calculation of relevant metrics by standard empirical equations or computational ray-tracing. These methods are not always sufficient for unusual shaped rooms, and can be supplemented with visualization and auralization techniques that improve our understanding of the behavior of sound in these spaces. Utilizing two- and three-dimensional room models with proprietary developed software, visualization of 360-deg particle emissions from a point source can provide useful information about sound propagation, spatial characteristics, and interactions with room boundaries. Following an iterative visualization process, design verification can be established using auralization techniques in a controlled listening environment. Such techniques facilitate better understanding of the acoustic phenomenon within the room and can help to refine acoustic solutions for room shapes or more efficient use of room surfaces for sound-absorbing finishes. Results from recent project studies show this approach to be useful to enhance communication with the architect and provides the acoustician with another useful analysis tool.

10:30

4aAA7. An exploration of virtual 3-D sound. Scott McDermott and Cheehung Henry Chu (Ctr. for Adv. Comput. Studies, The Univ. of Louisiana at Lafayette, Lafayette, LA 70503, sdm1718@cacs.louisiana.edu)

Many of the applications, virtual environments, and video games available to average computer users embrace stunning 3-D graphics and real-world visualizations. Developers spend an extraordinary amount of time and effort creating these immersive, realistic virtual environments, primarily focusing on the graphics components. Within these virtual realities, the user should be able to easily and accurately perceive the locations of sound sources, as well as the acoustic nature of the environment. However, for reasons of economy and simplicity, most developers apply readily available industry standards for generating pseudo-3-D sounds in their applications. This research explores the shortcomings of these standards and proposes an effective alternative. This project includes a number of computationally efficient, physics-based 3-D acoustics simulations, each of which will produce realistic aural reproductions. The goal is to evaluate and compare these algorithms against each other, non-3-D sound reproduction, and the current industry standards (e.g. Microsoft's DirectX's pseudo-3-D algorithm). Two hypotheses will be tested. First, users will find true, physics-based 3-D algorithms to render improved auralization reproductions compared against DirectX and/or OpenAL. Second, localization and spatialization improve with user training when using these algorithms.

10:45

4aAA8. An automated high spatial-resolution scanning system for room-acoustic scale models. Rolando de la Cruz, Gordon Rubin, and Ning Xiang (Grad. Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

Experimental study of acoustically complex environments is a lengthy and troublesome process, in part due to the great deal of effort in acquiring data. A custom Cartesian scanning system has been developed to perform automated room impulse response measurements in scale models. High spatial-resolution measurements are performed within a planar grid and mapped to observe the spatial distribution of the sound field. Phase relationship across the scanning grid is maintained between measurements using continuous pseudo-random signals. The efficiency and repeatability of the high spatial-resolution scanning system is demonstrated in acquiring a large number of room impulse responses over a predefined Cartesian grid, while maintaining a wide bandwidth in accordance with the spatial Nyquist theorem. This paper discusses some preliminary results obtained in an eighth scale model of a coupled-volume system.

11:00

4aAA9. Prediction of sound around a corner in long corridors. William W. S. Fung, Sai Tung So, Chun Wah Leung (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong), and Kai Ming Li (Purdue Univ., West Lafayette, IN 47907-2031, mmkmli@purdue.edu)

An L-shape section is a prime consideration in the design of corridors in modern buildings. Similar to a straight corridor, the direct sound field dominates the sight-line region in the near field. However, there is no contribution from the direct sound field when the receiver locates around the corner at the shadow zone. The reflected sound field becomes more important as the receiver moves further from the bend. The present study examines the propagation of sound along a long space with an L-shape section. A numerical model using an image-source method is developed for the prediction of sound transmission along the L-shape section of a long corridor. The reverberation times are predicted by the proposed numerical model. In particular, the respective sound fields at two octave bands (500 Hz and 2,000 Hz) are obtained. The noise levels at these two octave bands are needed for the computation of the rapid speech transmission index. This information is often used to conveniently assess the speech intelligibility of a communications channel. The proposed image source model is validated by comparing with field measurements conducted in an L-shape corridor and also conducted in a scaled model in an anechoic chamber.

11:15

4aAA10. Laser Doppler vibrometer-based acoustical measurement of reflection coefficient up to 20 kHz. Gordon Rubin and Ning Xiang (Grad. Prog. in Architectural Acoust., Rensselaer Polytech. Inst., Troy, NY 12180)

The well-established two-microphone technique for determining the acoustic reflection of a material in an impedance tube has physical limitations that hinder the measurement of this quantity at high frequencies. Tube geometry, microphone spacing, and perturbation of the sound field by measurement instruments make the determination of normal incidence reflection coefficients above 10 kHz difficult, if not impossible, using microphones and the traditional impedance tube methods. In a recent work, an extended impedance tube technique utilizing a scanning laser doppler vibrometer (SLDV) was presented [Vanlanduit et al., J. Acoust. Soc. Am. **3** (2005)]. In this study, a method using a single point laser Doppler vibrometer (LDV) is used to study the limitations and merits of acousto–optic sensing in impedance tubes. Foams of various thicknesses are measured using the transfer function method in a transparent impedance tube. Results are presented that convey the potential usage of these acousto–optic techniques in the measurement of acoustic reflection properties of materials up to 20 kHz and beyond.

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Biological Effects and Medical Applications of Stable Cavitation I

R. Glynn Holt, Chair
Boston Univ., Dept. of Aerospace and Mechanical Engineering, Boston, MA 02215

Chair's Introduction—8:35

Invited Papers

8:40

4aBB1. Stable cavitation description and characterization. Charles C. Church (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

In his now-famous work, ''Physics of Acoustic Cavitation in Liquids,'' [Physical Acoustics, Vol. 1-B, edited by W.P. Mason (Academic Press, New York, 1964), Chap. 9, pp. 57–172], Hugh Flynn coined two terms that have become commonplace in the lexicon of bubbologists, ''transient cavitation'' and ''stable cavitation.'' Although in routine use in the field of acoustic cavitation research, the original, rather precise meanings of these phrases have since been obscured. For example, ''inertial cavitation'' is now preferred owing to the perception that the former, ''transient cavitation,'' refers to motion that must inevitably result in the destruction of the bubble upon collapse, while in fact the original criterion was simply that the contraction must approximate that of a Rayleigh cavity until arrested by the internal pressure. Similarly, ''stable cavitation'' is often considered to be low-amplitude motion, but Flynn's criterion was simply that the bubble oscillate nonlinearly about its equilibrium radius. This talk will describe the original and modern meanings of these terms, discuss various cavitational phenomena, and differentiate the effects likely to be induced by each type of bubble motion. [Work supported by DAMD17-02-2-0014.]

9:00

4aBB2. Cavitation microstreaming patterns in single and multiple bubble systems. Andrew Ooi, Paul Tho (The Univ. of Melbourne, Parkville, Victoria 3010, Australia), and Richard Manasseh (CSIRO Manufacturing and Infrastructure Technol., Highett, Victoria 3190, Australia)

Cavitation microstreaming is a well-known phenomenon, yet there are very few flow visualizations or measurements of the velocity fields. In this talk, results from micro-particle image velocimetry measurements and streak photography, illustrating the flow field around a single and multiple oscillating bubbles resting on a solid boundary, will be presented. The different modes of bubble oscillations were also measured in terms of the variation in the radius of the bubble and the movement of the bubbles centroid so that the streaming flow field could be accurately related to the bubble motion. Thus, the resulting flow field can be correlated with the different vibration (volumetric, translating, and orbiting) modes. The flow field resulting from these oscillation modes contains closed streamlines, representing vortical regions in the vicinity of the bubble. Despite some inconsistencies, there is general agreement between these streaming patterns and those found in existing theoretical models. In addition, shape mode oscillations of single bubbles, as well as several different cases of multiple bubbles simultaneously oscillating with the same frequency and phase, were also investigated and show a rich variety of streaming patterns.

9:20

4aBB3. Subharmonic generation mechanisms in stable cavitation. R. Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215) and Joachim Holzfuss (Technische Universität Darmstadt, 64289 Darmstadt, Germany)

The observation of subharmonic emissions during an ultrasound experiment has often been taken as the sine qua non of stable cavitation. While the appearance of subharmonics is very likely the indication of the presence of bubbles, it leads to a thorny question: What do subharmonic emissions tell us about the nature of the cavitation in the experiment? The answer to that question will naturally be of pragmatic importance to the experimenter seeking to enhance, inhibit, or otherwise control the resulting bioeffect. In this talk, we review the physical mechanisms for bubble generation of subharmonics, giving where possible, examples of each mechanism and eliciting the consequences for biomedical experiments.

9:40

4aBB4. Ultrasound contrast agent microbubbles: Stable and transient subharmonic emissions. Leonardo Masotti, Elena Biagi, Luca Breschi, and Enrico Vannacci (Ultrasonic and Non Destructive Testing Lab., Electrons and Telecommunications Dept. via S Marta 3, 50139 Firense, Italy)

The paper describes a method able to perform an ultrasonic study of a single UCA bubble immersed in a tank filled with a liquid (water) by means of ultrasonic equipment able to insonify and acquire RF echo signals coming from an isolated bubble with a high frame rate. The work was focused on the study of subharmonic emission from an isolated bubble of SonoVue. In particular, the

acoustic pressure threshold for a subharmonic stable emission was evaluated for a subset of 50 microbubbles at 3.3 MHz and at 5 MHz of insonation frequencies. An unexpected second pressure threshold, which caused the standstill of the subharmonic emission, was detected at 3.3 MHz and 5 MHz. A transient subharmonic emission, which is hypothesized as being related to the formation of new free gas bubbles, was detected during the ultrasound-induced destruction of microbubbles. Several sequences of RF echo signals and the related spectra, acquired from a set of 400 isolated microbubbles, is investigated at different acoustic driving pressures and excitation frequencies. Specific bubble behaviors concerning the stable subharmonic emission, the inhibition of subharmonic generation, and the transient subharmonic emission were detected during experimentation and discussed in this paper.

10:00–10:15 Break

Contributed Papers

10:15

4aBB5. Monitoring and simulating stable cavitation during ultrasound-enhanced thrombolysis. Saurabh Datta, Azzdine Y. Ammi (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, MSB 6152, Cincinnati, OH 45267), Constantin C. Coussios (Univ. of Oxford, Oxford, UK), and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH 45267-0586)

The presence of stable cavitation has been shown to be highly correlated with thrombolytic efficacy for recombinant tissue plasminogen activator (rt-PA) mediated thrombolysis. A commercial contrast agent, Definity, was used with 120 kHz pulsed ultrasound to nucleate, promote, and sustain stable cavitation. The effect of stable cavitation on increased penetration of rt-PA into the clots is discussed. Also, the possibility of lowering the rt-PA dose using sustained stable cavitation adjuvant to thrombolytics is presented. To understand the bubble dynamics involved, the bubble response was studied using the Keller-Miksis model and stable and inertial cavitation thresholds were studied as a function of bubble radius. The largest mass loss (26.2%) was observed for clots treated with 120 kHz ultrasound (0.32 MPa peak-to-peak pressure amplitude, 80% duty cycle), rt-PA (96 μg/ml) and stable cavitation nucleated by Definity. A comparable mass loss of 22% was observed at a much lower concentration of 11 μg/ml in the presence of stable cavitation. In addition, the simulation results provide insight into the nuclei sizes relevant for this therapeutic application and possible mechanisms involved.

10:30

4aBB6. 120 kHz pulsed ultrasound enhanced thrombolysis with tissue plasminogen activator-loaded echogenic liposomes. Jason M. Meunier (Dept. of Emergency Medicine, The Univ. of Cincinnati, 231 Albert Sabin Way, PO Box 670769, Cincinnati, OH 45267-0769, meuniejn@uc.edu), Denise A. B. Smith, Christy K. Holland (The Univ. of Cincinnati, Cincinnati, OH 45221-0761), Shao-Ling Huang, David D. McPherson (The Univ. of Texas Health Sci. Ctr., Houston, TX 77030), and George J. Shaw (The Univ. of Cincinnati, Cincinnati, OH 45267-0769)

Echogenic liposomes (ELIP), phospholipid vesicles filled with gas and fluid, can be manufactured to incorporate the thrombolytic drug tissue plasminogen activator (tPA). Real-time thrombolysis of blood clots exposed to tPA-incorporating ELIP (t-ELIP) was monitored using video microscopy with an inverted optical microscope. Human whole blood clots on silk sutures were exposed to tPA alone (3.15 micrograms/ml), t-ELIP alone (3.15 micrograms/ml), t-ELIP and 120 kHz ultrasound (0.18 MPa peak negative pressure, 1.667 kHz pulse repetition frequency, 50% duty cycle), or tPA and ultrasound, for 30 min. The extent of thrombolysis was determined by assessing clot width as a function of time, using a time-lapse microscopic imaging technique. The average percent change in clot width at 30 min for clots treated with t-ELIP alone exceeded tPA alone (22.8% vs. 15.6%, respectively). Thrombolytic efficacy was similar for either tPA or t-ELIP exposed to 120 kHz ultrasound. Thus, the thrombolytic drug could be effectively released by exposure to 120 kHz ultrasound. [This work was supported by The Distinguished Chair for Clinical Research in Emergency Medicine Foundation Award, K02-NS056253, NIH 1RO1 NS047603, and NIH 1R01 HL074002.]

10:45

4aBB7. Effect of 810 kHz cw ultrasound on bacterial biofilms. Kofi Anguah, Roby Velez, Amy C. Vollmer, and E. Carr Everbach (Eng. and Biol. Dept., Swarthmore College, Swarthmore, PA 19081, ceverba1@swarthmore.edu)

Biofilms of florescent E. coli bacteria were grown in protein-rich media on coverslips that served as the floor of an ultrasound exposure chamber. Two opposite walls of the chamber were strips of lead zirconate titanate (PZT-4) driven at their resonance frequency of 810 kHz in continuous wave mode, thereby setting up acoustic standing waves within the 12 mm by 16 mm by 0.5 mm chamber. A convolution confocal florescence microscope was used to visualize the biofilm before, during, and after 10-minute exposures at various acoustic pressure amplitudes, with and without Optison microbubbles present. Quantifiable changes to the biofilm structure suggest that stable cavitation is a mechanism of interaction that may provide a method of enhancing antibiotic action in medical applications. [Work supported by HHMI grant #52005202.]

11:00

4aBB8. Simulation of diagnostic ultrasound image pulse sequences in cavitation bioeffects research. Douglas L. Miller, Chunyan Dou (Dept. of Radiol., Univ. of Michigan, Ann Arbor, MI 48109), and Roger C. Wiggins (Univ. of Michigan, Ann Arbor, MI 48109)

Research on cavitational bioeffects of diagnostic ultrasound typically involves a diagnostic scanner as the exposure source. However, this can limit the ranges of exposure parameters for experimentation. A single-element laboratory exposure system was used to simulate 1.5 MHz contrast-aided diagnostic ultrasound of rat kidney. Amplitude modulation of pulses (1.7 microsecond duration, 0.43 ms repetition period and 2.3 MPa amplitude) with a Gaussian envelope of 4.6 ms duration reproduced the image pulse sequence (IPS) of 1 s intermittent beam scanning. A 10 microliter/kg/min IV dose of Definity contrast agent was given during 1–5 min exposures. Glomerular capillary hemorrhage was assessed by histology. A fixed beam induced hemorrhage within the exposed area comparable to the diagnostic imager, but the use of 5 exposures at closely spaced positions more faithfully reproduced the effect produced within a scan plane. IPSs consisting of only one or two pulses approximated the effect of the simulated IPS. Use of a 100 ms duration envelope confirmed a previous observation that slow Doppler imaging mitigated the bioeffect. Thus, relatively simple laboratory exposure systems can simulate diagnostic ultrasound scanning in cavitation bioeffects research, allowing exploration of parameter ranges beyond those available on present clinical systems. [Work supported by NIH grant EB00338.]

11:15

4aBB9. Growth of gas bubbles by rectified diffusion at high supersaturation levels. Yurii A. Ilinskii, Preston S. Wilson, and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

There is continued interest in possible effects of under-water sound on the growth of gas bubbles in capillaries of marine mammals and humans. [Houser et al., J. Theor. Biol. **213**, 183 (2001)] showed that following a series of dives, gas supersaturation in capillaries of marine mammals can reach 300%. For such high supersaturation levels, standard mathematical

models of both static and rectified diffusion underestimate the rate of bubble growth by 10%–20%. The discrepancy is demonstrated by comparing predictions based on existing mathematical models with direct numerical solutions of the differential equations for gas diffusion in the liquid and heat conditions in the bubble. The Rayleigh-Plesset equation is used to describe the bubble dynamics. Underestimation of bubble growth by existing mathematical models is due to the underlying assumption that the gas concentration in the liquid is given by its equilibrium state for a bubble of constant radius. This assumption is violated when high supersaturation causes the bubble to grow too fast in relation to the time scale associated with diffusion. Rapid bubble growth results in an increased concentration gradient at the bubble wall, and therefore a growth rate in excess of predictions based on quasistatic gas concentrations. [Work supported by ONR.]

FRIDAY MORNING, 30 NOVEMBER 2007 NAPOLEON B1, 9:00 TO 10:00 A.M.

Session 4aEA

Engineering Acoustics: Topics in Engineering Acoustics

Daniel M. Warren, Chair
Knowles Electronics, 1151 Maplewood Drive, Itasca, IL 60134

Contributed Papers

9:00

4aEA1. Effect of leakage in Helmholtz resonators. Iljae Lee, Ahmet Selamet, and H. S. Kim (The Ohio State Univ., 930 Kinnear Rd., Columbus, OH 43212, Selamet.1@osu.edu), and Norman T. Huff (Owens Corning, Novi, MI 48377)

The effect of leakage in Helmholtz resonators has been experimentally and numerically investigated. Transmission loss of a Helmholtz resonator having a gap between the cavity and main duct is measured using an impedance tube setup. The effect of leakage on the transmission loss is examined using different amounts of gap openings. Experimental results are then compared with the predictions from the boundary-element method. The study shows that the leakage increases the resonance frequency substantially and widens the transmission loss. Hence, the leakage needs to be taken into account for accurate predictions of Helmholtz resonators.

9:15

4aEA2. Optimization of baffle configurations to prevent aeroacoustic instabilities in heat exchangers: Preliminary experiments. A. Miguel Moreira (College of Technol., Setubal Polytechnic Inst., Campus do Instituto Politcnico de Setubal, Estefanilha 2910-76, Portugal, mmoreira@est.ips.pt), Jose Antunes, Vincent Debut, Martins Paulino (Inst. of Nuclear Technol., Portugal), and Heitor Pina (Tech. Inst., 1049-001 Lisboa, Portugal)

Gas heat exchangers are prone to aeroacoustic instabilities, which often lead to severe noise levels, structural vibrations, and fatigue. Actually, this problem is solved by placing rigid baffles inside the container, which modify the acoustic modal fields and eventually inhibit the instability. For realistic industrial components using a restricted number of acoustical baffles, their optimal location is a challenging problem, as trial and error experimentation is often a costly and frustrating procedure. Recently, some strategies were proposed for the optimal location of a single baffle in a typical re-heater from a power station boiler, based on simulated annealing as well as genetic algorithm approaches. In this paper and using the above-mentioned optimization strategies, a more complex case of the problem—the optimal location of a given number (two or more) of baffles in a typical re-heater, was addressed. Some preliminary experiments were performed and compared with the simulation results. From the discussion, improvements in the developed optimization functional were proposed. [This work has been endorsed by the Portuguese FCT and POCI 2010, with funding participation through the EC programme FEDER (Project No. POCI/EME/5728/2004).]

9:30

4aEA3. Digital reconstruction of sound using array of microspeakers. Kassiani Kotsidou and Charles Thompson (ECE Dept., Univ. of Massachusetts Lowell, Lowell, MA 01854)

This work examines the digital reconstruction of sound using the method of matched asymptotic expansions. We will study an array of microspeakers and in particular, the role the cut off frequency and the damping ratio have in the successful reconstruction of sound. The transducers are fed by pulses whose duty cycle will be studied and its ramifications will be reported. We will consider how the coupling between the transducers affects the performance of the overall system.

9:45

4aEA4. Personal sound system design for mobile phone, monitor, and television set; Feasibility study. Chan-Hui Lee, Ji-Ho Chang, Jin Young Park, and Yang-Hann Kim (Ctr. for Noise and Vib. Cotrol, KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea, chance99@kaist.ac.kr)

Having a personal sound system that does not need to use an earphone or any wire connected microphone would have great interest and potential impact on the associated industries. A device that uses nonlinear characteristics of high intensity ultrasonic sound, such as a parametric array, enables us to localize the sound on any position we want. However, it requires a fairly significant amount of power consumption. Therefore, the mobile phone or note book computer system cannot accept this type of device. Instead, we have attempted to use a line array speaker system to localize the sound in our listening zone. It is known that sound pressure or intensity can be well focused in the zone of interest: For example, see [J.-W. Choi, Y.-H. Kim, J. Acoust. Soc. Am. **111** (2002)]. We applied the method to see the possibility to make a good bright zone around the human head or ears. Depending on the size of the zone and array parameters, for example, array size, speaker spacing, frequency characteristics of the speakers, the quality of the bright zone is changed. The quality means the ratio of sound energy between the bright and dark zone, regarded as acoustic contrast, analogous with what we use for an optical device.

Session 4aED

Education in Acoustics: Hands-On Experiments for High School Students

Uwe J. Hansen, Chair
Indiana State Univ., Physics Dept., Terre Haute, IN 47809

Approximately 20 acoustics experiments will be set up, ranging in complexity from simple resonance on a string to ultrasonic levitation. Students from local area high schools will perform the experiments with help from ASA scientists and students. Regular ASA conference participants are welcome to the session, as long as they do not interfere with student experimentation.

Session 4aMU

Musical Acoustics: Musical Acoustics and Instrumental Acoustics

Scott McDermott, Chair
1431 St. Claude Avenue, New Orleans, LA 70116

Contributed Papers

8:00

4aMU1. A theoretical dissipative analysis of optimized multimodal resonators coupled to room acoustics. José Antunes (Instituto Tecnologico Nuclear, Appl. Dynam. Lab., Estrada Nacional 10, 2686 Sacavem, Portugal, jantunes@itn.pl), Octávio Inácio (ESMAE-Inst. Politecnodo Porto, Porto, Portugal), François Axisa (Com à l'Energie Atomique, Gif-sur-yvette, France), and Vincent Debut (Inst. Tech. Nuclear, Sacavem, Portugal)

Helmholtz resonators are often applied for the sound equalization of control rooms in recording studios, through adequate leveling of the low frequency acoustic modal room responses. The number of controlled acoustic modes depends on the central frequency and damping of resonators, as well as on the modal density of the controlled system within the resonator's frequency range. In a recent paper, we proposed to improve the efficiency of such devices by, instead of using basic Helmholtz resonators, develop shape optimized multimodal resonators in order to cope with a larger number of intrusive room modes. In spite of the promising results thus obtained, further work is needed to demonstrate the feasibility of such an approach. The present paper is a further step in that direction by analyzing the acoustics of the fully coupled room/resonators system including dissapative effects. More specifically, using an acoustical substructure computational approach, we theoretically derive the coupled modes of rooms fitted with several optimized multimode resonators.

8:15

4aMU2. Auditory resolution in auditory virtual environments. Georgios Marentakis, Nils Peters, and Stephen McAdams (CIRMMT, McGill Univ., 555 Sherbrooke St. W., Montreal, QC, Canada H3A 1E3)

Minimum audible angles (MAA) were estimated for listeners in the sweet spot of four- and eight-loudspeaker arrays in the studio as a function of angle of incidence (0, 60, 90) and source position within the array (on the loudspeaker, midway between or one-third of the way between loudspeakers). Vector-based amplitude panning (VBAP) was used with a 250-ms broadband stimulus. MAAs increased with angle of incidence in direct comparison to real sources. For frontal incidence, MAA was 7.4, 2.9 for the four- and eight-speaker VBAP system (compared to 1 with real sources). A systematic effect of source position within the speaker array was found with MAAs being smallest for sources halfway between the speakers. The experiment was replicated in a concert hall for listeners at eight positions away from the sweet spot for eight- and 16-loudspeaker arrays and for angle of incidence of 0, 45, and 90 relative to a listener in the sweet spot, facing forward. Participants were able to perceive change in accordance to their relative angle to the stimulus. For frontal incidence, MAA at threshold was 3.24 and 2.0 for VBAP for eight and 16 speakers, respectively. The results are investigated using binaural localization models.

8:30

4aMU3. The enigma of Vitruvian resonating vases and the relevance of the concept for today. Rob Godman (Univ. of Hertfordshire, College Ln., Hatfield, Herts AL10, UK)

Anyone who has visited a Greek or Roman theatre cannot fail to be impressed by the overall clarity of sound without any form of enhancement. The seats arranged in curved rows around the circular orchestra form large horizontal reflecting surfaces. This ensures that the path of the sound waves travel from the source (the actor or singer) to each of the listeners in a direct path (i.e., without reflection). Vitruvius, however, claimed further enhancements could be made. ''In theatres, also, are copper vases and these are placed in chambers under the rows of seats in accordance with mathematical reckoning. The Greeks call them Echeia. The differences of the sounds which arise are combined into musical symphonies or concords: The circle of seats being divided into fourths and fifths and the octave.'' (Vitruvius, on Architecture, Book I,—on training of architects, Loeb). With the aid of modern-effect systems that attempt to mimic real and imaginary spaces, it may be difficult to imagine the importance of the Vitruvian idea. This paper demonstrates a number of digital reconstructions, examines the issue of intent, and questions an arguably more complex issue, that of a potential fusion between archaeology, science, and music.

8:45

4aMU4. Quantal computation of tonality in music. Alpar Sevgen (Dept. of Phys., Bogazici Univ., Bebek 34342, Istanbul, Turkey)

Each independent scale interval structure with N semitones and M notes contains a multiplet of N scales (for example, the major scale has the interval structure 2212221; and its twelve multiplet members are: C-major, D flat-major, . . . , B-major scales). Operators that step through these multiplet members (by sharping or flatting the notes of the multiplet members) are identical with those of N-level finite state quantum mechanics. It is possible to construct orthonormal basis states which represent the multiplet members of a given interval structure. A musical score can then be expressed as a mixture of various tonalities, using the density matrix formalism of quantum mechanics. Several examples help illustrate this quantitative approach.

9:00

4aMU5. Time series analysis of the Kurdish long necked lute, tanbour. Hedayat Alghassi (Dept. of Elec. and Comput. Eng., Univ. of British Columbia, Vancouver, BC, V6T 1Z4, Canada), Sohrab Ferdowsi, Roozbeh Alghassi (Tehran Univ., Tehran, Iran), and Babak Khademi (Stanford Univ., Stanford, CA 94305)

There are a variety of long necked, plucked string lutes in the East with the universal name tanbour/tanbur. This research concentrates on the three-stringed fretted tanbour of Yarsan Kurds of Gouran, which is the prevailing derivative of ancient tanbour originated from Persia. Contrary to most of Western stringed instruments, the vibration in tanbour is not limited to soundboard, especially at higher frequencies. This fact, along with string tension modulation, creates a nonlinear vibration of the body and neck. Therefore, the single channel sound pickup and linear modeling is not representing tanbour's complex sound field. Utilizing an array of four microphones, three spreading in front of soundboard and one in back of body, the system transfer function for tanbour's main plectrum was derived via suitably optimizing a nonlinear-auto-regressive-exogenous model with the envelope of signals extracted from five tanbours with different known body sizes. Besides neck length and body width, length, and depth, the location of bridge is incorporated in the model, since it has a major role in the vibration, quality, and final setup of tanbour. Employing this model and tanbour's physical sizes, one can estimate the optimal bridge location on the sound board to achieve finest vibration in any tanbour.

9:15

4aMU6. Implementation of non-stringed instruments without bodies. Sangjin Cho and Uipil Chong (Univ. of Ulsan, San29, Mugeo-Dong, Nam-Gu, Ulsan, Korea)

This paper describes the implementation of the non-stringed instruments using the TMS320F2812 DSP, which is possible to produce the sounds of musical notes. This basic concept of the instrument comes from the non-stringed instruments, which is not necessary to tune the string of the instruments before starting the play. In this paper, this system chose the guitar as an example for implementation. The system does not need a guitar body, but the pickup part is needed. The pickup of the system plays an important part in representing the up/down strokes and arpeggios. Therefore, the proposed system consists of three parts: Chord glove, pickup, and processor. Mapping between guitar chord and the glove takes place on the values of the force sensing resister and FLEX sensors. The extension of this system can be applied to stringed instruments in the future. [Work supported by KOSEF, R01-2005-000-10671-0.]

9:30

4aMU7. Experimental investigation of the effects of extra mass on a piano soundboard's vibration property. Hao Xing (Inst. of Acoust., Chinese Acad. of Sci., 21 Bei Si Huan Xi St., Beijing 100080, P. R. China), Bonan Zhao, and Haiyan Zhao (Tsinghua Univ., Beijing 100084, P. R. China)

A soundboard, as an important resonance component of a piano has been continuously studied in the past several decades. It is known that the variation of mass distribution on a soundboard could affect its vibration properties and the sound quality of a piano. However, those effects have been seldom investigated until now. In this paper, an experiment is carried out to study the effects of additional mass-blocks on a soundboard. Small pieces of metal blocks are pasted on several different locations on a soundboard, and a laser vibrometer is used to extract its mode shapes and spectrum under the excitation of an acoustic source with frequency ranges from 10 Hz to 1 kHz. The results show a notable difference of spectrum between mass-added cases and original in low frequencies, but little difference in high frequencies. It could be concluded that additional mass-block could greatly affect the energy of the low-mode of a soundboard.

9:45

4aMU8. Reed vibration, pressure, and airflow in Western free-reed instruments. Edward L. Toussaint (Lawrence Univ., Appleton, WI 54912, edward.l.toussaint@lawrence.edu) and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

Western free reed instruments, including the reed organ, harmonium, harmonica, and accordion, use asymmetric free reeds that are able to maintain sustained oscillation in the absence of a pipe resonator. Studies of the motion of air-driven American reed organ reeds and accordion reeds have been made that yield the volume air flow through the reed opening as a function of time. Reed vibration waveforms are obtained using a variable impedance transducer, and corresponding sound pressure waveforms are obtained from a probe microphone at positions near the reed opening on both sides of the reed tongue. The airflow waveform is obtained by integrating the pressure waveform and using a calculated area function and is calibrated using the measured average and minimum airflow rates. The results are compared with earlier preliminary studies as well as recent theoretical work. For pressure and airflow waveforms among different types of reeds or for the same reed at different blowing pressures, differences can be understood in terms of the configuration of the reed and reed frame system, as well as the amplitude of reed vibration. [Research supported by NSF REU Grant No. PHY-0354058.]

10:00–10:15 Break

10:15

4aMU9. Empirical physical modeling based music synthesis and representation. Xiaoxiao Dong, Mark Sterling, and Mark Bocko (Univ. of Rochester, 405 Comput. Study Bldg., 500 Joseph C. Wilson Blvd., Rochester, NY 14627, xidong@ece.rochester.edu)

We describe a method in which empirically-based musical instrument physical models are employed both to synthesize musical sounds and to form the basis of a compact representation of mono-timbral musical sound. Results are presented for a B^b clarinet. The physical model incorporates measured acoustic impedance spectra of a clarinet air column for all playable notes. Low bandwidth control parameter time histories, serving as inputs to the physical model, are inferred from audio recordings of actual clarinet music. The control parameters represent the fingerings, the blowing pressure of the player, and the mouthpiece clamping pressure of the player's embouchure. It is shown that, given an appropriate physical model, the control parameter time histories can serve as a highly compact representation (compression by a factor of several hundred) of a source recording. To serve as an assessment tool, we also describe a distance metric emphasizing timbral aspects of musical audio. The metric provides a method to quantitatively evaluate the parameter extraction routines, to

test the effect of minor alterations to the model, and also to compare the present physical modeling approach with other forms of musical sound synthesis. The framework described may be extended to other musical instruments and instrument families.

10:30

4aMU10. The influence of the acoustic feedback on the fluid-structure interaction within single-reed mouthpieces: A numerical investigation. Andrey Ricardo da Silva and Gary Scavone (Computational Acoust. Modeling Lab. and CIRMMT, McGill Univ., 555 Sherbrooke St. W., Montreal, Canada H3A 1E3, andrey.dasilva@mail.mcgill.ca)

The aim of this article is to complement a previous numerical study conducted to investigate the flow behavior in single-reed woodwind instruments during dynamic regimes with a moving reed [J. Acoust. Soc. Am. **120**, 3362 (2006)]. The code uses a fully-coupled scheme based on the lattice Boltzmann method and on a finite-difference scheme to represent the interaction between the reed and the fluid field. The previous study had suggested that, during a dynamic regime and in the absence of acoustic feedback, the flow behavior diverges significantly from what is predicted by the current quasi-stationary models based on the Bernoulli obstruction theory. The present work provides a further investigation on the same topic by taking into account the influence of the acoustic feedback from the bores open end and its interaction with the mean flow within the mouthpiece-reed system. This is achieved by coupling the previous mouthpiece-reed model to a digital waveguide whose reflection function is described by a low-order digital filter to represent the open-end boundary condition. [The first author would like to thank CAPES for supporting his doctoral research.]

10:45

4aMU11. Vibration modes of the snare drum batter head. Barry Larkin (Dept. of Music, Iowa State Univ., Music Hall, Ames, IA 50011) and Andrew Morrison (Illinois Wesleyan Univ., Bloomington, IL 61701)

Percussionists have always had to contend with an undesirable ringing sound while performing on the snare drum. It is usually referred to as the edge ring. A common method to eliminate this sound comes from placing some type of dampening material on the edge of the drum head. Manufacturers of drums have provided many ways of dealing with this problem, including internal dampening devices, customized drums heads, or material designed to be placed on the drum head. Using electronic television holography, was revealed the source of this ring to be the third mode of vibration that produces a pitch approximately one octave and a half-step above the fundamental. During this presentation, holographic images from the fundamental to the eighteenth mode will be displayed. Methods to dampen this mode will be displayed and demonstrated as part of this paper.

11:00

4aMU12. Acoustics of the mridangam: Study of a new design of a South Indian drum. Rohan Krishnamurthy (Kalamazoo College, Kalamazoo, MI 49006, rohan_krishnamurthy@rohanrhythm.com), Ian K. Hempe, and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

The acoustical properties of the ancient Carnatic drum, the mridangam, have been studied, including instruments of traditional design as well as instruments with a newly designed mounting method for the drumheads. The mridangam is comprised of three primary parts: The tonal head (valanthalai), the bass head (thoppi), and the central shell (katta), to which the two heads are traditionally fastened by leather rope. The new design is proposed as a convenient and practical way to mount and tune the drumheads. Measurements were made on drums with the drumheads mounted in the traditional way, and then repeated with the heads remounted using the new design. For measurements of the modes and mode frequencies of the drumheads, the modes were excited by a mechanical vibrator and vibration of the drumhead was measured using a laser vibrometer system. The frequencies of the first few modes of the tonal head were found, as expected, to be tuned to be approximately harmonic. Measurements of drumhead vibration and sound spectra were also made when the drumhead was excited by a skilled player using standard strokes. Some effects of coupling between the two drumheads were studied. [Work supported by the Kalamazoo College HHMI Undergraduate Research Program.]

Session 4aNS

Noise and Animal Bioacoustics: Advances in Measurement of Noise and Noise Effects on Animals and Humans in the Environment I

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Fluid Mechanics and Engineering, Secr TA 7, 10587 Berlin, Germany

Chair's Introduction—8:00

Invited Papers

8:05

4aNS1. Animals, ordinary humans, and scientific experts: Noise(s) or sounds? How semantics may help understanding animal behavior and human cognition as well. Daniele S. Dubois (LCPE/LAM/IJLRDA, 11 rue de Lourmel, 75015 Paris, France, ddubois@ccr.jussieu.fr)

If an anthropomorphic conception of animal behavior is inescapable, cross cultural studies, situated cognition, as well as research in ethology, have recently pointed that contemporary cognitive models of information processing that pretend to be universal of any living system, could be misleading as relying on a specific conception of cognition. Actually, such models can account for the cognition of experts trained in an analytic processing of the world, grounded in contemporary scientific (physical) knowledge, but not for the ways ordinary humans perceive and react to environmental noise for behaving and acting in the world. In every day life, humans process multimodal incoming stimulations in a holistic manner that gives meaning to their ways of being. Within this frame, acoustic stimulations are not processed per se, but as cues pointing to categorical knowledge already stored in memory and structured according to adaptation purposes. Such a situated model of cognition has been fruitfully integrated in soundscape research and in environmental acoustics. It now challenges the traditional psychophysics paradigms in developing an ecological exploration of categorization for sounds. We would like to present such methodological issues, expecting to provide new insights for human as well as for animal behavior research.

8:30

4aNS2. Noise and noise effects on humans and its meaning for measurement decisions. Brigitte Schulte-Fortkamp (TU-Berlin ISTA, Einsteinufer 25, D-10587 Berlin, Germany, brigitte.schulte-fortkamp@tu-berlin.de)

To discuss advances in measurements of noise and noise effects on animals and humans is challenging. Usually, it is physics that dominates the approach in the studies about noise and noise effects on humans; indicators in use are based on A-weighted sound pressure levels. By extension, psycho-acoustics has contributed to understanding the limitations of the A-weighted sound level in dB as a metric criterion or a criterion for measurements. In addition, noise annoyance appears to be differently perceived, expressed, and experienced in different places. For this reason, supplementing the acoustics and psycho-acoustics approaches by a more qualitative approach, as in soundscape research, seems to be appropriate in order to consider more the perception, the way people experience noise, and relate this to their well-being. Research shows that perceived and expressed annoyance are explained not only by acoustics, but also by other factors, and their relationships can explain the links between populations and their environment. Therefore, the way people react and evaluate noise as the real experts has to be taken into consideration with respect to conceptualization of measurement approaches and measures. The diverse approaches will be introduced with respect to decisions for measurements in defined environments.

8:55

4aNS3. Documenting noise exposures for wildlife studies: Review of large studies over the past 18 years. Robert Kull (Parsons, 5800 Lake Wright Dr., Ste. 101, Norfolk, VA 23502, bob.kull@parsons.com)

A couple of decades ago, researchers recognized the importance of improving understanding of noise exposure to wildlife for bioacoustic studies. Due to the differences in the ecology of species of concern, various approaches took place over these past 20 years to address this issue. The metrics used to determine effects and how the metrics were applied to observed behavioral and physiological responses also varied, based on several factors. This presentation will review methods used on several large projects, including big horn sheep, caribou, kit fox, Mexican spotted owls, peregrine falcons, and red cockaded woodpeckers. This review may help provide the direction for future bioacoustic studies.

9:20

4aNS4. The design of experimental studies of noise impacts on wildlife: Issues and outcomes. Ann E. Bowles, Donald Hunsaker II, Samuel L. Denes (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109), Robert Kull (Parsons Corp., Norfolk, VA 23502), Jeffrey R. Dunk (USDA Forest Service, Arcata, CA 95521), Lisa Hayward, and Samuel L. Wasser (Univ. of Washington, Seattle, WA 98195)

Historically, most studies exposing terrestrial wildlife to anthropogenic noise experimentally have failed to prove biologically-significant impact. These counter-intuitive outcomes have given rise to political debate, fueled by marginal effects or weak trends in some studies that might have proved to be significant with better sampling procedures or controls. A number of recent studies have attempted to address these weaknesses by developing adequate sampling designs and balancing exposures with respect to important environmental factors (e.g., habitat type). Examples include studies of: (1) Effects of low-flying NATO aircraft on Mexican spotted owl territory occupancy and reproduction; (2) effects of U.S. Marine Corps. helicopters on passerine reproductive success; (3) effects of operations on a U.S. Air Force bombing range on small mammal abundance; (4) impact of off-highway vehicles on reproduction of northern goshawks in the Plumas National Forest; and (5) impact of Enduro motorcycle races on northern spotted owl physiology in the Mendocino National Forest. To date, preliminary outcomes of the studies suggest that more sophisticated models of effect should be developed, and additional research should focus on long-term cumulative impact.

9:45

4aNS5. Response of gopher tortoises to military training operations: Preliminary data. David Delaney, Larry Pater (USACERL, P.O. Box 9005, Champaign, IL 61826), Tom Radzio, Joseph Hackler, Andrew Walde (ITS, Victorville, CA 92395), and Matthew Hinderliter (The Nature Conservancy, Hattiesburg, MS 39407)

The Department of Defense (DOD) needs defensible data to understand how gopher tortoises utilize military lands to address the question of potential military effects on tortoise populations, while also maintaining its ability to continue to train to standard. The objective of the research is to examine if there are differences in tortoise activity and behavior due to variation in military training activity. Automated radio-telemetry equipment is being used to monitor the activity patterns of gopher tortoises on Camp Shelby, MS. The automated radio-telemetry equipment enables one person to monitor the activity of many transmittered tortoises continuously over an extended time, regardless of weather, light level or terrain. Video surveillance and sound recording equipment are also being used to monitor tortoise behavior in proximity to burrows and quantify and characterize potential anthropogenic disturbances within study locations. The goal of this study is to benefit the recovery and management of gopher tortoise populations on DoD and non-DoD lands by providing natural resource managers detailed temporal and spatial data on gopher tortoise activity, behavior, and movement patterns under varying conditions. Funding was provided by the Engineering Research and Development Center. Preliminary examples of field data and database tables will be presented.

10:10–10:25 Break

10:25

4aNS6. The effects of helicopter noise on the reproductive success of the coastal California gnatcatcher. Don Hunsaker II (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, dhunsaker@hswri.org), Jacque Rice (US Naval Facilities Eng. Command, San Diego, CA 92132), and John Kern (Kern Statistical Services, Sauk Rapids, MN 56379)

Our laboratory conducted a five-year study on the potential effects of helicopter noise on the reproductive success of the coastal California gnatcatcher (*Polioptila californica californica*) on Marine Corps Air Station Miramar (MCAS Miramar) in Southern California. Seven-hundred twenty-one nests were monitored for reproductive success, predation, noise levels, and habitat quality. An array of Larson-Davis sound level meters was used to monitor habitat on MCAS Miramar for a total of 6,176 days during 620 runs at 328 locations. Most sites were exposed to noise in excess of 60 dBA SPL for less than 5% of the monitoring period, but some nests experienced levels in excess of 70 dBA for more than 20% of the time. Statistical models of nest success, nest site selection, and number of fledges per pair showed that the factors best predicting reproductive success were measures of suitable nesting habitat, not noise levels. Helicopter and other noise sources did not affect the reproductive success of gnatcatchers. [Supported by the Marine Corps Air Bases Western Area and the Naval Facilities Engineering Command, SW.]

10:50

4aNS7. Effects of the sounds from an artificial oil production island on bowhead whale calling behavior. Susanna B. Blackwell, Charles R. Greene, Jr. (Greeneridge Sci., Inc., 1411 Firestone Rd., Goleta, CA 93117), Trent L. McDonald, Ryan M. Nielson, Christopher S. Nations (WEST, Inc., Cheyenne, WY 82001), W. John Richardson (LGL, Ltd., King City, Ont. L7B 1A6, Canada), and Bill Streever (BP Exploration (AK) Inc., Anchorage, AK 99519-6612)

The westward migration of bowhead whales (Balaena mysticetus) was studied during autumn 2001–2004 to examine the effects of sounds from an artificial oil production island (Northstar) on whale calling behavior. Whale calls were recorded by an array of directional autonomous seafloor acoustic recorders (DASARs) located 6.4–21.5 km northeast of Northstar in the Beaufort Sea. Simultaneously, a continuous record of sounds produced by Northstar was obtained ~450 m north of the island. More than 130,000 bowhead calls were detected, and were classified as to their type, including five types of simple calls and one complex call category. In addition, the directional capability of DASARs allowed triangulation of an estimated whale position for a majority of the calls. The primary objectives of the study were to assess any effects of Northstar sounds and environmental-physical covariates on the duration, midfrequency, frequency range, type of call, and call detection rates. The analyses showed that an increase in transient sounds from

Northstar (i.e., boats) resulted in significantly shorter calls. In addition, call detection rates were significantly higher for whales upstream of the array, and the use of complex calls increased significantly as the whales swam westward past Northstar. [Study funded by BP Exploration, Alaska.]

Contributed Papers

11:15

4aNS8. Marine wildlife behavior database for estimating environmental impacts. Kathleen J. Vigness Raposa, Adam S. Frankel, Geoff Sisson, William T. Ellison (Marine Acoust., Inc., 809 Aquidneck Ave., Middletown, RI 02842, kathleen.vigness@marineacoustics.com), and Christopher Damon (Univ. of Rhode Island, Kingston, RI 02881)

Assessing the potential effects of underwater sound on the environment is an increasing concern. In order to predict an animal's exposure from an acoustic source, it is necessary to know how it moves through the sound field. While a number of tools have been developed in recent years to estimate environmental impacts, they require an estimate of the species present at the time of the operation, their abundance and distribution, as well as their diving, movement, and acoustic behaviors. To determine whether a species might be susceptible to certain sound exposures, it is also necessary to consider the probable hearing range and sensitivity of the species. The marine wildlife behavior database (MWBD) is creating a worldwide standard for measuring and characterizing behavior. It includes geo-referenced baseline data on diving, movement, and acoustic characteristics of fishs, marine mammals, and sea turtles. The database is designed to capture the distribution of the values of a behavior in a way that is amenable to models that replicate the movement or other behaviors of an animal. The MWBD is publicly available over the Internet through a registered node on the National Spatial Data Infrastructure at http://www.edc.uri.edu/MWBD.

11:30

4aNS9. Update on exposure metrics for evaluation of effects of sound on fish. Mardi C. Hastings (Appl. Res. Lab., Penn State Univ., State College, PA 16804) and Arthur N. Popper (Univ. of Maryland, College Park, MD 20742)

Several studies [A. R. Scholik and H. Y. Yan, Hear. Res. **152**, 17–24 (2001); A. N. Popper et al., J. Acoust. Soc. Am. **117**, 3958–3971 (2005); M. E. Smith et al., J. Exp. Biol. **209**, 4193–4202 (2006)] indicate that the equal energy hypothesis does not apply when evaluating auditory effects in fish. This is not surprising, as similar results have been reported for mammals [R. P. Hamernik and W. Qiu, J. Acoust. Soc. Am. **110**, 3163–3168 (2001)]. New data show that fish recover relatively quickly from large temporary threshold shifts (TTS) after exposure to impulsive sounds, but take several days to recover from smaller amounts of TTS due to exposure to band-limited noise. Recent studies [e.g., J. J. Govoni et al., J. Aquatic Anim. Health. **15**, 111–119 (2003)] also validate that peak sound pressure is not an appropriate metric for effects on nonauditory tissues, even if the peak level is negative. Large negative pressures have been thought to be more harmful due to swim bladder expansion, but there are no data to support this hypothesis. Physical injury correlates with a cumulative energy index, as has been found in terrestrial mammals [J. H. Stuhmiller et al., J. Biomechanics **29**, 227–234 (1996)].

11:45–12:00
Panel Discussion

FRIDAY MORNING, 30 NOVEMBER 2007 BORGNE, 8:00 TO 10:30 A.M.

Session 4aPA

Physical Acoustics: Acoustics of Soils and Porous Media

Wheeler B. Howard, Cochair
National Center for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Alexander Ekimov, Cochair
National Center for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Contributed Papers

8:00

4aPA1. An experimental study for the penetration of sound into a rigid porous medium. Hoyt Chang and Kai Ming Li (Ray W. Herrick Labs, School of Mech. Eng., Purdue Univ., 140 S. Intramural Dr., W. Lafayette, IN 47907-2031, mmkmli@purdue.edu)

In a recent study, a new analytic solution for computing the penetration of airborne sound into a rigid porous ground has been derived. Modeling the lower medium as a modified fluid, a simple analytic solution can be obtained that can be cast in a form analogous to the Weyl-van der Pol formula for the sound field in the upper medium. The current study is devoted to an experimental validation of the analytical model. A layer of glass beads is used in the experimental study for modeling the rigid porous ground. To test the analytical formulation, measurements of excess attenuation spectra above and within the layer of glass beads were conducted. The impedance, propagation constant, and the index of refraction of the glass bead aggregate were deduced from the excess attenuation spectra. The experimental results are compared with other methods for finding the propagation of the rigid porous medium. The proposed method suggests a new way of finding the propagation constant and the index of refraction of a rigid porous medium.

8:15

4aPA2. Refraction and attenuation of sound in a rigid porous medium. Kai Ming Li (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031)

An investigation was conducted to predict the penetration of sound into a ground surface irradiated from a point source located in the air above the air/ground interface. To simplify the analysis, the outdoor ground surface is treated as a modified fluid in which its acoustical properties are characterized by using a complex density and a propagation constant of the porous medium. By modeling the porous ground as a

modified fluid, the physical phenomenon for the refraction of sound from the upper medium (air) to the lower medium (porous ground) has been explored. It has been shown from the mathematical analysis that the refracted path can be determined by Snells Law. The sound field can be formulated in an integral form for subsequent evaluations. Two approaches are used to calculate the integral, numerically predicting the total attenuation of sound in the porous medium. A numerically intensive scheme, known as the fast field formulation, and an analytical method known as the steepest-descents integration, are used to calculate the sound fields in the lower medium. Numerical results computed from both schemes agree well with each other over the practical range of interest.

8:30

4aPA3. Acoustic to seismic signatures of layered near-surface soil. Wheeler Howard (MIL-TEC, Oxford Enterprise Ctr., 9 Industrial Park Dr., Oxford, MS 38655) and Craig J. Hickey (NCPA, Univ. of Mississippi, University, MS 38677)

The near-surface soil structure, which strongly influences agricultural productivity and soil erosion, is amenable to study by acoustic to seismic techniques. These techniques utilize the coupling between airborne sound and seismic vibrations of the soil to investigate the distribution of soil's mechanical properties. Two field sites were chosen over which the acoustic to seismic coupling signature of the ground, was measured. A predictive model, assuming acoustic plane waves incident on a vertically stratified ground was developed to predict the acoustic to seismic response. In this presentation, the acoustic to seismic signatures from these sites are compared to model results using the ground truth from seismic refraction and cone penetrometer surveys as input to the model.

8:45

4aPA4. The influence of water potential, moisture, and temperature on sound speed and the hysteretic nonlinearity parameter of granular materials. Zhiqu Lu (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, University, MS 38677)

In the 153rd ASA meeting (Salt Lake City, Utah), the results of a long-term survey of a field soil by using both linear and nonlinear acoustic techniques were reported. The study showed that soil water potential is the predominant factor governing the sound speed and the hysteretic nonlinearity parameter while moisture plays a minor role. It was also found that, unlike sound speed that is insensitive to temperature, the nonlinear parameter shows strong temperature-dependent behavior. The question remains: Is this temperature effect an intrinsic property of the hysteretic nonlinearity parameter, or is it caused by the degradation of the transducers since the measurement has been running for two years? To answer this question, the influence of water potential, moisture, and temperature on sound speed and the hysteretic nonlinearity parameter of granular materials (unconsolidated sands) has been studied in the lab with temperature-controlled conditions.

9:00

4aPA5. Theoretical modeling of active seismic detection of shallow buried tunnels. Elizabeth T. Küsel, Mark Andrews, and Purnima Ratilal (Northeastern Univ., 440 Dana Res. Ctr., Boston, MA 02115, kusele@ece.neu.edu)

Low-frequency seismic waves are widely used to detect oil/gas reservoirs, and study the properties of Earth's interior. Here, we investigate the application of low-frequency active seismic waves (tens of Hz to a couple of hundred Hz) for long-range wide-area detection of shallow underground tunnels. The main objective is to image wide areas of the ground at the longest possible ranges to identify potential underground tunnel hotspots. The tunnel is modeled as a circular air-filled cylinder. The incident field at the tunnel is calculated in terms of the horizontal (u) and vertical (w) displacements by the elastic parabolic equation. The scattered field from the tunnel is modeled by application of Green's theorem, approximating the radial dependence of the scattered field by the Hankel function, and satisfying the boundary conditions. Compressional and shear waves scattered from the tunnel will be calculated to determine which wave type is most suitable for detection purposes. Scattered field levels from the tunnel will be compared with seismic ambient noise and reverberation from surrounding medium heterogeneities. Tunnel detection ranges will be quantified as a function of tunnel diameter and length, seismic wave frequency, receiving array length and orientation, and medium properties.

9:15

4aPA6. Varied amplitude nonlinear acoustic method of landmine detection. Laurent Fillinger (Stevens Inst. of Technol., Hoboken, NJ 07030, laurent.fillinger@artannlabs.com), Brad Libbey (RDECOM CERDEC NVESD, Fort Belvoir, VA 22060), Alexander Sutin, and Armen Sarvazyan (Artann Labs., West Trenton, NJ 08618)

Acoustic methods of land mine detection are being developed for nonmetal mines where conventional electromagnetic methods fail. Nonlinear acoustic methods have shown higher sensitivity than linear methods. In previous studies, the nonlinear methods, based on interaction of two frequency waves and phase inversion of two pulses, were investigated. We propose an alternative nonlinear method based on the use of wide-frequency pulses with different amplitude and subtraction of the normalized received signals. This method evaluates the nonlinear effects in a wide-frequency band similar to the phase inversion method and it detects nonlinear components that are lost in the phase inversion method. In our tests, the suggested method was combined with time reversal acoustic focusing to concentrate seismic wave energy in proximity to the mine, which increased nonlinear effects. Experiments with various types of mines were conducted at Fort Belvoir using six loudspeakers in a box for excitation in the frequency band 50–500 Hz. The nonlinear mine response measured by the varied amplitude method provides the highest mine/no mine contrast among all the experiments performed using linear and nonlinear techniques. [Work was supported by the U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate.]

9:30

4aPA7. Evaluation of parametric acoustic array technology in application to the acoustic landmine and improvised explosive device (IED) detection problems. Murray S. Korman (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402), Antal A. Sarkady (U. S. Naval Acad., Annapolis, MD 21402), and James M. Sabatier (Univ. of Mississippi, University, MS 38677)

There has been recent interest in using the parametric acoustic array for obtaining highly directional low frequency sources in acoustic landmine and IED detection. Experiments (using a commercial parametric array (0.5 m diam) located 1.5 m directly over the target and driven by a swept sine audio modulated 60 kHz tone) were performed to compare the ''on'' to ''off'' target soil surface particle velocity of an inert VS 1.6 (plastic) anti-tank landmine buried 2.5 cm deep in masonry sand in a concrete box located in the anechoic chamber facility. The laser Doppler vibrometer response (normalized to the microphone) was measured from 100–1,100 Hz using a spectrum analyzer. A SPL of 40 dB $re\ 20\ \mu\ rmPa$ at 150 Hz was insufficient, where the mine was known to have its largest resonance. Resonances between 300–600 Hz also went undetected, but amplitude contrast ratios of 20 and 3 were measured at 850 and 1,050 Hz, respectively, upon signal averaging. Limitations in carrier SPL, auditory damage, saturation effects, distortion, and low efficiency need to be addressed to achieve the required 60–80 dB SPL needed near 100 Hz. Mines or IEDs buried at 15 cm require more SPL. [Work supported by ARL.]

9:45

4aPA8. Seismic excitation using a high power CO_2 laser. Brad Libbey, Jamie Krissoff, James Perea, and Gwen Newsome (Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Fort Belvoir, VA 22060)

Lasers provide one method of remotely exciting propagating waves in soil. The absorption of optical energy results in plasma formation, a dynamic change near the soil surface, and excitation of propagating waves. Experiments were performed using a high power CO_2 pulsed laser to

determine the pressure waveforms generated in desert sand and caliche clay. The laser energy ranged from 400–750 J with a pulse duration of 35 μs. The resulting pressure pulse was observed for duration of at least 400 μs and had a peak pressure of 100 kPa. The duration increased at greater depths, while the peak pressure and energy in a pulse decreased rapidly at these depths. Surface velocity reached a peak of 35 mm/s as acquired on geophones placed in proximity to the laser plasma event. This system, in combination with a laser Doppler vibrometer, has the potential to be a useful component of non-invasive mine detection equipment.

10:00

4aPA9. Directivity pattern of footstep sound at ultrasonic frequencies. Alexander Ekimov and James M. Sabatier (Univ. of Mississippi, NCPA, 1 Coliseum Dr., University, MS 38677)

The human footstep generates broadband frequency sound in air due to friction forces between the foot and the ground surface. An ultrasonic (25–26 kHz) footstep signal was measured on the ground at a distance of 17 m from the walker. Another aspect of footstep sound characterization is a directivity pattern. The sound wavelength at those frequencies corresponds to 13–14 mm, which is much less than the source size (the area of sound generation is approximately the size of a foot, 300 mm$\times$120 mm). The results of the experimental investigations are presented and discussed. An analysis of test results shows that the directivity pattern of footstep sound has inhomogeneous dependence on the azimuthal angle. A maximum value of the pattern corresponds to the direction of walking, while the minimum value is in the opposite direction. [Work supported by Department of the Army, Army Research Office Contract No. W911NF-04-1-0190.]

10:15

4aPA10. Seismic refraction tomography of a small earth dam. Craig J. Hickey (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38655) and Lakshmikanth Duddu (Univ. of Mississippi, University, MS 38655)

Over the past 50 years there were nearly 11,000 flood control dams constructed by the United States Department of Agriculture (USDA) nationwide in 2,000 watersheds (USBR, 1987). The failure of these earth dams is associated with erosion, seepage, piping, slope failures, and slides. Precursory evidence for certain failures can be identified by visual inspection; however, others are not so easily detected. Seismic imaging may provide unique and valuable information about the subsurface to assist in the evaluation of the integrity of an earth dam. In this presentation, results from a field study carried out on Drewery Lake Dam in the Upper Yocona River watershed to determine the feasibility of using seismic refraction tomography for imaging earthen dams is discussed. Attributes in the refraction tomograms are correlated with the location of the drainage pipe and the excavated surface given in the as-built plans indicating that seismic refraction tomography does have sensitivity to the internal structure of a dam. This study also suggests that the construction practices for earthen dams result in seismically uniform structures. A uniform background is very advantageous if seismic tomography is to be used to detect zones of internal weakness.

FRIDAY MORNING, 30 NOVEMBER 2007 NAPOLEON D2/D3, 8:00 A.M. TO 12:00 NOON

Session 4aPP

Psychological and Physiological Acoustics: More Potpourri (Poster Session)

Edward Goshorn, Chair

Univ. of Southern Mississippi, 118 College Dr., #5092, Hattiesburg, MS 39406-0001

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.

4aPP1. Are there systematic movements of poles/zeros with sound source location for pole/zero models of acoustic directional transfer functions? Bahaa Al-Sheikh, Matin Matin (Dept. of Elec. Eng., Denver Univ., 2390 S. York St., CMK 206, Denver, CO 80208), and Daniel Tollin (UCHSC, Aurora, CO 80045, daniel.tollin@uchsc.edu)

Humans and animals locate sound sources using three main acoustical cues: Interaural time (ITD) and level (ILD) differences and monaural spectral shape cues. The cues are generated by the spatial- and frequency-dependent filtering of the propagating sound waves by the torso, head, and external ears. These acoustic transformations can be captured in measurements of the directional transfer functions (DTFs), the directional components of the head related transfer functions (HRTFs). A set of DTFs can be used to create a virtual auditory environment by presenting over headphones arbitrary sounds filtered with DTFs. DTFs are generally measured at finite locations in azimuth and elevation, so models are needed to synthesize DTFs for source locations not measured. Here, we investigated the use of different pole/zero models with different orders to model DTFs measured in humans and cats. With models that successfully characterized the DTFs with little error, we examined the systematic movements of the poles/zeros in the unit circle with changes in source azimuth and elevation. The ultimate goal of this research is to find a simple relationship between the movements of the poles/zeros with source location. [NIH Grant No. R01-DC6865.]

4aPP2. A study on the input-output function of a nonlinear cochlear transmission-line model with the function of an outer hair cell model. Yasuki Murakami and Masashi Unoki (School of Information Sci., JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan, y-muraka@jaist.ac.jp)

A nonlinear cochlear transmission-line model with the physiological function of an outer hair cell (OHC) model is presented to investigate how OHCs produce the nonlinearity in the input-output (IO) functions of cochlear filtering. As the physiological function of the OHC model, the pressure produced by the somatic motility of the OHCs was studied. The somatic motility is induced by the transducer currents that change with the displacement of the hair bundle. The pressure produced by the OHC

model can be linearly approximated by the amount of current. The proposed model is comprised of a middle ear model and a cochlear transmission-line model, which includes a basilar membrane and lymph fluid model, a tectorial membrane model, hair bundles model, and the OHC model. The model's parameters were set as estimates for a human obtained from animal data. In simulation results, this model could account for the IO functions of cochlear filtering obtained from both physiological and psychological experiments. These results suggest that the somatic motility of the OHCs, which depends on the transducer currents, produces the nonlinearity in the IO functions of cochlear filtering.

4aPP3. Does Norwich's entropy theory of perception produce equal-loudness contours? Lance Nizami[a] (1312 Grayson Pl., Decatur, GA 30030, nizamii2@aol.com)

Norwich & Wong [Percept. Psychophys. 59, 1997; derived backwards in JASA, 97, 1995] modeled loudness L versus intensity I by $L(I,k,\gamma,I_{th},n)=(k/2)ln(1+\gamma(I/I_{th})^n)$. They assumed $L_{th}=L(I_{th})$ and $L=L-L_{th}$, producing *two* loudnesses, L or $L=L-L_{th}$, giving $L(I_{th})=L_{th}$ or 0, and $L(0)=0$ or $(-k/2)ln(1+\gamma)$, a *negative*. They took dL/dI, called it $\Delta L/\Delta I$, and rearranged it, giving $\Delta I/I(\Delta L,k,\gamma,I_{th},n)$. New terms $S_\infty(\Delta L,n,k)$ and $S_0(\Delta L,n,\gamma,k)$ were introduced, thus $\gamma=S_\infty/(S_0-S_\infty)$. The result was $\Delta I/I$ in S_∞, S_0, and n, identical to an equation of Riesz (Phys. Rev., 31, 1928). Riesz's empirical S_∞, S_0, and n provided γ's and n's for $L(I)=L(I,k,\gamma,I_{th},n)-(k/2)ln(1+\gamma)$. *Assuming k was constant*, Norwich & Wong equated $L(I)$'s for comparison and reference tones, generating equal-loudness contours. The contours contradicted empirical ones; to adjust, they replaced Riesz's n by Stevens' Index, x. The present author curvefitted the Entropy Equation and Stevens' Law to 37 loudness-growth plots. k depended on maximum loudness as $k=0.173L_{max}^{1.41}$ $n\neq x$, rather, $n=0.123+1.215x$. Altogether, the derived contours are invalid. [a]Research done at Univ. of Toronto in Mississauga, 3359 Mississauga Rd. N., Mississauga, ON, Canada.

4aPP4. Tonotopic map reorganization and spectral weights in high-frequency sensorineural hearing loss. Blas Espinoza-Varas (Dept. Commun. Sci. and Disord., OU Health Sci. Ctr., 825 NE 14th St., Oklahoma City, OK 73190), Hammad Akram, Titus Oleyadun (College of Public Health, Oklahoma City, OK 73190), Merlyn George (Univ. of Oklahoma, Norman, OK 73019), and Hyunsook Jang (Hallym Univ., Chuncheon 200-702, Korea)

Owing to reduced afferent input from high characteristic-frequency hair cells, high-frequency sensorineural hearing loss (SNHL) reorganizes the tonotopic maps in the central auditory nervous system: Map expansion is observed in the midfrequencies and map contraction in the high frequencies. Does the relative weighting of mid and high frequency information depend on this reorganization? With complex tones consisting of a 1,000- and a 3,127-Hz sinusoid, 80 and 60 ms each, the resolution of simultaneous increments in frequency (IF) at 1,000 Hz and in duration (IT) at 3,127 Hz was studied in participants having either normal hearing or moderate-to-severe SNHL above 2,000 Hz. In both groups, the sensation level at 1,000 Hz (1.5–8 dB) was 17–38 dB lower than at 3,127 Hz (25–39 dB). In normal-hearing participants, IF resolution was inferior to that of IT, suggesting stronger weighting for the more audible component. In hearing-impaired participants, IF resolution was superior or equal to that of IT, suggesting stronger weighting for the lower sensation level 1,000-Hz component impinging on the expanded tonotopic map region; in addition, IT resolution improved significantly following extensive training. [Work supported by PHF and NIH grants.]

4aPP5. Independence in the perception of complex non-speech sounds. Noah Silbert (Dept. of Linguist., Indiana Univ., Memorial Hall 322, 1021 E. Third St. Bloomington, IN 47405), James Townsend, and Jennifer Lentz (Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Little, if any, work has explicitly addressed independence in the perception of complex sounds. General recognition theory provides a powerful framework in which to address such issues. Two experiments were carried out to test within-stimulus, between-stimulus, and decision-related notions of independence in each of two stimulus sets. One set consisted of broadband noise stimuli varying in frequency and duration, the other set consisted of harmonic tone complexes varying in fundamental frequency and spectral shape. Model fitting analyses indicate that decision-related independence (decisional separability) holds for all participants with the noise stimuli and the majority of participants with the harmonic stimuli, that within-stimulus independence (perceptual independence) holds for all participants with the noise stimuli and between-stimuli independence (perceptual separability) holds for all but two participants with the noise stimuli, whereas considerable individual differences are exhibited in a variety of failures of either perceptual independence or perceptual separability or, occasionally, both, for participants with the harmonic stimuli. The noise stimulus results are consistent with expectations based on models of frequency and duration encoding, whereas the harmonic stimulus results present a challenge to a number of models of pitch and timbre encoding. [This work was supported by NIH Grant No. T32 MH019879-13.]

4aPP6. Development and achievement of hearing-aid fitting strategy based on loudness model. Ying-Hui Lai, Yang-Ming Jhou, and Sheunn-Tsong Young (Inst. of Biomed. Eng., Natl. Yang-Ming Univ. No. 155, Sec. 2, Linong St., Taipei City 112, Taiwan, young@bme.ym.edu.tw)

The purpose of this study was to develop a hearing-aid fitting strategy (named NYMU strategy), which aimed to compensate the loudness perception of hearing-impaired individuals as that of normal-hearing individuals. The loudness growth curve of normal-hearing individuals was derived by the equal loudness contour. The loudness growth curve of hearing-impaired individuals was constructed by fitting the curve through three points, which were the uncomfortable loudness level, the most comfortable loudness level, and the pure-tone threshold of hearing-impaired individuals for each frequency band. The NYMU strategy primarily calculated the hearing-aid gain for each frequency band according to the differences between the loudness growth curve of normal-hearing individuals and that of hearing-impaired individuals; the gain was set to guarantee the audibility of the lowest boundary of speech intensity with normal vocal effort. Furthermore, the NYMU strategy adopted an innovative approach, which adjusted the point of the most comfortable loudness level to fine-tuning the loudness growth curve, and it resulted in the fine-tuning gain of each band. The preliminary results reveal that the NYMU strategy provides more gain above 4,000 Hz and under 500 Hz, compared with the NAL-NL1 strategy. In addition, three of five subjects get higher word recognition scores when they were fitted by the NYUM strategy rather than the NAL-NL1 strategy.

4aPP7. Off-frequency listening: A population study. Veronica Eckstein and Bruce 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 in a college-age population. For different conditions, the central tone was 1,000 Hz with the side tones spaced by either 15 Hz or 45 Hz. The tones were 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. Psychopys. **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 the 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. Results for the population were consistent with pitch matches for expert listeners. [Work supported by the Center for Hearing Research, UCI.]

4aPP8. Predicting the ability to learn complex auditory tasks. Charles S. Watson and Gary R. Kidd (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

The goal was to determine whether the ability to learn complex auditory discrimination and identification tasks could be predicted from performance on a screening test battery. The screening battery included auditory tasks chosen on the basis of earlier factor-analytic studies of individual differences [Kidd et al. JASA **122**, 418–435 (2007)], visual acuity tasks, and a test of working memory. These tests were administered to 1,000 college students with normal audiograms. Based on screening test performance, 25 exceptional performers and 25 average performers were selected for prolonged training on three new auditory tasks. The training tasks included multi-tone pattern discrimination, identification of a vocabulary of sequences of three-tone syllables, and binary categorization of multi-dimensional profile-like stimuli. A broad range of individual differences in post-training performance was observed on each of these tasks, but differences were not well predicted by screening task performance or estimated general intelligence. Analysis of the screening task performance supported the previously reported independence of individual differences in spectral-temporal acuity and the recognition of familiar sounds (both speech and nonspeech). Intellectual abilities and visual processing abilities were also independent of each other and of auditory abilities. [Work supported by Grant No. N000140310644 from the Office of Naval Research.]

4aPP9. SPATS: Speech perception assessment and training system. James D. Miller, Charles S. Watson, Diane Kewley-Port, Roy Sillings, William B. Mills, and Deborah F. Burleson (Commun. Disord. Technol., Inc., 501 N. Morton St., Sta 215, Bloomington, IN 47404)

A software system (SPATS) to test and train important bottom-up and combined bottom-up/top-down speech-perception skills is described. Bottom-up skills are the abilities to identify the constituents of syllables: Onsets, nuclei, and codas in quiet and noise as produced by eight talkers. Top-down skills are the abilities to use knowledge of linguistic context to identify the words in spoken sentences. The sentence module in SPATS emphasizes combined bottom-up/top-down abilities in recognizing sentences in noise. The word-initial onsets, stressed nuclei, and word-final codas are ranked in importance and grouped into subsets, based on their importance. Testing utilizes random presentation of all the items included in a subset. Training in Quiet (SNR=40 dB) or in Noise (SNR=5 dB), is adaptively focused on individual listeners learnable items of intermediate difficulty. Alternatively, SNR-adaption training uses Kaerenbachs algorithm to find the SNR required for a target percent correct. The unique sentence module not only assesses and trains a listener's ability to infer words in sentences based on linguistic cues, but also trains the ability to combine bottom-up with top-down abilities to identify words in spoken sentences in noise. Scoring in the sentence module is objective and automatic. SPATS demonstrations are included.

4aPP10. Preliminary evaluation of the speech perception assessment and training system (SPATS) with hearing-aid and cochlear-implant users. James D. Miller, Charles S. Watson (Commun. Disord. Technol., Inc., 501 N. Morton St., Sta 215, Bloomington, IN 47404), Doris J. Kistler, Frederic L. Wightman (Heuser Hearing Inst., Louisville, KY 40203), and Jill E. Preminger (Univ. of Louisville, Louisville, KY 40292)

SPATS is evaluated as a testing and training system for hearing-aid (HA) users and cochlear-implant (CI) users. Criterion measures include the HINT, CNC tests, W22 tests, and Coxs CDT, parts of Gatehouses SSQ and a special SPATS inventory. SPATS measures include the identification of syllable constituents (onsets, nuclei, and codas) and measures of top-down and combined top-down and bottom-up recognizing spoken sentences. Control subjects were measured on criterion and SPATS tests, and then retested after a pause of several weeks. Trained subjects took all of the same tests, but in the time between first and second testing underwent either 12 or 24 hours of systematic training using special SPATS algorithms that focus training on items of intermediate difficulty in quiet and noise. Trained subjects show gains on speech-perception measures in quiet and noise, and in look-and-listen tasks, even though there was no training of visual speech perception. Subjects report that SPATS training and testing gave them a much clearer understanding of the severity of their hearing impairments and led to improved speech perception in everyday life through greater attention to detail, and to differences between talkers. [Supported by Grant No. R44DC006338 from NIH/NIDCD.]

4aPP11. The absolute threshold of hearing changes by perceiving a previous sound in the contralateral, ipsilateral, and both ears. Hiroshi Hasegawa, Junji Yoshida, Masao Kasuga (Grad. School of Eng., Utsunomiya Univ., 7-1-2 Yoto, Utsunomiya-shi, Tochigi-ken 321-8585, Japan), and Yu Watanabe (Utsunomiya Univ., Utsunomiya-shi, 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, and then the test sound was presented in either the contralateral or ipsilateral ear at an interval of 0.5 s. When the previous sound was presented first in both ears, the test sound was presented second in one ear (left or right ear). Both the previous and test sounds had the same frequency of 500 Hz, and the same duration of 3 s. The threshold change was obtained from the difference between the thresholds with and without the previous sound. The threshold was found to be decreased by 3.1 dB when the previous sound was presented in the contralateral ear and increased by 0.9 dB in the ipsilateral ear. On the other hand, the threshold was decreased 2.3 dB when the previous sound was presented in both ears.

4aPP12. Examination of bone-conducted sound pathways. Lynn M. Brault, Charissa R. Lansing, Ron D. Chambers (Dept. of Speech and Hearing Sci., UIUC, 901 S. Sixth St., Champaign, IL 61820), and William D. O'Brien (UIUC, Urbana, IL 61801)

Evidence of bone-conducted sound pathways through the skull have been explored with growth of loudness tests, auditory brainstem response, and otoacoustic emissions; however a comprehensive understanding of these pathways has not been achieved. Less is known about the usefulness of behavioral thresholds, sound pressure levels in the ear canal, and the auditory steady-state response (ASSR) as a function of ear canal occlusion in understanding propagation pathways of bone-conducted signals. Results acquired from multiple, repeated test sessions with a single-subject design, revealed lower behavioral thresholds and higher sound pressure level in the ear canal (Real Ear) in the occluded condition below 2,000 Hz; however, this trend was reversed above 2,000 Hz. The ASSR revealed lower thresholds in the occluded condition at all frequencies. Real Ear measured the bone-conducted signal while the behavioral thresholds show the product of the signal after following the traditional bone conduction pathway through the ear structures and the brain. The ASSR results indicate the responses recorded were suprathreshold, and are the product of bone conduction sound excitation at the level of the auditory nerve. Understanding bone-conducted sound pathways could provide guidance for designing enhanced hearing protection, new diagnostic protocols, and novel approaches for signal processing. [Work supported by AFOSR FA9550-06-0128.]

4aPP13. Infants' and adults' sensitivity to a tone masked by remote frequency tones or noise. Lynne A. Werner (Dept. Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246) and Lori J. Leibold (The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599)

It was previously reported that infants and adults exhibit masking of a tone by two tones of constant, remote frequencies. Neither informational nor energetic masking would be expected for adults under these conditions. In the present experiment, infants' and adults' detection of a tone masked by two constant frequency tones or two narrow bands of noise were compared. The target tone was 1,000-Hz, 300-ms duration with 16-ms rise/fall, at a level expected to produce $rmd' = 1$ in the two-tone masker condition for each age group. The maskers were either tones of 581 and 2,920 Hz or 50-Hz-wide noise bands centered at 556 and 2,895 Hz. The duration of the maskers was 300 ms with 16-ms rise/fall. The maskers repeated at 600-ms intervals. The target tone was presented simultaneously with the maskers on signal trials. Sensitivity was measured using an observer-based procedure. Listeners received feedback whenever they correctly detected the target. Infants achieved a higher rmd' in the two-noise-band condition than in the two-tone condition; there was no difference in rmd' between these conditions for adults. Thus, introducing a timbre difference between the target and masker reduced the amount of masking for infants, but not for adults.

4aPP14. Intelligibility differences between male and female bone conducted speech when presented in high noise environments. Maranda McBride, Meghan Hodges, and Jon French (Embry-Riddle Aeronautical Univ., 600 S. Clyde Morris Blvd., Daytona Beach, FL 32114)

Clear and intelligible verbal communication is essential in many organizations. When designing verbal communication devices, developers must not only consider the electrical and mechanical components of the devices themselves, but also various parameters of both the listening environment and vocal signals being transmitted. Bone conduction communication devices have recently been introduced into several tactical organizations; however, the effectiveness of these devices in transmitting vocal signals has not yet been comprehensively explored. This paper describes a study investigating the impact of voice type (male and female), vibrator location (condyle of the jaw and mastoid process), and background noise level (0, 83, 93, and 103 dB(A) SPL) on the intelligibility of bone conducted verbal communication. The fundamental frequencies of the male and female voices were approximately 100 and 185 Hz, respectively. The background noise used in the study was pink noise. Six male and six female students from Embry-Riddle Aeronautical University participated in the study. The results of the study indicate significantly higher speech intelligibility scores for the male voice than for the female voice, and the condyle was significantly more effective than the mastoid in transmitting intelligible speech in each of the high noise environments tested.

4aPP15. Psychoacoustic evaluation of broadband and conventional reversing alarms localization accuracy. Stephen Lakatos (Washington State Univ., Vancouver, 14204 NE Salmon Creek Ave., Vancouver, WA 98686, lakatos@vancouver.wsu.edu)

Three psychoacoustic tests assessed whether a recently developed class of vehicle reversing alarms with broadband signal characteristics would be more accurately localized than conventional equivalents. Vehicle reversing conditions were simulated by mounting sample alarms on a linear slide whose velocity and displacement could be controlled by computer. Reversals of 4 m displacement were recorded at two velocities (1 m/s and 2 m/s) and at azimuthal orientations varying in 50 increments. The first two tests presented direct hit trajectories recorded with a binaural microphone (Crown SASS-P) and Kemar Mannequin, respectively; a third test presented near miss trajectories with the recording centerline perpendicular to the alarm trajectory and at linear displacements of 0.5 m and 1 m. Participants listened to trajectories over supra-aural headphones and selected the closest match from computer depictions of all possible trajectories. Results for all three tests revealed no significant advantage in localization acuity for the broadband alarm.

4aPP16. Using multichannel communication systems to observe selective and divided attention. Misty Gripper (Biomed., Industrial and Human Factors Eng. Dept., Wright State Univ., 3640 Colonel Glenn Hwy, Dayton, OH 45435, misty.gripper@wright.edu), Tomasz Letowski, and Tim Mermagen (U.S. Army Res. Lab., Aberdeen Proving Ground, MD 21005)

The objectives of this study are to compare the effectiveness of bone- and air-conduction radio communication while listening to natural auditory signals through air conduction and to assess the feasibility of using a mixed bone- and air-conduction system for multichannel radio communication. There are many benefits to both air-conducted and bone-conducted systems, but little research on combining the two into one radio system. The goal of the study is to explore this possibility. The SATASK test developed at the U.S. Army Research Laboratory is used to observe human capabilities to distribute attention across multiple channel communication systems. This test consists of sentences structured as follows: NAME write the number NUMBER on the COLOR OBJECT. Five different multichannel communication systems (three three-channel and two four-channel systems) consisting of various combinations of air-conducted earphones, bone-conducted vibrators, and an array of loudspeakers are used for data collection. Each channel presents different sentences in different voices. Listeners are given a target name and they are to record the corresponding information associated with that name (selective attention) or they are to record information from all channels (divided attention). The results of the study will be discussed at the meeting.

4aPP17. Auditory detection of amplitude modulation in psychophysical notched noise task and electroencephalography. Allison I. Shim, Bruce G. Berg, and Ramesh Srinivasan (Dept. of Cognit. Sci., Univ. of California Irvine, 3151 Social Sci. Plaza, Irvine, CA 92697)

Notched tone-in-noise detection tasks have commonly been utilized to investigate the properties of auditory filters. A similar design is used to study filters for auditory temporal processes; amplitude modulated (AM) noise is used as the signal component rather than tone. The signal is composed of a 200 Hz wide noise-band with a 40 Hz modulation rate and centered at 1,000 Hz. Notched-noise maskers composed of 200 Hz unmodulated noise bands located an equal distance logarithmically on either side of the stimulus create the notch. Psychophysical thresholds were measured using a 2IFC adaptive staircase procedure. A common assumption is that thresholds for signal detection are unaffected until the masker impedes upon the edge of a filter. Data were fit using a leaky integrator model with three forms of the bandpass filter. Results support the existence of wide filters necessary for temporal processing. In an effort to integrate psychophysical and brain imaging methods, a similar study was conducted using electroencephalography. Cortical evoked responses at the gamma frequency (40 Hz) also appear to be affected by narrowing notch widths. [Work supported by the Center for Hearing Research, University of California, Irvine.]

4aPP18. Effects of frequency shifts on the identification of vowels and words in sentences. Peter F. Assmann (School of Behavioral and Brain Sci., Univ. of Texas at Dallas, Box 830688, Richardson, TX 75083) and Terrance M. Nearey (Univ. of Alberta, Edmonton, AB, Canada T6E 2G2)

Studies of the effects of frequency shifts on vowel identification have shown a drop in accuracy when the spectrum envelope is shifted up or down, and when the fundamental frequency (F0) is raised or lowered. We have found an interaction between F0 and spectrum envelope shifts: Performance is better for vowels with matched shifts (both F0 and spectrum

envelope shifted in the same direction) compared to mismatched shifts (F0 shifted up and spectrum envelope shifted down or vice versa). The aim of the present study was to determine the extent to which these effects persist in sentence context. The STRAIGHT vocoder was used to process a set of sentences from the HINT test using the same scale factors as in the vowel identification experiment. Word recognition scores generally followed the same pattern as vowel identification, with poorer performance for lowered F0 and raised spectrum envelope, and the lowest scores in conditions with high F0 and downward shifts in spectrum envelope, compared to the corresponding matched shifts. [Work supported by NSF and SSHRC.]

4aPP19. Audio visual interaction in an auditory streaming task involving vowels. Etienne Gaudrain, Nicolas Grimault (Universit Lyon 1, UMR CNRS 5020, 50 av. T. Garnier, 69366 Lyon Cedex 07, France), Fredric Berthommier (Institut de la Commun. Parle, Grenoble, France), Eric Healy (Univ. of South Carolina, Columbia, SC), and Jean-Christophe Bra (Inserm U556, Lyon, France)

Previous studies suggest that sequential segregation mechanisms might be primary in concurrent speech segregation tasks for hearing-impaired listeners. Other studies involving speech materials indicate that timbre cues and pitch cues primarily modulate sequential segregation. Finally, concurrent speech segregation has been reported to be enhanced by visual cues associated with speech reading. In this study, sequences of vowels having alternating fundamental frequencies (low and high) were presented to normal-hearing listeners who were asked to report the presentation order of the vowels. Lower performance in such a task has been related to increased sequential segregation. By directing listeners to the perception of a single stream, this task leads to an estimation of the pitch-based obligatory streaming evoked by the sequence. Simultaneous with the presentation of each high-frequency vowel, visual cues simulating lip movements were provided. Lip movements were either steady (control condition), open-closed (rhythm condition), or congruent with each particular vowel (congruent condition). As such, lip movements were used in this study as a potential distracter. The lower performance (increased streaming) in the congruent condition relative to control conditions suggest that visual cues may interact with sequential obligatory streaming mechanisms at a sensory level.

4aPP20. Effects of a response format modification on the diagnostic rhyme test. Edward L. Goshorn (Dept. of Speech and Hearing Sci., Univ of Southern Mississippi, 118 College Dr. No. 5018, Hattiesburg, MS 39406, edward.goshorn@usm.edu)

Detection theory measures of listener performance, as well as distinctive feature scoring, have been shown to provide useful additional information to percent correct scores when testing word recognition. The diagnostic rhyme test (DRT) uses a two-item multiple choice (MC) response format to provide percent correct and distinctive feature information but does not measure d. This study examined the effects of modifications to the DRT response format that allow a measure of d. The format was modified to a single-option in which the listener designates yes or no (Y-N) that the single-option matches the stimulus. DRT words were mixed in white noise and presented in sound field at 3 and 6 dB signal-to-noise ratios at 50 dBHL to eight normal-hearing subjects. Results showed a significant difference between response formats for percent correct and a significant interaction between signal-to-noise ratios and feature errors across formats. At 3 dB, there were more temporal-based feature errors (sustension and compactness) for the Y-N response format, but no significant differences between formats for voicing, nasality, sibilation, and graveness. At 6 dB, there were significant differences only for sustension and graveness. The ds for signal-to-noise ratios were significantly different for the Y-N format.

4aPP21. Coupled cross-modal activity in the human auditory and visual cortices. William Winter (Univ. of California, Irvine, 8730 Palo Verde, Irvine, CA, 92617)

From the perspective of systems analysis, the reactions of the brain to a steady state stimulus (steady state evoked fields) represent the brain's system response. In a linear system, these responses are expected to occur only at the stimulus frequency. However, in a nonlinear system, power can disperse to other frequencies, usually at integer combinations of the stimulus frequencies. By stimulating different components of a nonlinear system at various frequencies, it is possible to evaluate the strength of coupling between components. In this study, steady state auditory and visual stimuli are presented at various frequencies to human subjects to evaluate the nonlinear coupling between the auditory and visual systems. Subjects in whom strong responses are found will be further evaluated with simultaneous electroencephalography/functional magnetic resonance imaging and diffusion tensor imaging to attempt to determine the extent to which the cross-modal activity is due to field versus fiber propagation.

4aPP22. Evaluation of the consonant-nucleus-consonant word test as spoken by a female talker. Margaritis Fourakis, John Hawks, Heather Zingler, Lyndsey Persak (Dept. of Communicative Disord., UW-Madison, 1975 Willow Dr., Madison, WI 53706), John Hawks (Kent State Univ., Kent, OH 44242), Heather Zingler, and Lyndsey Persak (UW-Madison, Madison, WI 53706)

The most common recording of CNC words used to test speech perception by persons with cochlear implants utilizes a single male talker. However, research has shown that CI recipients often have difficulty perceiving speech by persons with high F0 and formant frequencies (Loizou, P., Dorman, M., and Powell, V. 1998. J. Acoust. Soc. Am. 103: 1141–1149). Presented here are CNC word lists recorded using an adult female talker. Ten normal hearing adults identified these words with 99% accuracy. Eight CI subjects also listened to the original (male voice) and new (female voice) recordings. The two star subjects identified the female speakers words 15% less accurately than the male speakers words (Male voice: 82% correct; Female voice: 67% correct). For the other CI subjects (mean performance around 40%) there was no effect of voice on their performance. Results of the acoustic analysis of all the stimuli will also be presented. [Research supported by a subcontract from Washington University in St. Louis, NIDCD R01 DC000581.]

4aPP23. The perception of auditory-tactile integration. E. Courtenay Wilson, Charlotte M. Reed, and Louis D. Braida (Harvard-MIT Div. of Health Sci. and Technol., Speech and Hearing Bioscience and Technol. Program; Res. Lab. of Electrons, MIT, 36-761, 50 Vassar St, Cambridge, MA 02139)

This project explores the perceptual interactions between certain auditory and tactile stimuli in an objective manner. Our hypothesis states that if the auditory and tactile systems integrate into a common pathway, the *d*-prime measure of the two sensory stimuli presented simultaneously will be significantly greater than the *d*-prime measure of the individual sensory stimuli. If the stimuli are judged independently, the resulting *d*-prime will be close to the root-squared sum of the individual sensory *d*-primes. If the stimuli are integrated into a single percept, the *d*-prime will be close to the sum of the individual *d*-primes. This experiment presents 250 Hz vibratory stimuli to the fingertip and 250 Hz tones binaurally at the subject's detection threshold. Experimental measurements of *d*-prime were obtained for auditory-alone, tactile-alone, and combination of auditory plus tactile. All testing used auditory white noise of 50 dB SPL. Effects of relative tactile-auditory phase and temporal asynchrony of the stimuli are explored. Preliminary data suggest that under certain conditions the two sensory stimuli integrate in a common pathway to form a single percept, that the stimuli must be synchronous for integration to occur, and that relative tactile-auditory phase has no effect on integration. [Hertz Foundation, NIH RO1-DC000126-25, ROI-DC000117.]

Session 4aSA

Structural Acoustics and Vibration: Modeling of Vibration and Radiation in Complex Structures

Courtney B. Burroughs, Chair
1241 Smithfield St., State College, PA 16801

Invited Papers

8:00

4aSA1. Numerical difficulties with boundary element solutions of interior acoustic problems. John Fahnline (The Appl. Res. Lab., The Penn State Univ., University Park, PA 16802)

Although boundary element methods have been applied to interior problems for many years, the numerical difficulties that can occur have not been thoroughly explored. Various authors have reported artificial damping due to acoustic radiation and low-frequency breakdowns. In this presentation, it is shown through a simple example problem that the numerical difficulties depend on the solution formulation. When the boundary conditions are imposed directly, the solution suffers from artificial damping due to sound radiation to the farfield. This difficulty can be alleviated by first computing an impedance matrix or admittance matrix, and then using its reactive component to derive the solution for the acoustic field. Numerical computations are used to demonstrate that this technique eliminates artificial damping, but does not correct errors in the reactive components of the impedance or admittance matrices, which then cause nonexistence/nonuniqueness difficulties at the interior resonance frequencies for hard-wall or pressure release boundary conditions, respectively. It is shown that the admittance formulation is better suited to boundary element computations for interior problems because the resonance frequencies for pressure release boundary conditions do not begin until the smallest dimension of the boundary surface is at least one half of the acoustic wavelength.

8:30

4aSA2. Development of a general statistical energy analysis subsystem formulation using finite element periodic structure theory. Vincent Cotoni, Phil Shorter (ESI Group, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130), and Robin Langley (Univ. of Cambridge, Cambridge CB2 1PZ, UK)

Statistical energy analysis (SEA) represents a field of study in which statistical descriptions of a system are employed in order to simplify the analysis of complicated vibro-acoustic problems. In SEA, a vibro-acoustic system is represented by a collection of subsystems that can receive, store, dissipate, and transmit vibro-acoustic energy. Traditionally, the SEA parameters for a given subsystem are formulated analytically based on consideration of wave propagation through the subsystem. While such analytical algorithms can be readily applied to the majority of systems encountered in practical problems, there are certain types of sections that are difficult to describe using existing analytical formulations. Examples include: Isogrid in launch vehicle fairings, extruded aluminum sections in train floors, and modern corrugated aircraft fuselage constructions. This paper describes the development of a generic SEA subsystem formulation based on the use of finite element (FE) periodic structure theory. A small unit cell of the section is created and computationally efficient algorithms are developed to calculate wave propagation through a large array of such cells. The resulting algorithms are used to calculate the SEA parameters for the section. The approach is described and a number of numerical validation examples are presented.

9:00

4aSA3. The response of a beam-like acoustic sensor to transient sound waves. W. Steve Shepard, Jr. and Bill B.B. Zhang (Dept. of Mech. Eng., The Univ. of Alabama, Box 870276, Tuscaloosa, AL 35487, sshepard@eng.ua.edu)

An analytical approach is presented for studying the response of a beam-like structure being considered to act as an acoustic sensor under the excitation of a transient sound wave. First, the response of a Bernoulli-Euler beam is modeled for a multi-cycle transient wave excitation moving at a constant speed. For verification, the model is compared to other approaches by looking at extremes of the input. Then, to examine sensing requirements for the beams structure, the influence of various parameters on the ability to excite the sensor in the midfrequency region is investigated. The parameters considered include beam length, beam materials, amplitude of the transient pressure excitation, duration of the excitation, and the number of sinusoidal half-cycles present in the transient excitation. Potential effects of the superposition of incident and reflected waves caused by the boundary conditions on the sensors response are discussed. [Work supported by the National Science Foundation.]

9:30

4aSA4. S-matrices for waves in random structures, unitarization of diffuse field waveforms, and mesocopics. Richard Weaver (Dept. of Phys., Univ. of Illinois, Urbana, IL 61801)

Diffuse field methods, such as reverberation room acoustics and statistical energy analysis, predict wave energy flow in complex structures. These methods are applied here to predict the root mean square s-matrix $<s^2(t)>^{1/2}$ for an undamped structure coupled to a single incoming and outgoing channel. Multiplying by a white noise random process produces a candidate s(t). The prediction is statistically exact if the internal scatterings are sufficiently phase incoherent, but it is not unitary and is thus inadmissible in the absence of such scatterings. Here it is shown that it can be made unitary in a minimally invasive manner by filtering with a minimum phase causal filter with modulus $1/|s(\omega)|$. This is illustrated with envelopes $<s^2(t)>$ predicated on a picture of the structure as a single reverberation room, as two coupled rooms, and as a quasi-one dimensional multiply scattering system. The resulting unitarized S matrices are found to exhibit a variety of familiar mesoscopic features and behaviors not present in the original diffuse $<s^2(t)>$. These include enhanced backscatter, quantum echo, power law tails, level repulsion, and Anderson localization. It is remarkable, and of potential practical interest, that quantitative mesoscopic features follow from simple principles. [Work supported by NSF CMS 05-28096.]

9:45

4aSA5. Modeling of statistics of structural elements vibration, induced by near-wall turbulence. Leonid R. Yablonik (Polzunov Central Boiler & Turbine Inst., 3/6 Atamanskaya str., 191167 St.-Petersburg, Russia) and Emmanuil M. Agrest (Johnson & Wales Unv., Charlotte, NC 28202)

The statistical characteristics of elastic elements fluctuations, generated by the dynamic effect of near-wall turbulent pressures field, are analyzed. The probabilistic properties of vibration of elements are determined by the characteristic functional of the random pressure field. A superposition of Gaussian and Poissonian field is used to model turbulent pressure in accordance with [E. B. Kudashev and L. R. Yablonik, Acoust. Phys. **48**(3), 321–324 (2002)]. The Poisson component of turbulent field, which is associated with the processes of the generation of spontaneous splashes in the near-wall zone of flow, has a basic impact on the statistical features of vibrations. The model relationships, which connect flow conditions with the parameters of the characteristic functional of Poisson component, are proposed. Characteristic functions of the amplitudes of vibrations with the different values of the parameters of the acting turbulent field simulations are represented. The statistics of the inertia-free pistons that are usually used in transducers for pressure turbulent field diagnostics is investigated in detail.

10:00–10:15 Break

10:15

4aSA6. A simplified finite-element modeling approach for constrained layer damping. Zhengchao Xie and W. Steve Shepard, Jr. (Dept. of Mech. Eng., The Univ. of Alabama, Box 870276, Tuscaloosa, AL 35487)

In formulating finite-element models for multilayer structures, designers usually use a detailed model that contains many degrees of freedom. In order to reduce the modeling complexity for structures containing constrained layer damping, some researchers have attempted to develop equivalent properties that are applied to regular single-layer finite elements. For example, the analytical formulation developed by Ross, Ungar, and Kerwin is sometimes used to find an equivalent bending stiffness. However, that stiffness is based on an infinite or simply-supported beam, which is often inappropriate. The research presented here uses the governing equations of a regular beam or plate to derive the frequency transfer function between two points on the structure. Curve fitting this transfer function to a reference value found using a more detailed model provides a complex bending stiffness at the first few peak frequencies. The goal is to develop an improved approach for extracting equivalent properties for the simplified modeling of multilayer structures. Numerical examples show that extracted properties can lead to good results when calculating the response of the structure. This approach may provide a means of reducing the modeling requirements for viscoelastically damped structures.

10:30

4aSA7. Dynamics of a hollow rubber-metal laminate structure. Robert W. M. Smith (Penn State Appl. Res. Lab., P.O. Box 30, State College, PA 16804)

Transverse frequencies of vibration of a hollow cylindrical rubber-metal laminate structure, subject to a mean pressure difference across the cylinder wall will be presented. Extending Timoshenko beam theory to model such a structure, we use two independent observations most often credited to J. A. Haringx: First, an analogy exists between an axially loaded Euler beam and a pipe without axial loading but subject to internal pressure, and second, for this type of rubber-metal laminate structure, the transverse component of the axial load acts perpendicular to the bending slope, rather than the deflection slope. When dynamics of the pressurized laminate are considered, theory predicts somewhat unexpected changes in the resonance frequency of transverse bending modes as a function of mean pressure difference across the structure. Experimental results illustrating static buckling behavior will be shown for both positive and negative pressure differentials, corresponding to the analogous axial loading cases of buckling in both compression and tension. In applications where oscillatory pressures are present across the structure, the unusual tuning curve has implications for prediction of parametric instabilities.

10:45

4aSA8. Radiation from a source embedded in a two-layer coating of a spherical shell. Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Analysis of the acoustic field generated by a small source embedded in a viscoelastic layer encapsulating a spherical shell gives rise to several analytical difficulties not encountered in the classical problem of a point source in the surface of a rigid sphere. The present investigation addresses a double layer configuration in which the layer closest to the shell is acoustically soft, whereas the outer layer, in which the source is embedded, has an impedance that matches that of the surrounding fluid, water. The analysis is carried out by treating the layers as dissipative fluids, while classical Love theory describes the interior shell. The analysis employs spherical harmonic expansions for each medium, which is known as Mie series analysis in related studies of electromagnetic scattering from coated series. Like the electromagnetic problem, computational difficulties associated with the occurrence of spherical Bessel functions of complex argument arise, but the remedies used there require modification.

11:00

4aSA9. Frequency spectra of elliptic cylindrical shells. Jeffrey E. Boisvert (NAVSEA Newport, 1176 Howell St., Newport, RI 02841) and Sabih I. Hayek (Penn State Univ., University Park, PA 16802)

The kinetic and strain energy densities were derived for the vibration of an elliptic cylindrical shell of constant thickness. The elastic strain energy density has seven independent kinematic variables: Three displacements, two thickness-shear, and two thickness-stretch. The shell has a constant thickness h, finite length L, and is simply supported at its ends $(z=0, L)$, where z is the axial coordinate. The elliptic cross section is defined by the shape parameter a, and the half-length of the major axis l. Using appropriate comparison functions that satisfy the boundary conditions for a simply supported elliptic cylindrical shell, the system is solved using the Galerkin method. Since the shell has an elliptic cross section, the

frequency spectra are categorized into distinct symmetric and antisymmetric displacement fields whose spectra coalesce as the elliptic cross section approaches a circular cross section. Numerical results are presented for two h/l and L/l ratios, and various shape parameters, including the limiting case of a simply supported cylindrical shell ($a \sim 100$). [Work supported by the ONR/ASEE Summer Faculty Research Program.]

11:15

4aSA10. Tilted aluminum cylinder acoustic scattering properties and holographic and synthetic aperture sonar (SAS) images. Kyungmin Baik, Christopher Dudley, and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

A solid bluntly-truncated aluminum cylinder in water is a convenient target for exploring the relationship between free-field scattering properties of short tone-bursts and acoustic images. These experiments concern situations where the product ka of the acoustic wave number and the cylinder radius exceed about 12. There are several prominent features in the tilt-angle dependence of the time-resolved backscattering. Some of these features are also present in prior measurements for tilted stainless steel cylinders using relatively long tone bursts [K. Gipson and P. L. Marston, J. Acoust. Soc. Am. **106**, 1673–1680 (1999); K. Gipson and P. L. Marston, J. Acoust. Soc. Am. **107**, 112–117 (2000)]. Prominent features include responses caused by guided generalized Rayleigh waves associated with meridonal rays, helical rays, and face crossing rays. These features are easily observed for tilted aluminum cylinders in the scattering of short tone bursts for cylinders having length-to-diameter ratios of 3 and 5. Guided Rayleigh waves also influence the appearance of the imaged cylinders in holographic and synthetic aperture sonar (SAS) images constructed using far-field bistatic data. This influence on the images is explained using ray theory. [Supported by ONR.]

11:30

4aSA11. Scattering from penetrable prolate spheroids. Jeffrey E. Boisvert (NAVSEA Newport, 1176 Howell St., Newport, RI 02841) and A. L. Van Buren (Cary, NC)

The problem of acoustic scattering from a fluid-filled elastic layer (shell) is considered. The inner and outer surfaces of the elastic shell are confocal prolate spheroids with the outer surface having major-axis length L and minor-axis length D. A plane wave ensonifies the shell at an arbitrary angle of incidence. The 3-D equations of elasticity in prolate spheroidal coordinates are solved using a Helmholtz decomposition for the displacement vector. The displacement vector is expressed in terms of a scalar potential (dependent on the dilatational wave number) and a vector potential (dependent on the shear wave number). These wave potentials are solutions, respectively, of the scalar and vector wave equations cast in prolate spheroidal coordinates. An eigenfunction expansion of spheroidal wave functions represents the solution of the scalar wave equation. The solution of the vector wave equation is facilitated using expansions of prolate spheroidal vector wave functions. Nearfield and farfield scattering results are presented as a function of incident angle, spheroid shape L/D, and internal fluid (air-filled or water-filled). Comparisons with the rigid case (impenetrable spheroid) are also presented. [Work supported by the NAVSEA Newport ILIR Program.]

11:45

4aSA12. Benchmark solutions for low-frequency structural acoustic target scattering. Mario Zampolli, Alessandra Tesei, and Finn B. Jensen (NATO Undersea Res. Ctr., Vle. S. Bartolomeo 400, 19126 La Spezia, Italy)

Low-frequency scattering problems, typically in the range up to 10 s of kHz for O(1 m) size targets, can be solved by a variety of numerical techniques, each of which has different tradeoffs. In September of 2006, a Target Strength Modeling Workshop was held at NURC with the purpose of establishing a set of benchmark solutions for problems of significance to the underwater structural acoustics modeling community. Problems for low-frequency scattering from spherical and cylindrical objects located in the water column or proud or buried in the seabed are presented. Solution techniques ranging from fully 3-D finite-element/boundary integral techniques, to Fourier decomposition axisymmetric finite-element techniques and semi-analytical methods have been applied to the problem set. The reported solutions show benchmark quality agreement in most cases, and hence could be used as reference solutions for testing low-frequency structural acoustic modeling tools in the future. The results and techniques presented reach beyond the realm of underwater acoustics, and should be of interest to researchers working in the broader field of structural acoustics.

FRIDAY MORNING, 30 NOVEMBER 2007 NAPOLEON B3, 8:00 A.M. TO 12:00 NOON

Session 4aSC

Speech Communication: Auditory and Somatosensory Feedback in Speech Production I

Anders Lofqvist, Chair
Haskins Laboratories, 300 George St., Ste. 900, New Haven, CT 06511

Chair's Introduction—8:00

Invited Papers

8:05

4aSC1. Psychophysical detection threshold of formant manipulations during speech production. D. W. Purcell (Natl. Ctr. for Audiol., Univ. of Western Ontario, 1201 Western Rd., London, ON, Canada, N6G 1H1) and K. G. Munhall (Queen's Univ., Kingston, ON, Canada, K7L 3N6)

In altered auditory feedback experiments, participants respond rapidly and alter their speech production to compensate for the induced auditory feedback error. This response does not occur with small feedback perturbations, and thus there must be a threshold for compensation by the auditory-vocal feedback system. What is unknown is the psychophysical threshold during speech production (i.e., how large are manipulations that are consciously detectable?). The purpose of this experiment was to compare the threshold for

conscious awareness of perturbation, and the minimal change that induces compensation. A real-time auditory feedback manipulation system was employed in a repeated measures design. The compensation threshold was determined using the change point test during a perturbation ramp of 4 Hz/utterance. In the psychophysical measurement, the size of the feedback manipulation followed a two-alternative forced-choice paradigm. With 17 individuals, the mean psychophysical threshold was 105 Hz (SE: 9 Hz). The mean compensation threshold was 64 Hz (SE: 8 Hz). The significant difference between these thresholds is consistent with the hypothesis that the auditory-vocal feedback system can operate without conscious control and suggests that it may have access to a more sensitive representation of speech formants.

8:30

4aSC2. Suppression of responses to pitch perturbations during Mandarin speech. Charles R Larson, Hanjun Liu (Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, clarson@northwestern.edu), and Yi Xu (Univ. College London, London, UK)

Previous research on the task dependency of voice fundamental frequency (F0) responses to pitch-shifted feedback has shown them to be larger during speech than in a vowel task. The present study investigated how responses to pitch-shift stimuli vary with the intonation of a speech utterance and the timing of the pitch shift stimuli. Mandarin speakers produced a four-word phrase in which the F0 was held relatively steady and then either increased or was held steady on the final word. Pitch shift stimuli (100 cents, 200 ms) were presented 160, 240, or 340 ms after the onset of vocalization. F0 response magnitudes for the 340 ms onset condition were significantly smaller than in the 160 or 240 onset conditions when the F0 level on the final syllable was rising. When the F0 on the final syllable was held constant, there was no difference in response magnitudes between the three onset conditions. The decrease in response magnitude in the 340 ms, F0 rising condition, corresponding to approximately 660 ms before the final syllable, suggests vocal responses can be modulated both by the timing of stimulus presentation and the speech intonation pattern. Possible explanations for this finding will be discussed.

8:55

4aSC3. Role of voice F0 production accuracy and pitch perception ability in modifying compensatory voice F0 responses during sustained vowel productions. Jay Bauer, Marylou Gelfer (Dept. of Commun. Sci. & Disord., Univ. of Wisconsin, Milwaukee, P.O. Box 413, Milwaukee, WI 53201), and James Bashford (Univ. of Wisconsin, Milwaukee, Milwaukee, WI 53201)

The purpose of the present study is to assess whether subject characteristics in terms of voice F0 production accuracy and pitch perception ability are related to the ability to compensate in real-time for altered auditory feedback. To assess the relationship of vocal production, pitch-perception, and voice F0 error-correction, a series of three tasks were completed by typical college-age participants: (1) A pitch-matching task where participants vocally matched the pitch of audible complex tones, (2) two-tone forced choice pitch-discrimination tasks where participants decided if complex tone pairs (200 ms duration for each tone) sounded the same or different, and (3) pitch-shifting tasks where participants produced sustained vowel phonations, while listening to experimentally altered auditory feedback (brief 200 ms pitch perturbations of $+/-20$, 60, or 100 cents were introduced into the feedback). Dependent measures of pitch-matching accuracy, pitch-discrimination sensitivity, and variations in the latency, stability, and magnitude of compensatory pitch-shift reflexes were assessed in all participants. Pilot data ($N=3$) indicate that production accuracy and discrimination sensitivity does not play a role in compensatory control of the voice. Poor pitch-matchers and/or discriminators actually respond to very small magnitude perturbations (20 cents) in a manner similar to accurate pitch-matchers and/or perceivers.

9:20

4aSC4. Modification of acoustic-vocal mappings for fundamental frequency control in singers and nonsingers. Jeffery A. Jones and Dwayne Keough (Ctr. for Cognit. Neurosci. & Dept. of Psych., Wilfrid Laurier Univ., Waterloo, ON N2L 3C5, Canada, jjones@wlu.ca)

Singing requires accurate control of the fundamental frequency (F0) of the voice. Previous work has demonstrated that F0 control involves an interplay between closed- and open-loop control. A frequency-altered auditory feedback (FAF) paradigm was used to examine the acoustic-motor representation of the mapping between F0 feedback and the vocal production system in singers and nonsingers. Participants sang a note while hearing their F0 shifted down one semitone. Both singers and nonsingers compensated for the F0 perturbations by increasing their F0. However, the magnitude of compensations was initially smaller in singers than nonsingers. Moreover, after participants heard their feedback returned to normal, the aftereffects observed for singers were larger than those observed for nonsingers. This pattern of aftereffects generalized to the production of another note different than the one participants produced while hearing FAF. Combined, these observations suggest that singers rely more on an internal model for F0 production during singing than nonsingers.

9:45

4aSC5. Effect of anticipation on voice F0 responses to auditory feedback pitch shifts. Theresa A. Burnett, Elise K. Hindmarsh, and Katie E. McCurdy (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Reflex-like voice F0 responses to anticipated and unanticipated auditory feedback pitch shifts were compared in healthy adults. Participants pressed a button to self-trigger a 100-cent, 100-ms upward shift in their voice auditory feedback over headphones while sustaining the vowel /u/ at a comfortable pitch and loudness. In separate blocks of 30 trials each, pitch shifts were self-triggered either immediately upon button press (Immediate) or after a random delay of 500 to 2000 ms (Random). In both conditions, participants were instructed to ignore the shifts as much as possible. To study any interaction between volitional voice F0 control and anticipation of the pitch shifts, the Immediate and Random conditions were repeated with pitch shifts that were long enough in duration for participants to volitionally oppose them (500 ms). Voice F0 response latency was significantly shorter when pitch shifts were

immediately triggered than when triggered after a random delay, but there was no interaction between stimulus duration and stimulus onset timing for voice F0 response latency, magnitude, or duration. Participants neither suppressed the voice F0 response nor amplified it to completely compensate for anticipated 500-ms pitch shifts. Similarities between the voice F0 response and the long-latency stretch reflex are discussed.

10:10–10:30 Break

10:30

4aSC6. The specificity of sensorimotor learning: Generalization in auditory feedback adaptation. K.G. Munhall, E.J.S. Pile, E.N. MacDonald (Dept. of Psych., Queen's Univ., Kingston, Canada), H.R. Dajani (Univ. of Ottawa, Ottawa, Canada), and D. W. Purcell (Univ. of Western Ontario, London, Canada)

An enduring question about sensorimotor learning is how specific the acquired input-output relationship is. In this presentation, we review a series of studies in which the stimuli and conditions during speech motor learning were manipulated to study when and if generalization occurs. In our studies, the first and second formants of vowels were shifted using a real-time signal processing system; when subjects spoke one vowel, they heard themselves saying another vowel. In response to this auditory feedback perturbation, talkers spontaneously compensated by producing formants in the opposite direction in frequency to the perturbation. These compensations persisted after feedback was returned to normal, indicating that a form of sensorimotor learning had taken place. When different vowels were tested following the perturbation of one vowel, they were found not to show evidence of the feedback perturbation nor did the new vowels have any influence on the perturbed vowels' return to normal baseline levels. Data from studies in which listening versus producing were compared and studies in which the similarity of the feedback voice quality was manipulated will also be presented. In general, the studies suggest that learning is quite local and, thus, that learning does not generalize beyond restricted conditions.

10:55

4aSC7. Speech motor learning without hearing. Sazzad Nasir and David Ostry (McGill Univ., 1205 Dr. Penfiled, Montreal, QC, Haskins Labs., New Haven, CT, nasir@motion.psych.mcgill.ca)

The nervous system receives both auditory and somatosensory information as we talk. Little is known about the role of somatosensory input in speech motor control. It is indeed a longstanding puzzle how post-lingually deaf speakers, in the absence of any substantial auditory input, maintain speech intelligibility. In this study we have assessed the role of somatosensory feedback in subjects with cochlear implants who were tested with their implants turned off. We used a robotic device to alter speech movements, and hence somatosensory feedback, by delivering mechanical loads to the jaw. We found that with training subjects corrected for the loads, such that the motion path approached that normally experienced under no-load conditions. Thus even under no auditory feedback, somatosensory input mediates speech movements in post-lingually deaf adults. In this study all five subjects with cochlear implants showed varying degrees of adaptation compared to only four out of six subjects in an age matched control group. This suggests a prominent role for somatosensory feedback in individuals with hearing loss. Finally, movement kinematic patterns for subjects with hearing loss were no different than those of subjects with normal hearing, which further underscores the reliance on somatosensory feedback in deaf adults.

11:20

4aSC8. Motor and sensory adaptation following auditory perturbation of /s/ production. Douglas M. Shiller, Marc Sato, Vincent L. Gracco, and Shari R. Baum (School of Commun. Sci. & Disord., McGill Univ., 1266 Pine Ave. W., Montreal, QC, H3G 1A8 Canada, doug.shiller@mail.mcgill.ca)

Studies of motor adaptation have provided valuable insights into the role of auditory and orosensory feedback in speech production. While perturbations to the structure of the vocal tract (e.g., pseudopalate), which alter both orosensory and auditory feedback, have been used to investigate the production of a range of consonants and vowels, studies of short-term speech adaptation utilizing purely auditory feedback manipulations (e.g., spectral shifting) have been mostly limited to investigations of vowel production. These studies have revealed that motor plans underlying vowel production are flexible, and that subjects will readily alter their motor output in order to achieve auditory perceptual targets. In the present study, we have extended this work by spectrally shifting the acoustic signal for the sibilant /s/ during the production of short /s/-initial words. Evidence of motor adaptation was observed following a brief, intensive period of practice under altered auditory conditions, indicating a role for auditory feedback in sibilant production. In addition to changes in speech motor output, subjects showed changes to their auditory-sensory representations of /s/ and /ʃ/ following the manipulation, suggesting that adaptation was not limited to the motor domain. These results will be discussed in the context of current models of speech motor control.

11:45–12:00
Discussion

Session 4aSP

Signal Processing in Acoustics, Acoustical Oceanography, and Underwater Acoustics: Session Honoring Leon Sibul I

R. Lee Culver, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Chair's Introduction—8:00

Invited Papers

8:15

4aSP1. A turbo approach to distributed detection and estimation. Michael Roan (Virginia Polytechnic Inst. and State Univ., Blacksburg, VA 24060)

A framework, based on the turbo principle, for distributed detection and estimation (DDE) using data from multiple sonar sensors is developed. Through the use of a Bayesian framework for DDE, an ideal observer is derived. This distributed observer makes optimal decisions, interpretations, or inferences based on measurements made at multiple sensors that observe the same scene. In this approach, a posteriori estimates at one sensor become prior probabilities for the next sensor. The posterior probability is proportional to the product of the probability of each possible state of the environment before observation (prior density) and the probability of the measurements given the state of the environment. It is the posteriors and the likelihoods representing knowledge about the environment that are shared among sensors. This iterative approach is mathematically similar to turbo coding in communications applications. This talk provides the background on the turbo DDE engine, and approaches for both practical implementation and performance quantification. Examples using Gaussian assumptions on the signal and noise distributions are given using simulations. The optimum approach is shown to be very good under ideal conditions, but not robust to sensor positioning errors. Sub-optimum approaches to improve robustness are introduced.

8:35

4aSP2. The impact of acoustic signal models on time-difference-of-arrival/frequency-difference-of-arrival estimation. Mark L. Fowler (Dept. of Elec. and Computer Eng., Binghamton Univ., Binghamton, NY 13902, mfowler@binghamton.edu) and Xi Hu (Binghamton Univ., Binghamton, NY 13902)

Much research has been done in the area of estimating time-difference-of-arrival (TDOA) and frequency-difference-of-arrival (FDOA) and their use in locating a radiating source. Early work in this area focused on locating acoustic sources using passive sonar processing. Only later was TDOA/FDOA-based location considered for the case of passively locating electromagnetic sources. As a result of this, it is tempting to use results derived for the acoustic case when answering questions about the electromagnetic case. This paper shows that such borrowing can lead to incorrect results. The key factor that drives the significant differences between these two cases is the difference between the signal model assumptions for the two cases: WSS Gaussian process is widely used as a valid model in the acoustic case, but is not usually appropriate in the electromagnetic case. Although the received signal equations may look identical (showing delay and Doppler shift), the resulting Fisher information, Cramer-Rao bound, and maximum likelihood estimator (MLE) are fundamentally different for the two signal scenarios. Thus, we show that there are distinct structures in processing and bounds that arise specifically due to using models appropriate for acoustic signals.

8:55

4aSP3. Model-based array signal processing: A tribute to Dr. Leon H. Sibul. James V. Candy (P.O. Box 808, L-156, Univ. of California, Lawrence Livermore Natl. Lab., Livermore, CA 94551) and Edmund J. Sullivan (EJS Consultants, Portsmouth, RI 02871)

Dr. Leon Sibul's contributions to acoustical signal processing are extensive and highly significant. They are marked by a deep intuition and foresight, leading to new and fruitful concepts, especially in underwater acoustic processing. In this paper we discuss how Dr. Sibul's contributions to adaptive array processing provided the conceptual basis for many of the modern approaches of model-based processing in ocean acoustics. In particular, we show how they were extended to incorporate sophisticated acoustical propagation models. Specifically, we discuss how his incorporation of propagation models in adaptive array processing schemes provided a new paradigm for such problems as source detection and localization. His concepts of an underlying physics-based estimation structure directly influenced efforts to develop environmentally adaptive processors that are capable of providing solutions to detection and localization problems. In this paper, we concentrate specifically on the model-based processing approach to the detection and localization problem, demonstrating how this area has progressed.

9:15

4aSP4. A local particle-like approach to wave propagation. Leon Cohen (City Univ. of New York, 695 Park Ave., New York, NY 10021)

Waves and particles are the main constituents of nature, but waves are our main method of communication, whether biological or technological. Historically, the study of waves has been separated into the stationary (standing wave) and nonstationary (pulse) case. Pulses are of fundamental consideration in radar, sonar, acoustics, fiber optics, among many other areas. A pulse has been called by many terms—transient, wave group, progressive wave, wave packet, nonrecurrent wave, traveling wave, nonperiodic wave—all to describe basically the same thing. The fact that so many names exist is an indication that there is no full understanding of their behavior. We present a simple particle-like view of pulse propagation, which makes pulses easier to understand and deal with intuitively. We have derived exact expressions for many physical quantities such as the spreading of a pulse, the conditions as to when a pulse contracts, the contraction time, and other important physical quantities. What emerges from these considerations is not only a clearer view of how pulses propagate, but a practical means to obtain their propagation in a dispersive medium.

Contributed Papers

9:35

4aSP5. Propagation-invariant classification of sonar signals. Patrick J. Loughlin and Greg Okopal (Dept. of Elec. and Comput. Eng., Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261)

Underwater sound waves can undergo frequency-dependent effects, such as dispersion and damping, that cause the wave to change as it propagates. In active sonar, these changes can be detrimental to automatic classification because the observed backscatter depends on the propagation environment and how far the wave has traveled. We aim to develop signal processing techniques to enable propagation-invariant classification. Toward that end, we have developed a local phase space view of propagation with dispersion and damping that is physically illuminating and suggests new features for classification. In this talk, we will present moment-like features extracted from the backscattered wave that are invariant to dispersion and damping, per mode. Environmental knowledge is not needed, beyond knowing the general form of the damping. Two forms of damping are considered. Classification results on simulations of the backscatter from different steel shells propagating in a Pekeris waveguide with damping and random variations will be presented. A dispersion-invariant matched-filter receiver approach will also be presented. [Supported by ONR Grant No. N00014-06-1-0009.]

9:50

4aSP6. Improving spatial resolution of a time-reversed focus using frequency-domain signal processing. Brian E. Anderson, Michele Griffa, T. J. Ulrich, Paul A. Johnson (Geophys. Group, MS D443, Los Alamos Natl. Lab., Los Alamos, NM 87545, bea@lanl.gov), and Robert A. Guyer (Univ. of Massachusetts, Amherst, MA 07030)

In the typical application of time reversed acoustics, the spatial resolution of a time-reversed focus is set by the diffraction limit, i.e., a focal spot size cannot be made smaller than a half of a wavelength. A frequency-domain signal processing method is introduced in which the information in the received signal of the forward propagation stage is shifted to higher frequency. The resulting time domain signal is time reversed and rebroadcast to form a focal spot size that is diffraction limited at the higher frequency (shorter) wavelength. This method requires neither a priori information about the original source location nor nearfield signal detection. A theoretical discussion of the problem, along with numerical and experimental results, will be presented. The limitations and potential applications will also be presented. [This work is supported by Institutional Support (LDRD) at Los Alamos National Laboratory.]

10:05–10:20 Break

10:20

4aSP7. Selective source reduction to identify masked sources using time-reversed acoustics. Brian E. Anderson, T. J. Ulrich, Michele Griffa, Paul A. Johnson (Geophys. Group, MS D443, Los Alamos Natl. Lab., Los Alamos, NM 87545, bea@lanl.gov), Marco Scalerandi, and Antonio S. Gliozzi (CNISM, Politecnico di Torino, 10129 Torino, Italy)

This paper describes a time-reversed acoustics (TRA) method of spatially illuminating a source signal, which has been masked by another source signal. The selective source reduction (SSR-TRA) method employs a subtraction technique where one focus is selectively reduced to illuminate the masked focus. In this paper, numerical and experimental results are presented to demonstrate to what degree the SSR-TRA method is successful. The SSR-TRA method is demonstrated for two elastic wave pulses emitted simultaneously from two spatially separated sources of differing amplitudes. Results are presented from experiments conducted with two different solid samples: Aluminum and doped silica glass. Applying the SSR-TRA method, a stronger source, up to 13 times stronger than a weaker one, may be reduced to reveal information about the weaker source. Spatial and/or temporal characteristics of multiple close proximity sources can be resolved with the use of the SSR-TRA method. The results show that the SSR-TRA methods' limitations are chiefly due to imperfect reconstruction of the source function in the time-reversed focal signal. [This work was supported by Institutional Support (LDRD) at Los Alamos National Laboratory.]

10:35

4aSP8. Identifying and integrating likelihood functions and signal parameter probability density functions (PDFs) for sonar signal processing. Richard Lee Culver, Colin W. Jemmott, and Jeffrey A. Ballard (ARL and Grad. Program in Acoust., Penn State Univ., PO Box 30, State College, PA 16804)

A likelihood ratio receiver developed using an estimator-correlator structure requires that the noise probability density function (pdf) belong to the exponential class, but need not be Gaussian. The receiver employs a composite hypothesis that takes into account knowledge (or predictions) of the signal parameter statistics. In order to implement the receiver, we must be able to integrate the a priori pdf for the random parameter times the likelihood function over the domain of the random parameter, i.e., $\int p(r-\theta)p(\theta)d\theta$. The signal fluctuation models used by Swerling (1960) for radar detection theory belong to the chi-squared family of distributions, which are part of the exponential class. Likewise, the maximum entropy method with moments $\ln\theta$ and θ could be used to fit $p(r-\theta)$ with a chi-squared pdf. If both noise and signal parameters are chi-squared distributed, a closed-form solution may exist. This paper considers ap-

proaches to fitting pdfs to noise and signal parameter samples and to solving the integral described above. [Work sponsored by ONR Undersea Signal Processing.]

10:50

4aSP9. A coded signal technique for the study of environmental noise propagation. Jamie A. Angus, David C. Waddington, and Phil J. Duncan (Acoust. Res. Ctr., School of Computing, Sci. and Eng., Univ. of Salford, Salford M5 4WT, UK)

This paper describes the design of a coded acoustic signal for the study of environmental sound propagation. An increasing awareness of environmental noise problems, and the need to better specify their needs and extent, means that a better knowledge of the effects of weather on sound propagation is necessary. Low signal-to-noise ratio is a common limitation in outdoor propagation investigations. A further problem is time variance. A coded acoustic signal approach consisting of a frequency carrier biphase-modulated by a specially-designed pseudo-random code sequence is proposed to overcome these limitations. This so-called ''inner and outer'' code sequence is specially designed for environmental sound propagation investigations, combining simultaneous fine time resolution and large range ambiguity. Experiment measurements are used to illustrate how the amplitude of the transfer functions between the receivers together with accurate times of flight may be calculated from the cross-correlation of the measured coded acoustic signal. It is further shown that the mean effect over the total propagation channel of factors such as wind, temperature gradients, turbulence, barrier, and ground reflections can be determined by measurement of the time-varying propagation transfer function using the proposed complex impulse response correlation technique. [Work supported by EPSRC UK.]

11:05

4aSP10. Performance classification degradation due to shallow-water propagation. Charles F. Gaumond and Ralph N. Baer (Acoust. Div., Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

The propagation-induced degradation of classification ability for echoes from a finite cylindrical shell and a from a fish school is investigated by use of the class specific method [P.M. Baggenstoss, ''Class-specific feature sets in classification,'' IEEE Trans. Sig. Proc. **47**, 3428–3432 (1999)]. The study is performed with previously displayed numerically simulated data. In prior work, degradation from additive noise was studied at a variety of signal bandwidths. Here, convolution noise caused by two-way propagation in a shallow water environment is shown to increase the duration of non-propagated echoes. The common propagation path between the two classes causes degradation when the classifier is based on non-propagated training data. Ameliorating effects of using training data comprised of either a fixed range or multiple distinct ranges are shown. The performance is quantitatively displayed as either classifier operating characteristic (CLOC) curves or as the integrals of the CLOC curves. The design trade-off space of bandwidth, signal-to-noise ratio (proportional to source level), and complexity of the training set is shown over a limited range of values. [Work supported by ONR.]

11:20

4aSP11. Accuracy enhancement of underwater target detection with time-frequency analysis techniques. Abdalla Osman, Naser El-Sheimy (Mobile Multi-Sensor Systems Res. Group, Dept. of Geomatics Eng., Univ. of Calgary, Calgary, AB, Canada, amosman@ucalgary.ca), Aboelmagd Nourledin (Royal Military Coll. of Canada, Kingston, ON, K7K 7B4), Jim Theriaul, and Scott Campbell (Dept. of Natl. Defense, Canada)

Underwater target detection is mainly based on acoustic emissions generated by the target of interest (TOI) as it propels itself through the ocean, and operates non-propulsion-related onboard systems. The spectral signature of these acoustic emissions is used for identification and localization of TOI. This paper focuses on underwater target detection using passive sonobuoys. Discrete Fourier transform (DFT) is used in most sonobuoy processing systems to provide spectral analysis of the received signals. Due to the relatively high noise level and the several sources of interference that may exist underwater, the DFT may not determine the spectral signature of TOI with adequate accuracy. The low signal-to-noise ratio (SNR) and the presence of strong interference sources with frequencies close to the TOI frequency jeopardize the detection accuracy, the bearing estimation performance, and target tracking capabilities. The aim of this paper is to: (1) examine the performance of both wavelet packet analysis and high resolution spectral estimation techniques, and (2) provide a comparative study between both methods. Based on underwater acoustic simulation developed in this research, results showed that the proposed methods can achieve robust target detection with low levels of SNR and interferences of nearby signatures that cannot be detected by DFT.

11:35

4aSP12. Cued beamforming of passive arrays in the context of active-passive data fusion. T. W. Yudichak and Minh Le (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Active-passive data fusion seeks to exploit the complementary information measured simultaneously by active and passive sonar sensors: For instance, active sensors can provide good range estimates for a contact, while passive sensors can provide information on both the bearing of and the spectrum radiated by that contact. One aspect of this subject is the use of feedback, in which prior information is collected about the state of a contact in the data fusion framework, and then exploited to enhance the signal processing of data measured in that framework. Cued beamforming is one such method, in which a dense set of passive beams are steered toward regions of high contact probability, as determined by the result of the fusion of initial active and initial passive state estimates. The denser cued beams may then provide a refinement of the position estimate of the contact, as well as better isolation of its spectral signature. Comparisons of the effect of using cued beams are compared to the effect of using standard beams (those used without prior knowledge) on the state estimate of contacts in various scenarios using simulated data. [Work supported by ONR.]

Session 4aUW

Underwater Acoustics: Underwater Reverberation Measurements and Modeling I

John S. Perkins, Cochair
Naval Research Laboratory, Code 7140, Washington, DC 20375-5350

Eric I. Thorsos, Cochair
Univ. of washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6606

Chair's Introduction—8:00

Contributed Papers

8:05

4aUW1. Overview of the reverberation modeling workshops. John S. Perkins (Naval Res. Lab., Washington, DC 20375) and Eric I. Thorsos (Univ. of Washington, Seattle, WA 98105)

Two workshops on reverberation modeling are being conducted under joint sponsorship from the Program Executive Office C4I, PMW 180 (as funded by the Office of the Oceanographer of the Navy) and the Office of Naval Research. The overall goal of these workshops is to evaluate recent progress made in reverberation-related research efforts and to make recommendations for further development. The first workshop was held in November 2006 and the second is scheduled for March 2008. The focus of the first workshop was on reverberation from the environment, while the second workshop will emphasize system characteristics. At the first workshop, fifteen different reverberation models were represented and extensive comparisons were carried out before, during, and after the workshop. We present the approach used to conduct the first workshop, discuss issues that have been identified, and outline tentative plans for the second workshop. [Work supported by ONR and PMW 180.]

8:20

4aUW2. Scattering models for Reverberation Modeling Workshop I problems. Eric I. Thorsos (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, eit@apl.washington.edu)

For many of the problems posed for Reverberation Modeling Workshop I, the boundary roughness was presented in terms of roughness spectra. Models and/or tables were also supplied for the bistatic scattering cross section and for the coherent reflection loss. Integral equation simulations were used to verify the accuracy of the bistatic scattering cross sections, and rough boundary PE simulations were used to verify the accuracy of the coherent reflection losses. The results of this work will be briefly summarized. The issue also arises for reverberation simulations whether the coherent reflection loss is appropriate for the boundary loss, or whether some other loss model is more appropriate, or whether the loss due to roughness should be simply ignored. This issue will be discussed in light of rough boundary PE simulations. [Work supported by ONR.]

8:35

4aUW3. Reverberation Modeling Workshop problem definitions. Kevin D. LePage (Naval Res. Lab, 4555 Overlook Ave. SW, Washington, DC 20375)

In November of 2006 the first Reverberation Modeling Workshop was held in Austin, Texas. In this talk, the problems defined for the workshop are presented, along with some selected collective results to give an idea of the level of agreement achieved between the participants. In all, 20 problems were defined, ranging from monostatic 2-dimensional range independent problems up through fully three-dimensional bistatic problems. [Work supported by ONR and PEO C41 and Space (PMW 180).]

8:50

4aUW4. Finite-element modeling of long range, range-dependent acoustic propagation in shallow water. Marcia J. Isakson, R. Abe Yarbrough (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029), and Preston S. Wilson (Univ. of Texas at Austin, Austin, TX 78712-2300)

The current methods of modeling long range, low frequency acoustic propagation in shallow water environments include ray methods, wave number integration, parabolic equations, and normal mode theory. Finite-element (FE) models have also been considered in the past, but computing power and algorithm cost have limited their application. Recent advances in computing power coupled with the advent of low cost, user-friendly, all-purpose finite-element codes have provided new opportunities for the application of FE modeling. We present both time and frequency domain solutions to canonical acoustic problems (scattering from a corrugated surface and propagation in a Pekeris waveguide), and to two-way, range-dependent problems, such as waveguides with surface and bottom roughness. For the latter environment, reverberation time series produced by Fourier synthesis of the FE model output can potentially provide benchmark solutions that no other currently available solution method can provide. [Work supported by ARL IR&D.]

9:05

4aUW5. Parabolic equation techniques for reverberation modeling. Joseph F. Lingevitch, David M. Fromm, and Kevin D. LePage (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Recent developments of parabolic equations for modeling reverberation will be presented. This method is applicable to the 2-D reverberation problems with general sound speed profiles, sediment layering, and interface roughness. The environment is sub-divided into range-independent regions, and the forward and back-scattered fields are coupled at each interface with a single-scattering approximation. This method incorporates a cumulative scattering loss into the forward field, but neglects the non-local coupling terms between the forward and backward scattered fields. The back-scattered field is stored at each interface on a forward sweep through the environment and back-propagated on backward sweep. Reverberation time series are Fourier synthesized from the frequency domain solutions. This technique is fundamentally different than scattering kernel based approaches and provides an independent method for obtaining the forward and back-scattered solutions for reverberation problems. As currently implemented, this model does not include multiple scattering effects, but the algorithm may be generalized to include multiple scattering with iterative sweeps through the propagation domain. The method applied to 2-D reverberation problems from the recent Reverberation Modeling Workshop (November 2006) and results are compared with ray and normal-mode based models. [Work supported by ONR.]

9:20

4aUW6. A mostly time-domain reverberation model with application to Reverb Workshop Problem II. Richard Keiffer (Naval Res. Lab., Stennis Space Ctr., MS 39529, keiffer@nrlssc.navy.mil)

A mostly time-domain reverberation model is described and applied in an isospeed shallow water ocean waveguide that has a flat sandy bottom and a rough air/sea boundary (Reverb Workshop Problem II) to predict the received intensity averaged over many realizations of the sea surface. The model described computes the scattering using a time-domain scattering approach called the wedge assemblage method and accounts for volume attenuation and bottom reflection loss under a ''constant Q'' approximation. Details of the model are discussed and results are presented for both the point and line source excitation of the waveguide. [This work was supported by ONR through NRL.]

9:35

4aUW7. An investigation into the validity of the Born approximation in reverberation calculations. David P. Knobles, Steven A. Stotts, Marcia J. Isakson, and Preston S. Wilson (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713)

Ocean reverberation models for active sonar applications assume the validity of the Born or single scatter approximation. In many areas of physics, a Born series does not converge due to a combination of the scattering strength and the spatial scale over which the scattering occurs. In these cases, the validity of the Born approximation is at best questionable. It is natural to attempt to examine the validity of the Born approximation and the convergence of the Born series in general for well-defined scattering problems in ocean acoustics. The problems examined in this study involve scattering from specific realizations of bottom roughness for 2-D waveguides. This study examines several idealized problems from the 1st Reverberation Workshop in November 2006 for the purpose of comparing a Born approximation solution to an exact solution provided by a 2-way coupled-mode integral equation approach. The coupled integral equations are solved exactly to within the limitations of finite differences and numerical quadrature. The exact solutions are then compared to those that solve the fundamental integral equation using the Born approximation and to solutions obtained with a finite-element model. [Work supported by ARL:UT IR&D.]

9:50

4aUW8. Rayleigh reflection loss non-linearity approximation and application to reverberation. Chris Harrison (NURC, Viale S. Bartolomeo, 400, 19126 La Spezia, Italy, harrison@nurc.nato.int)

This paper presents an extremely good approximation to the Rayleigh reflection loss for two half-spaces composed of water over sediment. This has some useful properties. First it demonstrates that the usual linear approximation for the reflection loss (inside the critical angle) can be extended to the nonlinear case. Second it shows that the non-linearity can be expressed as a separate function that multiplies the linear loss coefficient. Third this nonlinearity term depends only on sediment density and does not depend on sediment sound speed or volume absorption. Fourth the nonlinearity term tends to unity, i.e., the reflection loss becomes effectively linear, when the density ratio is about 1.27. There are similar findings for the reflection loss phase which lead to the well-known lateral shift or effective depth. A class of closed-form reverberation (and signal-to-reverberation) expressions has been developed at NURC, [Harrison (2003), JASA, 114, 2744–2756; Harrison (2005), JCA, 13, 317–340; Harrison (2005), JOE, 30, 660–675]. The new approximation enables one to convert these reverberation expressions from simple linear loss to much more general reflecting environments. Correction curves are calculated in terms of sediment density and applied to a test case taken from a recent ONR Reverberation Workshop (Nov 2006).

10:05–10:20 Break

10:20

4aUW9. A comparison of ray, normal-mode, and energy flux results for reverberation in a Pekeris waveguide. Dale D. Ellis (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS Canada B2Y 3Z7, dale.ellis@drdc-rddc.gc.ca), M. A. Ainslie (The Hague, The Netherlands), and C. H. Harrison (NATO Underwater Res. Ctr., La Spezia, Italy)

A number of problems were developed for, and presented at, a 2006 Reverberation Modeling Workshop sponsored by the US Office of Naval Research. The simplest of these known to the participants as Problem 11, was a Pekeris waveguide (isospeed water over a flat bottom half-space) with Lambert bottom scattering. The water depth was 100 m and frequencies of 250, 1,000, and 3,500 Hz were specified. A number of source-receiver combinations were specified, but the reverberation predictions are quite insensitive to sensor depth except at 250 Hz. With some benefit from hindsight and the results from other models, we compare our results from ray, normal-mode, and energy-flux approaches. All three approaches agree at intermediate times, say 3 to 50 s. At short times, the steep-angle paths and fathometer returns cause the mode and energy-flux models to underpredict the reverberation. At longer times, the ray models run out of steam: i.e., there are too many contributing ray paths for them to handle so they underpredict the reverberation. By combining the model predictions together with analytical results from an energy-flux model, we propose a composite benchmark solution. [Work supported in part by ONR.]

10:35

4aUW10. Issues associated with solutions to a rough-bottom Pekeris waveguide. Steven A. Stotts, David P. Knobles, and Robert A. Koch (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758)

The solution to reverberation time series in a 2-D shallow-water ocean waveguide with bottom roughness (c.f. Problem 1, Reverberation Workshop 2006) has been obtained using both plane wave Born approximation (PWBA) and two-way coupled-mode (CM) approaches. The source and receiver are range monostatic. The CM solutions obtained can be used to determine the accuracy of the Born approximation. Specified time intervals of the received field from the CM solution are correlated and compared to both the patch size and bottom roughness correlation length in the PWBA solution, thus connecting the two approaches. The minimum number of bottom interactions required for accuracy in the PWBA approach is determined. Propagation effects associated with time bin averaging in the PWBA approach are also discussed. The importance of including continuum components in the CM approach, via leaky modes evaluated in the complex horizontal wave-number plane, is also described. Convergence of the CM solution in an expansion of an orthogonal vector space is determined. [Work supported by ARL:UT IR&D.]

10:50

4aUW11. A Matlab and normal mode based reverberation model and some data/model comparisons. John R. Preston (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804) and Dale D. Ellis (DRDC Atlantic, Dartmouth, NS, B2Y 3Z7 Canada)

A Matlab and normal mode based reverberation model has been developed that uses Ellis' algorithm [J. Acoust. Soc. Am. **97**, 2804–2814] for reverberation predictions and the ORCA normal mode model [Westwood et al., J. Acoust. Soc. Am. **100**, 3631–3645] to compute the eigenvalues and mode functions. The model is currently range independent, but handles bistatic geometries, and towed array beam patterns. Group velocity corrections similar to LePage's [J. Acoust. Soc. Am. **106**, 3240–3254] have recently been added to this model. The matrix formulation in Matlab allows the bistatic calculations to be performed more efficiently. An overview of the model is presented that includes sample model-model comparisons for some problems from the 2006 ONR Reverberation Workshop. The model predictions are also compared with data obtained with the NURC (formerly the NATO Undersea Research Centre, SACLANTCEN)

towed arrays during several reverberation experiments on the Malta Plateau. Select data from these experiments are used to compare with model predictions using best estimates of bottom properties in the area. [Work supported by US Office of Naval Research, Code 321OA, Grant No. N00014-05-1-0156.]

11:05

4aUW12. The importance of retaining inter-path coherence in reverberation prediction. Kevin D. LePage (Naval Res. Lab, Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

The R-SNAP and BiStaR normal mode-based reverberation models are used to solve a selection of problems from the 1st Reverberation Modeling Workshop held in Austin, TX in November of 2006. R-SNAP and BiStaR are codes that include the effects of inter-mode decorrelation to predict the coherent aspects of reverberation, as well as allowing for the option to compute the ''incoherent'' reverberation. Comparisons between the incoherent R-SNAP and BiStaR predictions and model predictions obtained by other methods show good agreement. However, it is argued here that as defined, the problems require a coherent solution, and differences between the incoherent and coherent solutions can be significant. Comparisons with coherent 2-way PE results for problem I computed by Joe Lingevitch (NRL) show that coherent structure in reverberation exists, is predictable, and should therefore be an integral part of any reverberation calculation. [Work supported by ONR.]

11:20

4aUW13. Propagation modeling via ray bundles. Gary H. Brooke (General Dynam. Canada Ltd., 3785 Richmond Rd., Ottawa, ON, Canada K2H 5B7)

GD Canada has developed an acoustic propagation model, GDRAY, based on a Gaussian ray bundle propagation engine similar to that used in the CASS/GRAB performance prediction software [H. Weinberg and R.E. Keenan, J. Acoust. Soc. Am. **100**, 1421–1431 (1996)]. The GRAB-type model has proved to be useful for a wide-range of underwater acoustic scenarios, including active and passive prediction in shallow and deep water over a wide range of frequencies. Surprisingly, despite its extensive use in USN applications, the theoretical origins of the GRAB model are not well understood. This paper is intended to examine those origins and to outline some of the unique features of the GDRAY model implementation. It will be shown that Gaussian ray bundles, as such, have no particular physical significance in terms of the propagation, but are simply one of a family of possible ray bundle types that can be used to extract eigenray information. Several simple test cases will be employed to demonstrate the underlying concepts and issues. [Work is funded as an internal R&D program at GD Canada.]

11:35

4aUW14. An update on the multi-static model, SPADES, and its application to the reverberation workshop test suite. Gary H. Brooke (General Dynam. Canada Ltd., 3785 Richmond Rd., Ottawa, ON, Canada K2H 5B7)

GD Canada has developed a multi-static sonar performance prediction capability called SPADES. SPADES was applied to the Reverberation Workshop test cases [November 9–13, 2006, Austin, TX] and reported consistent behavior with other models in all cases except those involving refraction. This paper outlines recent progress with the SPADES model to include multi-static target echo, Nx2D propagation environments, and its current status with respect to the reverberation workshop test suite. In particular, it is shown that the model now aligns closer with the norm in all of the test cases. Sensitivities in the SPADES model to workshop problem parameters are detailed with a view to understanding the remaining discrepancies. Included also are new results for range-dependent test cases. [Work is funded as an internal R&D program at GD Canada.]

FRIDAY AFTERNOON, 30 NOVEMBER 2007 GRAND BALLROOM E, 1:30 TO 2:50 P.M.

Session 4pAA

Architectural Acoustics, Noise, and ASA Committee on Standards: Acoustics of Modular Construction

Edward C. Duncan, Chair

Resource Systems Group Inc., 55 Railroad Row, White River Junction, VT 05001

Chair's Introduction—1:30

Invited Papers

1:35

4pAA1. Acoustics of modular construction—Industry overview. Thomas E. Hardiman (944 Glenwood Station Ln., Ste. 204, Charlottesville, VA 22980, tom@modular.org)

This session will provide an overview of the issues and efforts impacting the commercial modular construction industry throughout North America, with particular focus on acoustics in relocatable classrooms. The Modular Building Institute is the international nonprofit trade association representing manufacturers and dealers of commercial modular facilites, both temporary and permanent, serving educational, health care, retail, industrial, military, and multi-family markets.

1:55

4pAA2. Issues in acoustic field testing of quiet modular classrooms. David Lubman (DL Acoust., 14301 Middletown Ln., Westminster, CA 92683) and Louis C. Sutherland (Acoust., Rancho Palos Verdes, CA 90275)

Modular classrooms are important to American education: About 300,000 modular classrooms are currently in use by public school systems here. Good acoustical conditions for learning are no less vital for students in modular classrooms than stick-built classrooms. In an effort to promote good acoustics in modular classrooms, ANSI S12 Working Group 46 is seeking to standardize acoustic field testing. Their efforts are in response to key acoustical issues of modular classrooms: Excessive noise from HVAC (heating ventilating and air conditioning) systems, and poor airborne sound insulation from exterior noise sources. In a recent and notable advance, an HVAC system provider reported good progress in modular HVAC noise reduction: A ducted wall mounted system was used instead of the usual free blowing system with exposed fans. HVAC noise in the unoccupied room was near the maximum 35 dB level required by ANSI S12.60. Interior noise levels were so low that efforts to confirm their values were confounded by noise contamination from exterior sources. The relatively high interior ambient noise levels were due to poor airborne sound insulation. Lessons learned from recent field testing will be discussed. Results of airborne sound insulation tests, now in planning stages, will be reported if available.

2:15

4pAA3. Performance and design considerations for modular panel systems used for studios and acoustic test chambers. Douglas F. Winker (Acoust. Systems, a Div. of ETS-Lindgren, 1301 Arrow Point Dr., Cedar Park, TX 78613, Douglas.Winker@ets-lindgren.com)

Modular panel construction is used to produce a variety of rooms including studios, medical rooms, hemi- and full-anechoic chambers, and reverberation chambers. Prefabricated panels provide many advantages over conventional construction, including the ability to relocate chambers, minimize on-site construction time, and maintain panel-to-panel performance through tight quality control, while maintaining the ability to customize an enclosure for a particular need. The design and performance of the panel systems will be discussed as they relate to their intended applications. The performance of each panel system was evaluated for both normal incidence and random incidence sound absorption in accordance with ASTM E1050 and C423, respectively. The panel's sound transmission loss was tested in a reverberation chamber suite in accordance with ASTM E90. Fully functioning systems were tested by ASTM E596 noise reduction of an isolating enclosure in one reverberation chamber. The performance of the modular panel systems will also be compared to conventional construction methods.

Contributed Paper

2:35

4pAA4. Noise in modular dwellings. Sergio Beristain (IMA, E.S.I.M.E., I.P.N., P.O. Box 12-1022, 03001, Mexico, D. F., Mexico)

Construction of modular dwellings in Mexico usually has had the problem of the joints; examples are given in some cases where this problem shows up as a low-quality construction, due to bad joints that no one will see. Sometimes, in order to make low cost construction, with the posibility of selling high, modular isolating panels are employed, but the details are seldom observed. Builders do not care to hire an acoustician, and there are infrequent complaints by owners.

FRIDAY AFTERNOON, 30 NOVEMBER 2007 MAUREPAS, 1:30 TO 3:20 P.M.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Biological Effects and Medical Applications of Stable Cavitation II

Saurabh Datta, Chair
Univ. of Cincinnati, Dept. of Biomedical Engineering, 231 Albert Sabin Way, Cincinnati, OH 45242-0586

Invited Papers

1:30

4pBB1. Stable cavitation in ultrasound image-guided high intensity focused ultrasound therapy. Shahram Vaezy, Wenbo Luo, Michael Bailey, Lawrence Crum (Univ. of Washington, Box 355061, Seattle, WA 981195), Brian Rabkin (Joint Genome Inst.), and Vesna Zderic (George Washington Univ.)

Microbubble activity is significantly involved in both diagnostic and therapeutic aspects of ultrasound image-guided HIFU therapy. Ultrasound interrogation techniques (A-, B-, M-mode, Doppler, harmonic and contrast imaging, and passive and active cavitation detection) were integrated with HIFU. Our results using HIFU devices of 1–5 MHz, and focal, derated intensities of 1,000–10,000 W/cm^2, show the formation of microbubbles (about 100 bubbles/mm^3, 5–100 microns in size) at the HIFU focus. Boiling, stable, and inertial acoustic cavitation activities were detected during therapy. The presence of bubbles allows the observation of the treatment spot as bright hyperechoic regions in ultrasound images, providing an effective method for guidance and monitoring of therapy. The stable cavitation of microbubbles may provide a mechanism for enhanced HIFU energy delivery, as well as induction

of biological responses for stimulation and regulation of specific physiological events such as coagulum and thrombus formation for hemostasis applications, apoptotic activity in treating tumor margins, and stimulation of immune response. Stable cavitation of extrinsic bubbles (contrast agents) is used in detection and localization of internal occult bleeding, using harmonic imaging. There appears to be benefits in utilizing stable cavitation in both diagnostic and therapeutic objectives of ultrasound image-guided HIFU. Funding: DoD, NIH, NSBRI.

1:50

4pBB2. Cavitation-enhanced ultrasound heating in vivo: Mechanisms and implications in MR-guided high-intensity focused ultrasound therapy. Shunmugavelu Sokka (Philips Medical Systems, 3000 Minuteman Rd., MS 027, Andover, MA 01810) and Kullervo Hynynen (Sunnybrook Health Sci. Ctr., Toronto, Canada, ON M4N 3M5)

Focused ultrasound is currently being developed as a noninvasive thermal ablation technique for benign and cancerous tumors in several organ systems. Although these therapies are designed to ablate tissue purely by thermal means, cavitation can occur. These bubbles can be unpredictable in their timing and location, and often interfere with thermal therapies. Therefore, focused ultrasound techniques have tried to avoid bubbles and their effects. However, gas bubbles in vivo have some potential useful features for therapy. In this research, we design and test in vivo ultrasound exposures that induce cavitation at appropriate times and take advantage of their absorption-enhancing properties while yielding reliable lesion sizes and shapes. In addition, MRI and acoustic methods to monitor and potentially control cavitation induction and the associated therapy are investigated. Finally, histology of the resulting cavitation-enhanced heating lesions is performed to assess the type of tissue damage. We demonstrate that cavitation-enhanced heating can be reliable and useful with the appropriate therapy protocol and application. If induced and monitored properly, cavitation in focused ultrasound therapy could potentially be very beneficial. Early MR-guided HIFU clinical systems that can utilize and monitor cavitation approaches will be presented.

2:10

4pBB3. Therapeutic potential of stable cavitation: From enhanced drug delivery to faster hemorrhage control. Vesna Zderic (Dept. of Elec. and Comput. Eng., The George Washington Univ., 801 22nd St. NW, Washington, DC 20052), Shahram Vaezy, and Wenbo Luo (Univ. of Washington, Seattle, WA)

Current studies indicate that mechanical effects of therapeutic ultrasound resulting in a cavitation activity may be advantageous in variety of clinical applications including tumor treatment, treatment of stroke and vascular diseases, targeted gene and drug delivery, and hemorrhage control. Our drug delivery studies using 880 kHz ultrasound have shown that enhancement of drug delivery through the cornea correlated well ($R\hat{2}=0.92$) with stable cavitation activity. The delivery enhancement ranged from two to ten times and the power of subharmonic ranged form 5 dBm to 20 dBm in the intensity range of 0.2 0.6 W/cm^2. Broadband noise, as an indication of inertial cavitation, was also detected at the highest applied intensity. Changes in the front surface layer of the corneal epithelium indicated the presence of both stable and inertial cavitation activity. Our hemostasis studies have shown that introduction of external microbubbles during application of 5.5 MHz high−intensity focused ultrasound (HIFU) led to faster hemorrhage control of solid organ injuries, appearing to result from both stable and inertial cavitation activity at the location of HIFU focus in a bleeding incision. Microbubbles also allowed easier targeting of an incision site under ultrasound guidance, thus facilitating faster localization and sealing of bleeding incisions.

2:30

4pBB4. Image-guided cavitational ultrasound therapy histotripsy. Zhen Xu, Timothy L. Hall, Charles A. Cain (Dept. of Biomed. Eng., Univ. of Michigan, Ann Arbor, MI 48109), J. Brian Fowlkes, and Williman W. Roberts (Univ. of Michigan, Ann Arbor, MI 48109)

Our recent investigations have demonstrated that mechanical tissue fractionation can be achieved using successive short (3–50 μs), high intensity ultrasound pulses delivered at low duty cycles (0.1–5%). This technique can be considered as soft tissue lithotripsy, which we call histotripsy. The acoustic pressures effective for histotripsy are similar to those used in lithotripter shockwave pulses. In bulk tissue, histotripsy produces extensive fractionation of cellular structures, resulting in a liquid cavity within the treated volume. At a tissue-fluid interface, histotripsy physically removes tissue. Histotripsy lesions have sharp boundaries, with several microns between intact and fractionated cells. The primary mechanism of histotripsy is thought to be cavitation. The production of tissue fractionation depends on initiation and maintenance of cavitation detected by acoustic and optical monitoring. Cavitation bubbles generated by histotripsy are easily detected on ultrasound imaging, which is used for targeting and real-time monitoring of the treatment progress. Histotripsy induced lesions are also clearly identifiable using ultrasound and magnetic resonant imaging. Our investigations suggest that histotripsy has potential to be developed into a new noninvasive ablation tool for many clinical applications where tissue ablation, remodeling, or removal is needed. [This research is supported by grants from NIH R01-HL077629 and Wallace Coulter Foundation.]

2:50

4pBB5. Observations of cavitation and boiling in a tissue-mimicking phantom due to high intensity focused ultrasound. Michael S. Canney, Wayne Kreider, Michael R. Bailey (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Vera A. Khokhlova (Moscow State Univ., Moscow 119992, Russia), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

Bubbles generated by acoustic cavitation or boiling are often observed during high intensity focused ultrasound (HIFU) medical treatments. In this work, high-speed video imaging, a 20-MHz focused acoustic transducer, and the driving voltage to our 2-MHz HIFU source are used to distinguish between cavitation and boiling in a tissue-mimicking gel phantom at peak focal intensities up to 30,000 W/cm^2. Bubble dynamics are modeled using a reduced order model that accounts for evaporation and condensation, heat and gas transfer across the interface, and temperature changes in the surrounding liquid. The model includes vapor trapping, whereby the noncondensable gas slows diffusion of vapor to the interface, thereby limiting condensation. At the transducer focus, evidence of cavitation is observed in the first millisecond before disappearing. Boiling is observed several milliseconds later, after sufficient heating of the focal volume to 100 °C. The disappearance of cavitation can be explained in part by the observed motion of bubbles away from the focal region due to radiation-pressure forces and in part by the softening of bubble collapses by vapor trapping. Thus, at clinical HIFU amplitudes, bubble dynamics and their impact on image-feedback and/or therapy change dramatically in only milliseconds. [Work supported by NIH DK43881 and NSBRI SMS00402.]

3:05

4pBB6. Magnetic resonance imaging of boiling induced by high intensity focused ultrasound. Tatiana D. Khokhlova (Ctr. for Industrial and Medical Ultrasound, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105 and Dept. of Optics, Phys. Faculty, Moscow State Univ., Moscow, Russia), Michael R. Bailey, Michael S. Canney, Vera A. Khokhlova, Donghoon Lee, and Kenneth I. Marro (Univ. of Washington, Seattle, WA 98105)

Bubble activity in high intensity focused ultrasound (HIFU) medical therapy is commonly but not rigorously divided between mechanically induced cavitation (μ size gas bubbles) and thermally induced boiling (mm size vapor bubbles). Our goal was to confirm that boiling occurred at 100 °C. A 2 MHz focused transducer (42 mm aperture, 44 mm focal length) was used to heat tissue phantoms in a 4.7 Tesla magnet. Temperature was measured by magnetic resonance imaging (MRI) proton resonance frequency shift and calculated from acoustic absorption. The MRI voxel was $0.3\times0.5\times2$ mm, and acquisition time was 1.3 s. Boiling was observed as a dark spot in MRI images and fluctuation in the transducer drive voltage. At 30 MPa peak pressure, boiling occurred in 7 s. Calculations yielded 100 ^{circ}C in 7 s and a temperature half maximum width of 1 mm. Averaging the calculated temperature field over the MRI voxel yielded a maximum of 73 ^{circ}C, which was the peak temperature measured in the last MRI slice before boiling. In conclusion, boiling appeared when the peak temperature reached 100 °C, and the results warn that MRI monitoring alone may underestimate the peak temperatures. [Work supported by NIH DK43881, NIBIB R21EB005250, and NSBRI SMS00402.]

FRIDAY AFTERNOON, 30 NOVEMBER 2007 GRAND COUTEAU, 1:30 TO 5:30 P.M.

Session 4pMU

Musical Acoustics and Speech Communication: Session in Honor of Max Mathews with Electronic and Computer Music Concert

Thomas D. Rossing, Cochair
Stanford Univ., CCRMA, Dept. of Music, Stanford, CA 94305

Julius O. Smith, III, Cochair
Stanford Univ., CCRMA, Dept. of Music, Stanford, CA 94305

Chair's Introduction—1:30

Invited Papers

1:35

4pMU1. A study of ambiguities in the acoustic-articulatory mapping. Bishnu Atal (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98185)

At Bell Laboratories, in 1976, Max Mathews, John Tukey, J. J. Chang, and I investigated the relationship between the shape of the vocal tract and its acoustic output for studying articulatory compensation. Although the mathematics of the forward transformation, from the vocal-tract shape to its acoustics, is reasonably clear, the same cannot be said about the inverse transformation from acoustics to the corresponding vocal-tract shape. Max, recognizing the increasing power of digital computers, suggested an interesting approach to the inverse problem, based on computer-sorting. Inversion by computer-sorting consisted of computing the output for many values of the input and sorting the resulting output-input pairs into a convenient order, according to the output data. The paper will highlight this approach and present some interesting results of this study on articulatory compensation. We were successful in determining articulatory regions (fibers) that map into a single point in the acoustic space and produce the same sound [Atal et al., J. Acoust. Soc. Am. **63**, 1535–1555 (1978)].

2:00

4pMU2. Max V. Mathews, the computer music pioneer. F. Richard Moore (Dept. of Music, Univ. of California, San Diego, La Jolla, CA 92093-0326, frm@ucsd.edu)

Max Vernon Mathews (born November 13, 1926) is the pioneer of computer music sound synthesis and control. After receiving a Sc.D in electrical engineering from M.I.T. in 1954, Mathews directed the Acoustical and Behavioral Research Center at Bell Laboratories at Murray Hill, New Jersey from 1962 to 1985, where he designed and implemented the first important computer programs to synthesize musicals sounds, starting with MUSIC I, in 1957. Mathews' ingenious strategies traversed many computer generations of improvement, and now form the backbone of much of the music synthesis industry. In 1968, Mathews turned to real time control of musical synthesis, resulting in the GROOVE system, which, like the MUSIC programs, provided a conceptual basis for further developments in computer-based musical performance. Mathews has influenced many musicians and researchers, including Pierre Boulez, who based major research aspects of his *Institute de Recherche et Coordination Acoustique/Musique* (IRCAM) in Paris on the work of Mathews and his colleagues. Mathews has received awards from the NAS, NAE, ASA, AES, and IEEE, as well as the Republique Francaise, and has been a professor of musical research at Stanford University since 1987.

2:25

4pMU3. Unit-generator architectures in computer music. Julius O. Smith III (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305, jos@ccrma.stanford.edu)

Max Mathews is well known as the "Father of Computer Music," having written the first program for generating digital sound samples from high-level "note" and "instrument" specifications. Instruments were specified as interconnections of elemental building blocks called *unit generators*. This presentation reviews some of the history and impact of the unit-generator concept in computer music over the past half-century. Widespread dissemination of the unit-generator architecture for sound synthesis began with the Music V Program (the first Fortran version), written by Max Mathews in 1968. Derivatives soon appeared at research universities, such as csound (Vercoe, MIT), Mus10 (Chowning's group, Stanford), and cmusic (Moore, UCSD). Hardware acceleration for unit generators appeared in the 1980s. Later, Music-V descendants included Lisp-based music compilers such as Common Lisp Music (CLM), the Synthesis Tool Kit in C++ (Cook *et al.*, Princeton), and graphical programming languages for real-time music processing/synthesis, such as Max/MSP and Pure Data (PD) started by Miller Puckette. More recent descendants include SuperCollider (McCartney, UCSB) and ChucK (Wang *et al.*, Princeton). Thus, the unit-generator architecture, introduced by Max Mathews half a century ago, has firmly taken root as a preferred modular approach to the problem of constructing virtual musical instruments and digital audio effects.

2:50

4pMU4. The Mathews legacy on the $100 Laptop. Barry Vercoe (Media Lab., M.I.T., 20 Ames St., Cambridge, MA 02139, bv@media.mit.edu)

The Mathews Music 4 program was the inspiration for a series of audio-processing languages by this author: Music 360 (1968), Music 11 (1973), Csound (1985), XTCsound (1997), and CsoundXO (2007). Instruments modeled in any one of these systems will run almost unmodified in any other, a testament to the long reach of Max's original concepts. But Max's longest reach is yet to come, for CsoundXO has been developed for the $100 Laptop—a machine designed by One Laptop Per Child (OLPC) which aims to put a laptop in the hands of every child on the planet (www.laptop.org). These machines are mesh-networked so that a child in one village can connect to a child in a neighboring village, and so on, until an international access point is reached. But CsoundXO on each machine is also designed for collaborative performance, so that a child in one country can "share" his/her original composition with a child in a similar or different culture. And if the work is interactive, each child can perform collaboratively with others who may be near or far. This paper will describe how all this works, along with a live demonstration.

3:15

4pMU5. Tapping into the internet as an acoustical/musical medium. Chris Chafe (CCRMA/Music, Stanford Univ., Stanford, CA 94305, cc@ccrma.stanford.edu)

Recent work in network audio transport transforms advanced networks into a new kind of acoustical medium in which sound waves propagate, as if traveling through air, water, or solids. Waves sent through the medium are reflected or altered as they bounce between hosts. Propagation delays are used to create echo chambers and build the resonances for "distributed musical instruments." As a side-effect, tones created by network resonances can be used to monitor the quality of the underlying network. The presentation presents three areas of research: (1) Auditory methods for monitoring QoS, especially for networks supporting real-time, interactive, bidirectional flows; (2) remote musical collaboration using professional-quality, low-latency audio; (3) empirical study of human factors affected by some unique acoustical properties of the medium network latency; jitter and delay asymmetry affect the speed of sound and are never uniform. By creating distributed virtual sound objects like instruments and rooms, and by studying distributed ensembles, we can begin to understand this new sound world. Some effects have been measured empirically and the results contain some surprises.

3:40

4pMU6. Max Mathews in real-time. David Wessel (Ctr. for New Music and Audio Technologies (CNMAT), Dept. of Music, Univ. of California Berkeley, Berkeley, CA 94720, wessel@cnmat.berkeley.edu)

Some of the most important contributions of Max Mathews stem from his emphasis on live-performance computer music. His performer-oriented perspective has won over musicians ranging from Pierre Boulez to Stevie Wonder. The most prominent real-time computer language in use today, Max/MSP, bears his name. In the 1960s, after making an overwhelmingly convincing argument for

the digital computer as a musical instrument, he and F. R. Moore at Bell Labs built a hybrid system, GROOVE, consisting of analog synthesis modules under real-time digital control. Though not a fully digital system—a practical move—it anticipated much of what we in the field are currently doing with laptops and gestural controllers in performance. Indeed his GROOVE conducting program impressed both Pierre Boulez and Stevie Wonder when it was demonstrated to them in the mid-1970s. I will strive to honor Max's contributions to musical performance with examples from his work and those influenced by him. Along the way there will be a few stories.

4:05

4pMU7. Max Mathews' magical music machines. John Chowning and Maureen Chowning (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305)

For nearly 40 years, Max Mathews has worked on real-time control of computers that he first programmed to make music in 1957. Beginning with a simple telegraph key to control tempo, followed by more sophisticated controllers that he developed while at Bell Telephone Laboratories and at Stanford's Center for Computer Research for Music and Acoustics (CCRMA), his latest and most sophisticated controller is his radio baton. We will demonstrate the extraordinary impact of Mathews' work on the world of music performance as the newest of musical instruments—computer and radio baton—joins with the oldest of musical instruments—the singing voice—in the performance of compositions where the soprano controls her computer generated accompaniment and computer processes. The presentation will include *Sea Songs* by Dexter Morrill, and Richard Boulanger's *Solemn Songs for Evening*, which uses the Bohlen-Pierce scale.

4:30–4:35 Break

4:35–5:00
Mini-Concert of Electronic and Computer Music

FRIDAY AFTERNOON, 30 NOVEMBER 2007 NAPOLEON A1/A2, 1:00 TO 5:15 P.M.

Session 4pNS

Noise and Animal Bioacoustics: Advances in Measurement of Noise and Noise Effects on Animals and Humans in the Environment II

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin Inst. of Fluid Mechanics and Engineering, Secr TA 7, 10587 Berlin, Germany

Ann. E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Chair's Introduction—1:00

Invited Papers

1:05

4pNS1. Quantifying lost opportunities to hear natural sounds. Kurt Fristrup (Natural Sounds Program, Natl. Park Service, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO 80525)

Hearing provides an omnidirectional alerting sense for wildlife that seems to be universal: No deaf vertebrate species are known and invertebrates display a remarkable diversity of hearing mechanisms. Anthropogenic noise elevates ambient sound levels, which masks natural sounds that would otherwise be heard. The costs of this masking can be assessed by calculating the loss of listening area or alerting distance that results. Listening area metrics are appropriate when a search function might be affected (e.g., foraging), while alerting distance metrics are appropriate when the distance to the sound source mediates the function (e.g., avoiding predation). Analytical approaches for calculating loss of listening area and alerting distance should incorporate available hearing data to account for the effects of hearing thresholds and critical bandwidths. A range of models for masking can be used. Very simple models may sacrifice accuracy to suggest metrics that are readily calculated using existing noise models. More complex models can capture the idiosyncrasies of each species hearing capabilities to render more detailed results. Examples of applying these metrics to National Park Service contexts are discussed, to illustrate the use of these concepts to render environmental acoustic data for resource managers and NPS leadership.

1:30

4pNS2. Noise impacts on birds: Assessing take of endangered species. Edward W. West (Jones & Stokes, Inc., 2600 V St., Sacramento, CA 95818), Robert J. Dooling, Arthur N. Popper (Univ. of Maryland, College Park, MD 20742), and David M. Buehler (Jones & Stokes, Inc., Sacramento, CA 95818)

Section 9 of the Endangered Species Act makes it illegal to take (harass, harm, pursue, hunt, shoot, wound, kill, trap, capture, or collect) an endangered species. Harass includes any action that would significantly disrupt normal behavioral patterns including breeding, feeding, or sheltering. ''Harm'' includes any action that actually kills or injures a listed species. To better understand how and to what extent anthropogenic noise can harm or harass birds, we examined the potential for noise effects at three levels: Physiological stress, hearing impairment, and interference with communication through signal masking. Pile driving and blasting ($\approx$100 dBA at 15 m) can potentially cause temporary or permanent hearing impairment in birds. Chronic intense noise (e.g., oil field compressor station: 92 dBA at 20 m) may induce physiological stress in some species if they cannot avoid exposure. Finally, highway noise as low as 45 dBA can potentially mask acoustic communication and modify breeding and other behaviors in many species. Thresholds of effect and effect distances are largely species-specific, reflecting differences in tolerance capacity, spacing requirements, and bioacoustic profiles. Indices of take can be physiological, behavioral, or ecological, but must be verifiably correlated with significant changes in species viability.

1:55

4pNS3. Environmental noise: Is there any significant influence on animals? Klaus Genuit and André Fiebig (HEAD Acoust. GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

It is stated that there is no convincing evidence that animals are deterred by intense sounds (Heffner 2007). Nevertheless, it must be thoroughly examined to what extent animals are impaired by specific environmental noise with regard to their orientation, communication, behavior, etc. Such interferences would significantly influence the animals' life. Similar to the auditory sensitivity of animals, the human hearing system shows a high performance in amplitude, frequency, and time resolution. It pays attention to more factors and noise features than only to absolute SPLs. Psychoacoustic parameters as well as further hearing-related parameters can capture such specific factors and features of noise, which allow us to draw conclusions concerning noise annoyance or, respectively, the pleasantness of noise. The adequate measurement and use of such parameters resulting in an advanced analysis and description of environmental noise—beyond simple sound pressure level considerations—provide valuable information with respect to noise annoyance, stress-related reactions, and interference of daily life activities of humans. The presentation will highlight the relevance, the determination, and interpretation of psychoacoustic and other hearing-related parameters in the context of environmental noise, with respect to the hearing sensation of humans.

2:20

4pNS4. Role in science in assessing noise impacts on wildlife under National Environmental Policy Act (NEPA). Sheyna Wisdom (URS Corp., 2700 Gambell St., Ste. 200, Anchorage, AK 99502)

Data uncertainty is an unfortunate reality of wildlife management. Data gaps can lead to decision uncertainty when assessing noise impacts on terrestrial and marine wildlife through the National Environmental Policy Act (NEPA) process and other regulations. How can managers take a precautionary approach to effects of noise on wildlife in light of imperfect information without imposing unacceptable restrictions on activities? This talk will describe the existing policies and/or guidelines used to assess noise impacts to marine and terrestrial wildlife under NEPA. It will provide specific examples of how noise impacts on wildlife have been assessed using GIS-based technology and industry-accepted acoustic propagation models, such as the Federal Highway Administrations Traffic Noise Model. There will also be a discussion of important gaps in our understanding of wildlife susceptibility to noise. Using principles of adaptive management, wildlife managers can implement policies that address data gaps, protect wildlife, and yet allow human activity.

2:45

4pNS5. Monitoring, prediction, and management of sonic booms in a valued ecosystem. Kenneth J. Plotkin (Wyle Labs., 241 18th St. S., Ste. 701, Arlington, VA 22202, kenneth.plotkin@wylelabs.com), Louis LaPierre (Inst. for Environ. Monitoring and Res., Moncton, NB E1C 9X5, Canada), and J. Wayne Boulton (RWDI Air Inc., Guelph, ON N1K 1B8, Canada)

Goose Bay, Labrador, is a sensitive ecosystem under airspace that has been host to military flying operations since World War II. Since 1995, the Institute for Environmental Monitoring and Research has documented and helped mitigate the effects of low altitude flight operations, serving to protect the welfare of aboriginal people and the survival of wildlife species in the area. There are current plans to conduct supersonic operations in part of the airspace. Based on experience with this type of operation in other places, there is an expectation that resultant sonic booms can be safely accommodated, but it is necessary to monitor effects. The Institute has sponsored the development of a sonic boom forecast model that combines real-time three-dimensional weather forecasts with sonic boom ray trace modeling. A set of new digital noise monitors has been developed to record noise and sonic booms in the airspace. Field biologists will observe behavioral response of key species. These elements will be key components of an adaptive management system that will ensure preservation of this highly valued ecosystem.

3:10

4pNS6. Hierarchical method for single-observer, continuous sound source observer logging. Richard Horonjeff (Acoust. and Noise Control, 81 Liberty Square Rd. #20-B, Boxborough, MA 01719) and Grant Anderson (Acoust. and Statistics, Concord, MA 01742)

Sound environments, whether urban or remote, often contain a variety of audible sound sources, each with different onset and offset times. Obtaining a continuous log of these sources can be a daunting task, especially for a single observer. In situations where it is impractical to deploy several observers to a single location, a hierarchical, forced-choice method has been developed. The method stratifies the sources into layers, each layer containing sources with common attributes, such as human-related or natural-indigenous. The method collects sufficiently detailed information for study objectives, without taxing the observers' cognitive and recording abilities, and thereby introducing large observation or recording errors. So long as the onsets and offsets of audible sources in one layer are independent of those in another, statistical descriptors, such as the percent of time audible, can be determined for each layer. The highest layer in the hierarchy becomes an exhaustive sample, whereas lower layers become statistical samples with random blocks of time removed. The method has been applied in a number of national parks over the past 15 years, by several organizations and sponsors. Advantages and disadvantages of the method, along with inter-observer logging consistency results, are presented.

3:35–3:50 Break

Contributed Papers

3:50

4pNS7. Environmental noise studies in Puget Sound. David Dall'Osto and Peter H. Dahl (Appl. Phys. Lab. and Dept. of Mech. Eng., Univ. of Washington, Seattle, WA 98110)

The ambient noise environment at a site in North Puget Sound, Washington, depth ~100 m, located within the nearby Smith Island marine sanctuary, is studied. The measurement system consists of a buoy for which the surface expression houses a microphone at nominal height 2 m above the sea surface, with multiple underwater acoustic sensors suspended in the water column. The recording bandwidth for the air system corresponds to audio band, whereas the underwater system records ambient noise at frequencies up to 50 kHz. The two systems are recorded coherently. One goal of this pilot study is to examine different components of the noise budget, including injection of noise from airplane flyovers, and correlation between pressures above and below water. Another goal relates to properties of the noise field that could possibly impact echolocation by southern resident killer whales; these properties likely being restricted to noise levels at frequencies above 25 kHz. Besides representing an important marine mammal habitat, a key advantage of this site is the availability of meteorological and sea surface wave data, obtained from Smith Island and a nearby NOAA buoy, and vessel and air traffic data. Results from the summer 2007 field work will be presented.

4:05

4pNS8. A procedure for the assessment of residential low-frequency noise complaints. David C. Waddington, Andrew T. Moorhouse, and Mags D. Adams (Acoust. Res. Ctr., School of Computing, Sci. and Eng., Univ. of Salford, Salford M5 4WT, UK)

Although the number of people that complain about low-frequency noise (LFN) is comparatively small, those that do tend to suffer severe distress. LFN is a recognized problem in many developed countries, and this paper describes a procedure for the assessment of LFN complaints that has recently been developed in the UK. Human reaction to sound is dependent not just on the sound itself, but on a complex array of other factors including personal associations with the sound. Consequently, the procedure does not provide a prescriptive indicator of nuisance, but rather gives guidance notes and a pro-forma report with step-by-step instructions to help environmental health practitioners to form their own opinion. In particular, an interview-based questionnaire is used to complement physical recordings of low-frequency sounds in the complainant's home, together with an event log completed by the sufferer. Examples of field measurements and application of the procedure are presented. [Work funded by the Department for Environment, Food and Rural Affairs (Defra) UK.]

4:20

4pNS9. Speech intelligibility in a passenger train compartment. Tze Leung Yip, William W. S. Fung, Chun Wah Leung (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong), Glenn H. Frommer (MTR Corp., Kowloon Bay, Hong Kong), and Kai Ming Li (Purdue Univ., West Lafayette, IN 47907-2031, mmkmli@purdue.edu)

An effective delivery of an informative announcement through a public address system in a mass transit train is very important from the viewpoints of safety and passengers' comfort. A high quality of speech intelligibility is needed in the compartment to facilitate the effective communication between train operators with passengers. The reverberation time is often used as one of the key elements for assessing and determining the quality and perception of the speech sound in an enclosure. Not only does it indicate the level of absorption and reflection in the train compartment, it also affects the overall noise levels and the clarity of speech. In the current study, a simple ray model is presented to estimate the sound field in a train compartment. It is then used to predict the decay curve of noise levels in which the reverberation time is calculated. Full-scale field measurements were conducted in a train compartment operating in a tunnel, for validation of the numerical model. This study aims to understand the speech intelligibility of a train compartment in a set of realistic operating conditions for setting achievable contractual specifications for the design of underground trains operating in tunnels.

4:35–5:15
Panel Discussion

Session 4pPA

Physical Acoustics: Atmospheric Acoustics

D. Keith Wilson, Chair

U.S. Army Cold Regions Research Lab., Engineering Research and Development Center, 72 Lyme Rd., Hanover, NH 03755-1290

Contributed Papers

1:30

4pPA1. Experimentally measured diffraction of sonic boom waves around a house. Victor W. Sparrow (Grad. Prog. in Acoust., Penn State, 201 Appl. Sci. Bldg., University Park, PA 16802, vws1@psu.edu), Jacob Klos, and Ralph D. Buehrle (NASA Langley Res. Ctr., Hampton, VA 23681)

During June 2006, a series of sonic boom flight tests were conducted as a part of the NASA low-boom/no boom program. One objective was to evaluate structural acoustic building response on a house. This presentation gives some results of that test regarding the outdoor pressure response on and near the house. The passage of the sonic booms over a nearby wall and the house itself have been examined to determine the maximum pressure loading. Both conventional *N*-waves and low-amplitude *N*-waves have been considered. The results indicate that the walls of the house facing the direction of the incoming sonic boom receive the maximum loading in agreement with image theory. The findings also confirm the temporal and spatial dependence of the structural loading. [The first author is supported by NASA and the FAA/NASA/Transport-Canada PARTNER Center of Excellence.]

1:45

4pPA2. Analysis of reverberation characteristics for sniper localization in an urban environment. Brad Libbey and James Perea (Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Fort Belvoir, VA 22060)

Acoustic sniper localization algorithms have proven useful to soldiers, but are difficult to implement in urban environments, due to reverberation and diffraction. Research is being performed to assist traditional localization methods by interpreting reverberant responses present in urban environments. Measured sniper responses are processed to establish if the first acoustic path is direct or indirect, based on characteristics of the received signal. If the first arrivals on several receivers are direct, an accurate relative time delay can be determined for each, which leads to accurate source localization. Indirect microphone arrivals can adversely affect the localization algorithm. Various metrics were used to characterize received signals: Arrival energy, reverberation decay time, initial arrival crest factor, and relationships between first and subsequent arrivals. Combining these factors aids in the identification of obstructed receivers, providing information about the effectiveness of a channel in a localization algorithm.

2:00

4pPA3. Acoustic surface waves and long-range detection of mortars and artillery. Xiao Di, Carrick L. Talmadge, Roger Waxler, and Kenneth E. Gilbert (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Coliseum Dr., University, MS 38677)

A natural terrain surface, because of its porosity, can support an acoustic surface wave that is a mechanical analog of the familiar vertically polarized surface wave in AM radio transmission. At frequencies of several hundred Hertz, the acoustic surface wave is attenuated over distances of less than about 100 m. At lower frequencies (e.g., below 100 Hz) the attenuation is much less, allowing surface waves to propagate thousands of meters. Thus, understanding the low-frequency surface wave is important for detecting and locating mortars and artillery. Data from recent field experiments will be shown demonstrating the persistence of the surface wave under different meteorological conditions. Propagation calculations will be presented using typical source signatures for mortars and artillery, and the implications for long-range detection will be discussed. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, New Jersey.]

2:15

4pPA4. Refraction corrections to source localization using an acoustic array suspended below an aerostat. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., 325 Broadway, Boulder, CO 80303 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003), Michael V. Scanlon (U.S. Army Res. Lab., Adelphi, MD 20783), and D. Keith Wilson (U.S. Army Engineer Res. and Dev. Ctr., Hanover, NH 03755)

Acoustic sensor arrays suspended below tethered aerostats have several advantages in detection and localization of sources on the ground in comparison with ground-based sensors. Aerostats can be elevated up to 1-2 Km above the ground where the temperature and wind velocity can significantly differ from their values close to the ground. Therefore, due to refraction, a ray trajectory that connects a source on the ground and an aerostat can bend significantly resulting in different directions of sound propagation along this trajectory. However, so far, refraction has been ignored in source localization using aerostats. In this paper, a theory is developed that allows one to find the source coordinates more accurately by taking into account refraction of sound signals emitted by the source. The source coordinates are expressed in terms of direction of sound signal propagation near the aerostat, the coordinates, and orientation of the aerostat, and the vertical profiles of temperature and wind velocity. Other meteorological parameters may also be incorporated, if available, such as time-of-day, geographical features, and historical weather trends.

2:30

4pPA5. Acoustic tomographic study of the mesoscale coherent structures in the lower atmosphere. Igor Chunchuzov Petrovich, Sergey Kulichkov, Vitali Perepelkin (Obukhov Inst. of Atmospheric Phys., 119017, 3 Pyzhevskii Per., Moscow, Russia), Astrid Ziemann, Klaus Arnold, and Anke Kniffka (Univ. of Leipzig, Germany)

The results of acoustic tomographic monitoring of the coherent structures in the lower atmosphere and the effects of these structures on the acoustic signal parameters (travel time, duration and direction of propagation) are studied. From the measurements of acoustic travel time fluctuations (periods 1 min-1h) by a net of distanced receivers the temporal fluctuations of the effective sound speed are retrieved along different ray paths connecting a pulse source and the receivers. By using a coherence analysis of the fluctuations near spatially distanced ray turning points, the internal wave-associated fluctuations are filtered and their spatial characteristics (coherences, horizontal translation velocities, and spatial scales) are obtained. A possible explanation of the presence of the dominant periods in the observed mesoscale variations of the acoustic parameters, wind shears and turbulent fluxes of momentum near ground is done.

2:45

4pPA6. Effects of internal gravity waves on acoustic pulse propagation in the atmospheric wave guides. Igor Chunchuzov, Sergey Kulichkov, and Vitali Perepelkin (Obukhov Inst. of Atmospheric Phys., 119017 Moscow, 3 Pyzhevskii Per., Russia)

Recently developed theoretical model of a random internal wave field in the atmosphere is used for calculating the statistical characteristics of the parameters of acoustic pulses propagating in the atmospheric wave guides. The variances, frequency spectra and spatial structure functions of the internal wave-associated fluctuations of the acoustic pulse travel times and grazing angles are calculated and compared to those obtained in the experiments on pulse sounding of the atmosphere. The effects analyzed here include: a scattering of acoustic waves from highly anisotropic wind speed and temperature inhomogeneities into the so-called acoustic shadow zones; a discrete character of the observed frequency spectra of the fluctuations of acoustic pulse parameters in the range 1min-1h; an error caused by internal waves in determining of the angle of arrival of the acoustic signal in the waveguide.

FRIDAY AFTERNOON, 30 NOVEMBER 2007 GRAND BALLROOM E, 3:00 TO 4:35 P.M.

Session 4pSA

Structural Acoustics and Vibration and Noise: Ground Vibration Impact on Buildings and General Structural Acoustics

James E. Phillips, Chair
Wilson Ihrig and Associates, Inc., 5776 Broadway, Oakland, CA 94618-1531

Invited Papers

3:00

4pSA1. Assessment of ground vibration impact from automotive and transit sources on future biotechnology research center using finite element analysis (FEA). James E. Phillips (Wilson, Ihrig & Assoc, Inc., 5776 Broadway, Oakland, CA 94618)

A new science and biotechnology research center was to be built in a metropolitan industrial area. There was concern that ground vibration from a nearby freeway, street traffic, and trains on an adjacent railroad would impact vibration sensitive research equipment inside the building. Ground vibration was measured at the project site prior to construction. Finite Element Analysis (FEA) was used to develop a computer simulation of the building structure using the measured ground vibration as input to the FEA model. The study determined the building floor vibration due to exterior sources would achieve the project's criteria for the building structure as designed.

3:25

4pSA2. Groundborne noise and vibration control at the Toronto Four Seasons Centre for the Performing Arts. Steven L. Wolfe (Wilson, Ihrig & Assoc, Inc., 5776 Broadway, Oakland, CA 94618)

The Four Seasons Centre for the Performing Arts, Canada's first purpose-built opera house had its grand opening on June 14, 2006. Occupying an entire block of the financial and theatre district in downtown Toronto (across from the Sheraton Hotel), the facility with its 2,000 seat auditorium, is the new home of the Canadian Opera Company and the performance venue for the National Ballet of Canada. A major challenge for the project's downtown location was isolating the auditorium from the vibration and noise from the adjacent subway system on University Avenue and surface streetcars on Queen Street. A ''building within a building'' vibration isolation design was developed that set the entire auditorium, stage, and rehearsal hall on rubber isolation bearings. This paper provides design details and the procedures used to ensure that the groundborne noise from outside activities would be inaudible. Results of follow-up measurements are presented which confirmed that the vibration design effectively mitigated the outside vibration and noise intrusion inside the auditorium to achieve background noise levels from exterior sources equivalent to the threshold of human hearing.

Contributed Papers

3:50

4pSA3. An adaptive tunable dynamic vibration absorber. Barbara Tiseo, Antonio Concilio, Antonio Gianvito (CIRA, The Italian Aerosp. Res. Ctr., Via Maiorise, Capua (CE), Italy), and Gary Koopmann (The Pennsylvania State Univ., State College, PA 16801)

This paper examines a new concept of adaptive tuned dynamic vibration absorbers (ATDVAs) using shape memory alloy (SMA) elements instead of spring elements. Shape memory is a class of alloys that shows a reversible change in crystalline structure. In the martensite phase, the material exhibits a low elastic modulus and yield strength. Subsequent heating of the material induces the change to austenite, with a corresponding higher elastic modulus and yield strength. The result of the phase transformation is a sizable change in the geometry, in the internal tension, a considerable deformation, and a concomitant frequency shift, as in the case where the SMA is used as the stiffness component of a tuned dynamic absorber. In this research, experimental investigations have been focused to verify the capability of an ATDVA to control vibration on a free-free aluminum panel.

4:05

4pSA4. A simplified method for approximating the natural frequencies of acoustic-structure systems. R. Benjamin Davis (Pratt School of Eng., Duke Univ., Durham, NC 27708)

It is well known that the natural frequency of a flexible structure in contact with a fluid-filled cavity can be approximated with a closed-form expression that considers the coupling of only two component modes (i.e., a rigid-wall acoustic mode and an in vacuo structural mode) [F. Fahy, *Sound and Structural Vibration* (Academic, London, 1985), Chap. 6, pp. 249–256]. However, this expression requires volume and area integrations over the component mode shapes. For many practical configurations, the effort to determine the component mode shapes and compute the necessary integrals counteracts much of the time and computational savings afforded by the closed-form expression. Here, the closed-form expression is recast as a function of the nondimensional frequency separation between the component modes and a new nondimensional coupling parameter, Ψ. Design curves representing the value of Ψ for common geometries and boundary conditions are presented. With the use of these design curves and a knowledge of the component modes frequencies, one can quickly compute reasonable approximations of the coupled natural frequencies by hand.

4:20

4pSA5. Vibration of circular plate and its frequency parameters. Li-Feng GE (School of Electron. Sci. and Technol., Anhui Univ., Hefei 230039, China)

The vibration of circular plates with inplane forces has been classically discussed [T. Wah, J. Acoust. Soc. Am. 34(3), 275–281, 1962, and A. Leissa, Vibration of Plates, ASA, 1993, originally, NASA, 1969]. In these literatures each of two characteristic values is expressed as a function of two stiffness factors, i.e., the inplane tension and the bending stiffness of the plate. The author has extended it to a general case in which the stretched plate is also supported by a massless elastic medium (or foundation). The corresponding two characteristic values are expressed as a function of three stiffness factors, which includes an additional stiffness of the foundation [L.-F. Ge, 16th ICA/135th ASA Meeting, Proc. 1081-1082, 1998, and J. Acoust. Soc. Am. 103, 2870, 1998]. In this paper the two characterized values and the frequency parameter have been calculated and checked further. It is found that some values given in the previous literatures have some errors, the biggest one of which is 4.0%. Furthermore, this paper developed a 3-D plot of the characterized values, and gave the corrected tables of the characterized values and frequency parameter for both simply supported and clamped plates. [Work supported by NSFC (60374044).]

FRIDAY AFTERNOON, 30 NOVEMBER 2007 NAPOLEON B3, 1:30 TO 3:45 P.M.

Session 4pSC

Speech Communication: Auditory and Somatosensory Feedback in Speech Production II

Anders Lofqvist, Chair
Haskins Laboratories, 300 George St., Ste. 900, New Haven, CT 06511

Chair's Introduction—1:30

Invited Papers

1:35

4pSC1. Compensatory responses to unexpected jaw loading during speech observed with magnetometry. Mark Tiede (Haskins Labs & MIT R.L.E., 300 George St., New Haven, CT 06511), Takayuki Ito (Haskins Labs, New Haven, CT 06511), and David Ostry (McGill U.)

Observations were made of articulator and formant trajectories during speech perturbed by unexpected mechanical loads applied to the jaw. Subjects produced multiple repetitions of the phrase ''see X avis,'' where X was one of ''raw,'' ''rob'' or ''rod.'' Perturbation forces were applied using a jaw-coupled robot, and triggered using a thresholding criterion applied to real-time tracking of the initial low to high transition in the first formant. Perturbations were delivered one out of every five repetitions, selected at random, with half applied upwards and half downwards, and forces sustained throughout the remainder of the utterance. Three EMMA sensors located on the anterior midsagittal surface of the tongue in addition to lower mandibular and upper lip sensors were used to track articulator positions. Audio and surface EMG (masseter, ABD, and OOS) were collected concurrently. Formants show initial deviation from control trajectories and then recovery that begins approximately 75 ms after the onset of perturbation. Compensation in most instances is nearly complete, even though jaw position does not recover its unperturbed trajectory. Perturbed tongue sensor trajectories are strongly distinguished by coda type, but show a pattern converging on the unperturbed tongue shape consistent with the tongue's presumed role in effecting formant recovery.

2:00

4pSC2. Sensory movement goals and control mechanisms in speech production. Joseph Perkell (Res. Lab. of Electrons, Massachusetts Inst. of Technol., Rm. 36-591, 50 Vassar St., Cambridge, MA 02139-4307)

An overview of speech production is described in which the goals of phonemic speech movements are implemented in auditory and somatosensory domains and the movements are controlled by a combination of feedback and feedforward mechanisms. Findings of motor-equivalent trading relations in producing /u/ and /r/, cross-speaker relations between vowel and consonant production and perception, and speakers use of a saturation effect in producing /s/ support the idea that the goals are in sensory domains. Results of production experiments in which auditory feedback was modified and interrupted, provide insight into the nature of feedback and feedforward control mechanisms. The findings are all compatible with the DIVA model of speech motor planning [Guenther et al., Brain & Language **96**, 280–301(2006)], which makes it possible to quantify relations among phonemic specifications of utterances, brain activity, articulatory movements, and the speech sound output. [Research supported by NIDCD/NIH.]

2:25

4pSC3. Somatosensory influence on the perception of speech sounds. Takayuki Ito, Mark Tiede (Haskins Labs., 300 George St., New Haven, CT 06511, taka@haskins.yale.edu), and David J. Ostry (McGill Univ., Montreal, QC, Canada)

Along with auditory information, sensory information associated with articulatory movements may also be involved in the neural processing that subserves speech perception. Although somatosensory organs in the orofacial system provide this kind of input, there have been few studies assessing the influence of somatosensory afferents on speech perception. To explore this idea, we examined the possibility that by altering somatosensory information, we might modify the perception of speech sounds. We tested native speakers of American English in a speech identification task using a 10-step synthesized continuum between gheadh and ghad.h Subjects listened to the stimuli one at a time through headphones and classified them as gheadh or ghad.h For half of the trials, during the presentation of the auditory stimulus, we used a robotic device to stretch the facial skin stretch lateral to the oral angle in order to modulate somatosensory information. The mechanical perturbation was timed so that its start and end coincided with that of the auditory stimulus. The results show a systematic modification of speech perception due to altered somatosensory input, which varied systematically according to the direction of the skin stretch perturbation. This suggests that somatosensory information influences the neural processing of speech signals.

2:50

4pSC4. Neural mechanisms underlying sensory feedback control of speech. Jason Tourville and Frank Guenther (Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, jtour@cns.bu.edu)

The DIVA model predicts that perturbation of speech will result in activation of sensory error cells that encode the mismatch between the expected and realized sensory consequences of speech production. Cells coding auditory and somatosensory errors are hypothesized to lie in posterior superior temporal gyrus and supramarginal gyrus, respectively. Activation of these error cells drives compensatory articulator movements, marked by increased activation of ventral motor cortex. To test these hypotheses, blood oxygen level dependent responses during normal and perturbed feedback conditions were assessed in two fMRI experiments. In one experiment, transient, unexpected auditory perturbations were induced by shifting F1 upward or downward; in another, a small balloon was inflated during speech production to induce a somatosensory perturbation. Increased bilateral activation in higher-order sensory cortices in response to the feedback perturbation was consistent with model predictions. Increased activation of frontal regions was found primarily in the right hemisphere. The results suggest that detection of feedback error by bilateral sensory error cells results in a shift from left-lateralized feedforward control of speech to feedback-based control mechanisms that rely more heavily on right hemisphere lateral frontal regions. [This work was supported by NIDCD Grant No. R01 DC02852 (F. Guenther, PI).]

3:15

4pSC5. How is auditory feedback processed during speaking? John F. Houde (Dept. of Otolaryngol.—Head and Neck Surgery, Univ. of California San Francisco, San Francisco, CA 94143) and Srikantan S. Nagarajan (Univ. of California San Francisco, San Francisco, CA 94143)

Several studies have shown that speaking suppresses the normal response to speech sounds in auditory cortex and associated regions. Our own studies using magnetoencephalography (MEG) suggest a model of how auditory feedback is processed during speaking where speaking-induced suppression (SIS) arises from a comparison between actual auditory input and a prediction of that auditory input. Over the past several years, we have been testing this model in various experiments. We have studied the specificity of SIS by examining how SIS is modulated by different types of feedback alterations. We have also looked at how choice of speech target produced modulates SIS. Our model also makes predictions about how auditory cortex would respond to transient perturbations of speech feedback, which we have tested in experiments examining the auditory response to pitch feedback perturbations. Finally, we postulate that our model of speech feedback processing is really just a special case of how auditory feedback is processed in all motor tasks with associated auditory feedback. Our recent experiments looking at responses to auditory stimuli induced by a self-initiated motor acts are consistent with this more general interpretation of our model. [Work supported by NSF Grant No. BCS-0349582 and NIH/NIDCD Grant No. R01-DC006435.]

Contributed Paper

3:40

4pSC6. A neuroimaging investigation of auditory-motor learning. Kevin J. Reilly, Frank H. Guenther, Jason A. Tourville (Dept. of Cognit. & Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215), and Jason W. Bohland (Cold Spring Harbor Lab., Cold Spring Harbor, NY 11724)

Brain activity was measured in 11 subjects while learning a novel speech-like auditory-motor task. Subjects learned to move a dot on a screen to one of three target locations by changing their formant frequencies during production of vowel-like sounds. The targets corresponded to regions of the formant space not associated with American English vowels. Functional images were acquired using a sparse sampling technique that allowed performance of the task in the absence of scanner noise and avoided artifactual signal changes resulting from articulator movements. The accuracy of subjects' production of the novel vowel sounds increased significantly over the course of the scanning session. Significant correlations between performance error and brain activity were observed in a number of areas, including superior temporal gyrus, inferior frontal gyrus, primary motor cortex, supplementary motor area, and medial and lateral cerebellum. Brain regions exhibiting a positive correlation with performance error were generally consistent with regions that have been associated with feedback error detection/correction for the production of learned speech sounds. Greater right hemisphere activation of frontal and temporal areas raises the possibility of hemispheric specialization for this aspect of speech motor control.

3:55–5:00
Panel Discussion

Session 4pSPa

Signal Processing in Acoustics, Acoustical Oceanography, and Underwater Acoustics: Session Honoring Leon Sibul II

R. Lee Culver, Chair
Pennsylvania State Univ., Applied Research Lab., P.O.Box 30, State College, PA 16804-0030

Contributed Papers

1:30

4pSPa1. Classification of impulsive-source active-sonar echoes based on a model of auditory perception. Jason E. Summers, Charles F. Gaumond, Derek Brock, and Ralph N. Baer (Naval Res. Lab., Washington, DC 20375-5350)

Models of auditory perception are investigated as the basis for a biomimetic classifier of impulsive-source active-sonar echoes. Multidimensional scaling estimates the perceptual space in which listeners perform classification [J. E. Summers et al., J. Acoust. Soc. Am. **120**, 3125 (A) (2006)]. In the resulting space of perceptual dimensions, stimuli form distinct clusters and target is discriminated from clutter along a single perceptual dimension. Dimensions in this space do not correspond to features having simple algorithmic representations. Consequently, conventional methods to develop a mapping from signal space to feature space fail. Instead, dimensions reflect untrained categorical perception manifested through the mixtures of top-down and bottom-up processes used by listeners: A high-level cognitive process for interclass dissimilarity ratings and a low-level signal-based process for intraclass comparisons. Behaving as expert systems, listeners rapidly assign stimuli to categories based on prior experiences, a process analogous to the statistical description of classes in the class-specific method. In contrast, within-class comparisons reflect signal-derived features found to be most efficacious for differentiating between the signals, a process similar to generation of features by singular-value decomposition. Implications of these findings for design of hybrid generative/discriminative human-mimetic classifier architectures are discussed. [Work supported by ONR.]

1:45

4pSPa2. Acoustic dopplergram. T.C. Yang (Naval Res. Lab., Washington, DC 20375)

Lofargram (or spectral gram) is widely used in sonar for detection of narrowband tonal signals and/or wideband transient signals. Such signals at a level comparable to the noise level are difficult to detect using snapshots of the signal spectrum, as they look like the noise. However, as noise is random in time while the signals are persistent over time, they can be detected by line-association (often referred to as eyeball integration) using the Lofargram. This concept for passive signal detection is carried forward to active signal detection using the Dopplergram. In an active system, the challenge is to detect a target echo from the noise and/or the reverberation signals. In a noise-limited situation, there is only so much power one can put out due to hardware and environmental (marine mammal) limitations. In a reverberation-limited environment, the problem is the high probability of false alarms due to reverberations. Active detections using a Dopplergram are rid of these problems in a natural way. [Work supported by ONR.]

2:00

4pSPa3. The Cramer-Rao lower bound (CRLB) on the bearing estimate for a moving array of vector sensors. Edmund J. Sullivan (EJS Consultants, 46 Lawton Brook Ln., Portsmouth, RI 02871)

The Cramer-Rao lower bound (CRLB) on the variance of the bearing estimate for a moving array of vector sensors is derived. It is shown that the motion itself provides a reduction in the variance of the estimate by utilizing the bearing information inherent in the Doppler. This is referred to as the passive synthetic aperture effect. In order to achieve this, it is necessary to jointly estimate the source frequency along with the bearing. Further, depending on the estimator used, it is necessary to provide a priori information on the source frequency. This can be done by using the observed frequency as a first estimate of the source frequency. It is shown that the CRLB decreases with the processing time and array speed, thereby significantly outperforming the conventional bearing estimator. It is then shown that by including vector sensors colocated with the pressure sensors in the bearing estimate, the CRLB decreases even further. This is due both to the increase in the effective number of sensors, and the inherent directivity provided by the vector sensor.

2:15

4pSPa4. Removal of non-Gaussian, broadband interferers using blind source separation and spatial processing. Elizabeth Hoppe and Michael Roan (131 Durham Hall, Virginia Tech., Blacksburg, VA 24061)

A method is introduced where blind source separation of acoustical sources is combined with spatial processing to remove non-Gaussian, broadband interferers from space-time displays such as bearing track recorder (BTR displays). This is quite different from the standard technique of placing nulls in the direction of interferers. The algorithm is implemented and tested against standard adaptive beamforming techniques, such as minimum variance distortionless response beamforming. In order to illustrate the utility of the approach, experiments were carried out in an anechoic chamber using two acoustic sources. The interferer source remained stationary at boresight, while the signal source moved, in steps of two deg, from negative to positive 20 deg. At each step, one second of data was recorded using a 64-element microphone array that had been calibrated at each subband frequency. All 64 channels of the array were sampled simultaneously at a rate of 22,050 Hz. Spatial processing is implemented using a broadband beamformer. The broadband beamforming algorithm first divides the signals into subbands, basebands, resamples, and finally low-passes the subbands, so that they can be processed using standard frequency domain narrowband beamformers.

2:30

4pSPa5. Performance analysis and modeling of underwater acoustic communications at low input signal-to-noise ratios using direct sequence spread spectrum. T.C. Yang and Wen-Bin Yang (Naval Res. Lab., Washington, DC 20375)

Direct sequence spread spectrum (DSSS) uses a code sequence to spread the symbols at the transmitter and a de-spreader at the receiver to recover the transmitted symbols. The de-spreader (a correlator or a

matched filter) provides a processing gain (matched filter gain), which enhances the symbol energy over noise, thus allowing communications at low input signal-to-noise ratio (SNR). For underwater communications, additional processing is needed to mitigate the effect of the rapid temporal fluctuation of the propagation channel. This paper analyzes DSSS data collected during the TREX04 experiment, which uses an m-sequence as the spreading code. More than 1,000 packets have been analyzed at input SNR, varying from -15 dB to $+23$ dB. Zero-bit errors were achieved for input SNR as low as -8 to -10 dB for two processors used. Performance loss due to inaccurate synchronization, inaccurate channel estimation, and signal fading are quantitatively modeled as a function of decreasing SNR. [Work supported by ONR.]

FRIDAY AFTERNOON, 30 NOVEMBER 2007 GRAND BALLROOM B, 3:00 TO 5:20 P.M.

Session 4pSPb

Signal Processing in Acoustics: Three-Dimensional Arrays, Machine Noise and Vibration

Natalia A. Sidorovskaia, Chair
Univ. of Louisiana at Lafayette, Dept. of Physics, P.O. Box 44210, Lafayette, LA 70504-4210

Chair's Introduction—3:00

Contributed Papers

3:05

4pSPb1. The analogy between acoustic holography and sound field reproduction. Ji-Ho Chang and Yang-Hann Kim (Ctr. for Noise and Vibration Control, Korea Adv. Inst. of Sci. and Technol., Sci. Town, Daejon 305-701, South Korea)

Acoustic holography is to predict sound field by sound pressure values on measurement plane while sound field reproduction is to generate sound field by loudspeakers. They have absolutely different objectives but handle sound field so that they have a common mathematical approach according to the shape of the sound field of interest. Hence several methods of sound field reproduction, WFS, ambisonics, BPC, mode matching method, etc., correspond with each category of acoustic holography: plane holography, cylindrical holography and spherical holography, etc. In this paper, mathematical and technical similarities are introduced and discussed. [This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the National Research Lab. Program funded by the Ministry of Science and Technology (M10500000112-05J0000-11210).]

3:20

4pSPb2. Alternative closeness functions for eye microphone array. Hedayat Alghassi (Dept. of Elec. and Comput. Eng., Univ. of British Columbia, Vancouver, BC, V6T 1Z4, Canada), Shahram Tafazoli and Peter Lawrence (Univ. of British Columbia, Vancouver, BC, V6T 1Z4, Canada)

A new signal processing algorithm, accompanied by a novel hemispherical microphone array structure for sound source localization in three-dimensional spaces was presented [H. Alghassi et al., ''Acoustic source localization with eye array,'' JASA **120**(5) (2006)]. This localization methodology, which has some analogy to the eye in localization of light rays, uses concepts of two-microphone (pinhole) or three-microphone (lens) cell structures alongside a special closeness function (CF) to approximate the proximity of the sound source direction to each of the hemisphere microphone directions based on particular similarity measures among signals. The CF plays a major role in the accuracy of the final source direction estimation. The earlier multiplicative CF (MCF) operates based on vector multiplication of spatial derivative and time derivative of microphone signals. This work presents two additional categories of CFs and compares them with MCF. The difference CF (DCF) is based on subtraction of delayed reference signal and shell microphone signals, while the correlative CF (CCF) is based on multiplication of delayed reference signal and shell microphone signals. Similar to MCF, both DCF and CCF perform demonstrated linear output versus deviation angle. Although DCF and MCF have not shown improved experimental accuracy compared to MCF, they attained lower computational complexity.

3:35

4pSPb3. A space domain complex envelope. Choon-Su Park and Yang-Hann Kim (Novic Ctr., Dept. of ME, KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea, yanghannkim@kaist.ac.kr)

Sound visualization tools, for example, beam forming and acoustic holography, exhibit the spatial look of sound in time or frequency domain. However, they normally require a significant amount of computation time to draw well the sound picture in space. The picture contains a great deal of information: Sound pressure distribution, intensity pattern, or energy information with respect to space. The information is often far more than what we need in practice. For example, when we want to know only the location of the sound sources and somewhat averaged sound pressure distribution, we need a means that can provide only what we need. The complex envelope in time domain can be a good starting idea to deal with the problems we have. A method to generate a spatial domain envelope has been theoretically developed. We found that the method not only has an advantage to show rather simple spatial sound distribution, but also reduce significantly the computation time. The latter makes it possible to see the sound picture faster than before: About ten times faster. This method has been applied to many practical examples: For example, sound from a musical instrument and sound from machinery.

3:50

4pSPb4. Acoustic source identification using a generalized regressive discrete Fourier series for tomographic reconstruction. Joris Vanherzeele, Roberto Longo, Steve Vanlanduit, and Patrick Guillaume (Dept. of Mech. Eng., Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium)

When measuring three-dimensional phenomena such as acoustic fields using an interferometric technique, one is prone to measure different angles of view to obtain a full three-dimensional representation of the phenomenon under investigation. This is due to the fact that an interferometric technique measures a line integral across the optical path. To obtain the full three-dimensional view, the different angles of view are passed through a tomographic algorithm. The most widely used tomographic method is filtered back projection. However, this process suffers from a series of drawbacks, the most important one being the fact that substantial truncation errors occur in the back projection step. In this article, a method is devised to eliminate these errors, based on a parametric frequency do-

main approach called generalized regressive discrete Fourier series (GRDFS). The method will be applied to laser doppler vibrometer measurements on a loudspeaker. This source will be rotated to obtain continuous angle views of the acoustic field, hence eliminating the tedious process of rotating the measurement set-up. By demodulation of a measured signal, it is possible to determine the position of the loudspeaker in space. The results obtained with the GRDFS will be compared to the classical filtered back projection method.

4:05

4pSPb5. Ultrasonic Doppler vibrometry using direct undersampling. Asif Mehmood, Paul M. Goggans (Dept. of Elec. Eng., Univ. of Mississippi, Anderson Hall Rm. 302B, University, MS 38677, amehmood@olemiss.edu), and James M. Sabatier (Natl. Ctr. for Physical Acoust., University, MS 38677)

In ultrasonic Doppler vibrometry (UDV) systems, sound returned from a moving point scatterer is frequency modulated by the component of the scatterer's velocity in the direction of the system's colocated ultrasonic transducers. Because of this, time-frequency analysis of the receiver transducer output can be used to reveal the velocities of multiple scatterers moving within the sensor's field of view. This principle can be used, for example, as in [Zhang et al., EL110 J. Acoust. Soc. Am. 121 (2007)] to study human gait. Because it is band-pass limited, the received signal can be expressed equivalently in terms of low-pass limited in-phase and quadrature (I&Q) components. To reduce the amount of data to be stored and analyzed, it is advantageous to sample the I&Q signals rather than the received signal. This paper reports an implementation of undersampling [J. L. Brown, Jr., IEEE Trans. Information Theory IT−26, 613−−615 (1980)] to capture samples of the I&Q using a data acquisition card without the use of external oscillators or mixers. Undersampling yields I&Q samples that are interlaced rather than simultaneous in time. Here, a spectrogram using interlaced I&Q samples is derived using Bayesian spectrum analysis.

4:20

4pSPb6. On-line failure detection of a vibrating structure: A model-based approach. B. L. Guidry, J. V. Candy, K. A. Fisher, D. H. Chambers, and S. K. Lehman (LLNL, P.O. Box 808, L-154, Livermore, CA 94551)

Model-based failure detection is based on the principle that the MBP for a normal or pristine structure is developed first and tuned during the calibration stage assuring a statistically validated processor. Once developed, the MBP is used as the integral part in a sequential detection scheme to decide whether or not the structure under investigation is operating normally. When an abnormality is detected, a failure alarm is activated and the type of failure is classified based on a library of potential failure modes. Here we use high-order parametric models to capture the essence of the structures over a limited frequency band known to be sensitive to structural changes. These estimated or identified models for normal operations are then used to develop the MBP which in this instance is a recursive Kalman filter. The filter is known to produce zero-mean/white residuals when optimally tuned to the data. Failure is declared when these properties are no longer valid. Once the detection is accomplished, the next step is to classify the type of failure mechanism and eventually the locations. Here we show results of the designs on both simulated and measured data.

4:35

4pSPb7. Fault localization of moving machinery in a noisy environment. Jong-Hoon Jeon, Choon-Su Park, and Yang-Hann Kim (Ctr. for NOVIC, Dept. of Mech. Eng., KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea)

Faults of rotating parts of a machine normally generate unexpected frequency band or impulsive sound, which has a period when it moves with a constant speed. The former can be detected by the moving frame acoustic holography method [S.-H. Park and Y.-H. Kim, ''An improved moving frame acoustic holography for coherent bandlimited noise,'' J. Acoust. Soc. Am. **104**, 3179–3189 (1998)]. We have attempted to apply the method to the latter case in the test site. The keywords are, therefore, the periodic impulsive sound, which is obviously different from those which can be visualized by the method, and the signal-to-noise ratio, which determines the success of early-fault detection. This research shows how the problems related with these keywords can be resolved. The main idea is that periodic impulsive signal can be expressed by an infinite set of discrete pure tones. This enables us to obtain alot of holograms that visualize periodic impulsive sound field and noise; therefore, holograms can be averaged to improve the signal-to-noise ratio until having reliable information that exhibits where the impulsive sources are. Theory, experiment, and application results to the train on a test rig are described. [Work supported by BK21 and KRRI.]

4:50

4pSPb8. Design of multi-input interleaved multisine excitation signals for ultrasonic testing. Roberto Longo, Steve Vanlanduit, Joris Vanherzeele, and Patrick Guillaume (Dept. of Mech. Eng., Vrije Universiteit Brussel (VUB), Pleinlaan 2,1050 Bruxelles, Belgium)

Multisines are periodic signals consisting of harmonically-related sine waves. By optimizing the phases of the different sine waves, a periodic signal is obtained with a small crest factor resulting in high signal-to-noise ratios. One disadvantage of multisines when several ultrasonic actuators are used simultaneously is that, in general, it is very difficult to distinguish from a measured ultrasonic signal the contribution of the different actuators. This is easier when pulses are used and when they do not overlap. In this contribution, a new approach for periodic continuous wave signals will be presented to separate the contribution coming from the different actuators (senders) at every receiver. The proposed approach is based on the use of multi-input interleaved multisines. Interleaved multisines contain energy at different spectral lines allowing an easy separation by means of a discrete Fourier transform. The separation becomes more complex when non-linear effects are present, but even then it is possible to apply this approach by properly selecting the excitation lines. The use of multi-input interleaved multisine offers new applications for ultrasonic NDT (e.g., combined transmission and reflection measurements) as well as in the field of material identification and biomedical applications.

5:05

4pSPb9. Bayesian spectrum estimation of termite signals using laser Doppler vibrometry. Asif Mehmood (Dept. of Elec. Eng., The Univ. of Mississippi, P.O. Box 4206, University, MS 38677), Orwa M. Tahaineh, Tom Fink, Vijay. P. Ramalingam, John. M. Seiner (Natl. Ctr. for Physical Acoust., University, MS), and Alan R. Lax (USDA-ARS Southern Regional Res. Ctr., New Orleans, LA)

Laser vibrometry provides a sensitive non-invasive means of measuring substrate vibration. These measurements include landmine detection, automotive testing, production testing, aerospace, and structural testing. We employed laser vibrometry for termite detection. The vibratory signal generated by termite head banging is picked up by the laser Doppler vibrometer. In laser vibrometry, the surface motion is monitored by heterodyne laser Doppler vibrometry, and the received heterodyne signal is sampled to produce time-series data. The time-series data thus obtained is the velocity signal of the vibrating object. We consider here a statistical signal-processing approach to termite head banging data, which is based on Bayesian inference. In this approach, the quantity of interest is the frequency of vibration caused by termite head banging, while phase and quadrature amplitudes are considered nuisance parameters. We employ a single-frequency model to determine this frequency. Because of the optimal prior knowledge about the signal of interest, the frequency can be measured with much greater precision and greater noise immunity than using Fourier transform. Our results show that the Bayesian method of processing an acoustic signal exhibits excellent performance in determining the vibrational frequency.

Session 4pUW

Underwater Acoustics: Underwater Reverberation Measurements and Modeling II

John S. Perkins, Cochair
Naval Research Laboratory, Code 7140, Washington, DC 20375-5350

Eric I. Thorsos, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6606

Contributed Papers

1:00

4pUW1. Reverberation modeling with BiRASP—The bistatic range-dependent active system prediction model. David Fromm (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5350, david.fromm@nrl.navy.mil)

The calculation of bistatic reverberation is a computationally intensive exercise. Efficient computation of timely results typically involves numerous physical and algorithmic assumptions or approximations. Consequently, it is not unusual for predicted time series from different models to differ in level or structure. At a recent Reverberation Modeling Workshop [9–13 November 2006, Austin, TX], a set of test cases were proposed with the intent of establishing benchmark reverberation problems and solutions. This talk will focus on the test case performance of BiRASP, the Naval Research Laboratory's bistatic reverberation model. BiRASP is a ray-based code, originally developed for deep-water, long-range, low-frequency reverberation predictions and data analysis. Results for the various test case problems will be shown. The consequences of some of the typical algorithm and physical assumptions/approximations will be demonstrated and discussed. [Work supported by the Office of Naval Research.]

1:15

4pUW2. Constrained comparison of ocean acoustic reverberation: Theory and observations. Charles Holland (The Penn State Univ., Appl. Res. Lab, State College, PA)

Measurements of long-range (order 10^4 m) shallow-water reverberation in the Straits of Sicily at 900 and 1800 Hz are compared with theoretical predictions. All of the required environmental inputs for the theory are obtained independently, that is to say there are no free parameters. The reflection coefficient and the scattering strength are measured by direct path methods; both quantities show strong frequency dependence. The theoretical reverberation predictions using these measurements are in good agreement with directional reverberation data, i.e., within the expected uncertainty bounds. The good agreement suggests that the supporting environmental measurement techniques are robust and that the physics associated with reverberation in a waveguide is reasonably well understood, at least in simple environments. The ability to independently measure the seabed scattering strength and reflection coefficient is a crucial step for the advancement of inverse methods using reverberation (e.g., rapid environmental assessment) inasmuch as it provides the means for quantitatively measuring the robustness of those methods. [Research supported by the Office of Naval Research OA321 and the NATO Undersea Research Centre.]

1:30

4pUW3. Reverberation modeling issues highlighted by the first Reverberation Modeling Workshop. Eric I. Thorsos (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, eit@apl.washington.edu) and John S. Perkins (Naval Res. Lab., Washington, DC 20375)

Participants at Reverberation Modeling Workshop I presented solutions to posed reverberation problems using a wide variety of methods. The reverberation problems were posed in both two- and three-space dimensions, and included cases with different levels of boundary roughness, different sound speed profiles, and in some cases, range dependence. A number of important reverberation modeling issues became evident when the results of this workshop were considered in detail following the workshop. These include the important role of coherent effects in determining reverberation structure at short ranges, and the important role of boundary reflection loss in affecting the reverberation level at long ranges. A summary of modeling issues highlighted by this workshop will be presented. [Work supported by ONR and PMW-180.]

1:45

4pUW4. A comparison of two modeling approaches for reverberation in a shallow-water waveguide where the scattering arises from a sub-bottom interface. Dale D. Ellis (DRDC Atlantic, Dartmouth, NS Canada, dale.ellis@drdc-rddc.gc.ca) and Charles W. Holland (The Penn State Univ., State College, PA)

In shallow water environments with a low-speed sediment layer overlying a higher-speed sub-bottom, the observed reverberation may be dominated by scattering from the sub-bottom. Here, reverberation predictions from normal mode and energy flux models are compared for the case where the scattering arises from a sub-bottom half-space under the low-speed sediment layer. It is shown that in such an environment, the position of the angle of intromission, in addition to the angular dependence of the scattering kernel, is a factor controlling the vertical angle distribution. It is also shown that the reverberation from a sub-bottom horizon is typically governed by higher grazing angles than the case where the scattering occurs at the water-sediment interface. There was generally very close agreement between the models as a function of frequency (200–1,600 Hz), layer thickness (0–8 m), and range (1–15 km). The model comparisons, showing some differences, illuminate the effect of different approximations in the two approaches. [Work supported in part by the Office of Naval Research.]

2:00

4pUW5. A model of narrow-band normal-mode reverberation in shallow water. Anatoliy Ivakin (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

A unified model of reverberation in a shallow-water waveguide caused by the volume heterogeneity and rough interfaces is proposed. Normal modes are used to describe propagation of a narrow-band signal from a

point source to the scattering volume (a vertical column of the waveguide including both water and sediment) and from the scattering volume to the receiver. A local scattering matrix describes a process of reradiation from one normal mode to another within the scattering volume. The case of statistically axial symmetry, where the source and receiver are separated only in the vertical direction and the medium is statistically homogeneous and isotropic in the horizontal plane, is considered in more detail. A simple relationship of the temporal dependence of the reverberation intensity with the scattering matrix, attenuation, and group velocities of the normal modes is obtained. Contributions of different components of reverberation due to rough air-water and water-sediment interfaces, volume heterogeneity of the water column, and the sediment are discussed.

2:15

4pUW6. The effect of different environments on shallow-water reverberation: Measurements and modeling. Roger Gauss (Naval Res. Lab., Code 7144, Washington, DC 20375-5350) and John Preston (Appl. Res. Lab., The Penn State Univ., State College, PA 16804)

Low-frequency reverberation in shallow water is a complex and difficult quantity to predict as its mean levels can be due to a combination of seafloor, sea surface, and fish scattering, coupled with propagation and system-dependent effects. In this paper the importance of having a physics-based modeling capability and the correct spatially-dependent environmental description will be demonstrated via data-model comparisons corresponding to monostatic data the authors have collected over the last 15 years in a range of distinct acoustic environments, from silt to sand to chalk to basalt. Additionally, the use of in situ measurements to both refine predictions and mitigate uncertainty is discussed. [Work supported by ONR.]

2:30

4pUW7. Frequency and angular dependence of bottom scattering strength in shallow water with a sandy seabed. 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)

The bottom scattering strength at low frequencies and small grazing angles is difficult to directly measure in shallow water. It is generally derived from mid- and long-range reverberation measurements. Reliable estimation of the bottom scattering strength from shallow-water reverberation requires a correct reverberation model, quality reverberation data, and an appropriate seabed geo-acoustic model that controls two-way sound propagation. In this paper, an effective Biot model for sand-silt bottoms, derived from low-frequency field measurements at 18 locations in different coastal zones around the world is used [J. X. Zhou and X. Z. Zhang, J. Acoust. Soc. Am. **117**, 2494 (2005) and **119**, 3447 (2006)]. A simple closed form expression for the reverberation in shallow isovelocity water [J. X. Zhou, Acta Acustica **5**(2), 86–99(1980) in Chinese] as well as the normal-mode expression for the reverberation in shallow water with an arbitrary sound speed profile are used to derive the bottom scattering strength. The bottom scattering strength as a function of frequency and angle, derived from broadband reverberation data, will be reported and discussed. [Work supported by ONR.]

2:45

4pUW8. High frequency acoustic bottom backscatter at shallow grazing angles. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas at Austin, Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

Variability is an important feature of high frequency acoustic bottom backscatter. A relationship had been developed between grain-size and backscattering strength in the laboratory, but it rarely applies in the field. The main problem is that real ocean sediments have surface roughness and volume scatterers or inclusions that are continually evolving. The result is a rather large range of variations as a function of space and time. In some cases, estimates of the variability may be obtained from the extant database of published measurements. [Work supported by the Office of Naval Research, Ocean Acoustics.]

3:00

4pUW9. Demonstration of coherent gain from synthetic aperture for a midfrequency sonar in shallow water. Kevin D. LePage (Naval Res. Lab, Code 7144, 4555 Overlook Ave SW, Washington, DC 20375)

The Clutter 07 experiment was carried out this May in the Straits of Sicily on board the R/Vs Alliance and Oceanus with the participation of NATO Undersea Research Centre, NRL, Applied Research Lab-Pennsylvania State University, and Defense Research and Development Canada. High repetition rate monostatic active sonar data collected off the alliance on the Malta Plateau during this experiment is processed to show significant inter-ping correlation indicative of the utility of performing coherent synthetic aperture sonar (SAS). Strip SAS images of scattering from a rocky outcrop on the bottom known as the Ragusa ridge show that the lateral extent of returns from compact objects can be reduced through SAS techniques for real shallow water environments using autofocusing techniques. [Work supported by ONR.]