CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 125, No. 5

May 2009

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

GENERAL LINEAR ACOUSTICS [20]

UNDERWATER SOUND [30]

(Continued)

CONTENTS—*Continued from preceding page*

ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35]

STRUCTURAL ACOUSTICS AND VIBRATION [40]

NOISE: ITS EFFECTS AND CONTROL [50]

ARCHITECTURAL ACOUSTICS [55]

ACOUSTIC SIGNAL PROCESSING [60]

(Continued)

CONTENTS—*Continued from preceding page*

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Sound speed and density characterization of milk adulterated with melamine

Luis Elvira and Jaime Rodríguez
Instituto de Acústica, CSIC, Serrano 144, Madrid 28006, Spain
lelvira@ia.cetef.csic.es, jaimerl@ia.cetef.csic.es

Lawrence C. Lynnworth
Lynnworth Technical Services, 77 Graymore Road, Waltham, Massachusetts 02451-2201
larry@kynosoura.com

Abstract: Milk contaminated with melamine resulted in an important health hazard that affected many babies in China recently. Ultrasonic characterization of adulterated milk may detect gross levels of melamine contamination. Sound speed and density measurements were made in skim milk as a function of melamine adulteration. An ultrasonic measurement technique to implement milk quality control is discussed.

© 2009 Acoustical Society of America

PACS numbers: 43.20.Hq, 43.35.Bf, 43.58.Dj [AN]
Date Received: December 17, 2008 **Date Accepted:** March 3, 2009

1. Introduction

Shortly after the conclusion of the 2008 Olympics, international health alerts were posted in response to milk adulteration in China. That adulteration resulted in kidney illness of varying degrees affecting, according to some reports, about 300 000 babies, 6 of whom died. In the global economy, food products often contain ingredients coming from different countries, which make quality control difficult. Although international regulations try to assure the safety of foods all over the world, this illegal, intentional contamination provides compelling evidence that better, reliable techniques are needed to guarantee the safety of foods we consume.

In the particular case of the tainted milk cited above, adulteration was not detected because the traditional milk quality sensors that were used were based on nitrogen quantification as an indirect method to determine protein content.[1] Although water was added to milk to fraudulently increase the productivity, melamine, which is an organic molecule rich in nitrogen, was added too, hiding the milk adulteration. Following Barbano and Lynch,[2] the development of multi-sensor quality sensing devices, based on the measurement and detection of different parameters, would improve the odds of detecting fraud. This is not so different in principle from an ordinary person checking a suspicious beverage by smelling, tasting one small sample, viewing the color, or jostling the beverage to intuitively judge viscosity or density or some other characteristic.

Ultrasonic non-destructive assays are commonly used in industry as reference techniques to assess the quality of many products and detect incipient or eventual failures in structural elements, materials, and processes. Ultrasound is widely used in industrial process control, e.g., in measuring the flow velocity of liquids, gases, and some mixtures, also liquid level and in other specialized analytical measurements such as composition of mixtures. Many industrial applications benefit from the transducers being external to the boundaries of the process. In NDT/NDE, e.g., thickness gauging or weld inspection, the interrogation is non-destructive even when the medium must be contacted by the transducers. Medical uses of ultrasound are familiar in clinics and hospitals around the world. Nevertheless there is limited application of ultrasound in the quality control of foods, drugs, and biological media, where chemical or biochemical techniques are commonly preferred. Recent developments are showing that ultrasonic-based sensing systems are well suited to achieve quality control in these industrial sectors,[3] and research is being conducted for the control of fermentation and gelation processes, enzymatic reaction monitoring, or microbiological growth detection in milk.

Ultrasonic techniques have been implemented in non-invasive on-line measuring systems. They can avoid the important hazard of product contamination, with the advantage of making real-time corrections in the production chain and may permit easy automation of quality control. Another important feature of ultrasonic measuring techniques is that, in most cases, neither reactive nor replaceable elements need to be added to the medium under test. This implies that the ultrasonic inspection and quality control system may be implemented in an environmentally friendly and economical manner. On the other hand, ultrasonic inspection of milk based on sound speed c is subject to interfering variables (temperature, fat content, etc.) that can mask detection of an adulterant such as melamine.

The European Food Safety Agency[4] (EFSA) states a limit of 0.5 mg/kg body weight for a tolerable daily intake (TDI) of melamine. Nevertheless, melamine concentration in milk up to 2.5 g/kg was found in the cited adulteration, exceeding the TDI by a factor of 5000. In the present work, it is shown that acoustical characterization of tainted liquid milk, through density and/or sound speed measurements, can be used to detect gross melamine contamination, provided uncertainty in interfering variables such as milk fat or temperature does not mask m, the melamine concentration in adulterated milk. To the extent that an ultrasonic inspection can be implemented more easily and economically than alternatives, it might be a worthwhile first step in screening. Products passing the ultrasonic test might still contain dangerous levels of contaminant below ultrasonic detection limits. Therefore, products which get a passing grade from the initial screening ought to be subjected to other tests (e.g., mass spectrometer) that can reliably detect smaller but nevertheless potentially hazardous concentrations, even though such subsequent tests might be more complicated or more expensive to conduct on-line or in samples of batch-produced product, compared to a clamp-on or non-invasive ultrasonic test.

2. Experimental setup

2.1 Sample preparation

A liquid mixture was prepared containing distilled water, lactose, and melamine. The proportion used was 47 g of lactose and 8 g of melamine for each kilogram of water. This liquid has the same lactose concentration as milk. The melamine concentration was adjusted to reach the same nitrogen level as ordinary milk. Adding such a mixture to milk maintains constant sugar content and nitrogen levels of the resulting liquid. Skim milk was used in the experiment to diminish the influence of fat dilution when the melamine mixture was added.

A magnetic stirrer with a heating plate was used to dissolve melamine and lactose in water. This mixture was maintained at 40 °C for the experiments to avoid melamine precipitation, as the solubility of melamine in water at room temperature, 20 °C, is 3.1 g/l. Then it was added to skim milk up to 30% by weight for the analysis. Using Eq. (1) the melamine concentration in milk, m, expressed in g/kg can be obtained as a function of the mixture addition M in percent:

$$m = 0.075M. \tag{1}$$

An amount of 2.25 g of melamine per kilogram of milk corresponds to 30% mixture concentration (Fig. 1). Although a 40 °C temperature was not necessary to dissolve 2.25 g of melamine in milk, the experiment was made at this temperature, to avoid precipitating the mixture before it was added to milk.

2.2 Measurement techniques

Milk samples were prepared containing different concentrations of the melamine-lactose-water mixture. A DMA 4100 Anton Paar density analyzer determined the density of the adulterated milk at 40 °C.

As the melamine mixture was continuously added to the milk, time of flight variations were measured using the ultrasonic device developed by Elvira *et al.*[5] The liquid sample was put

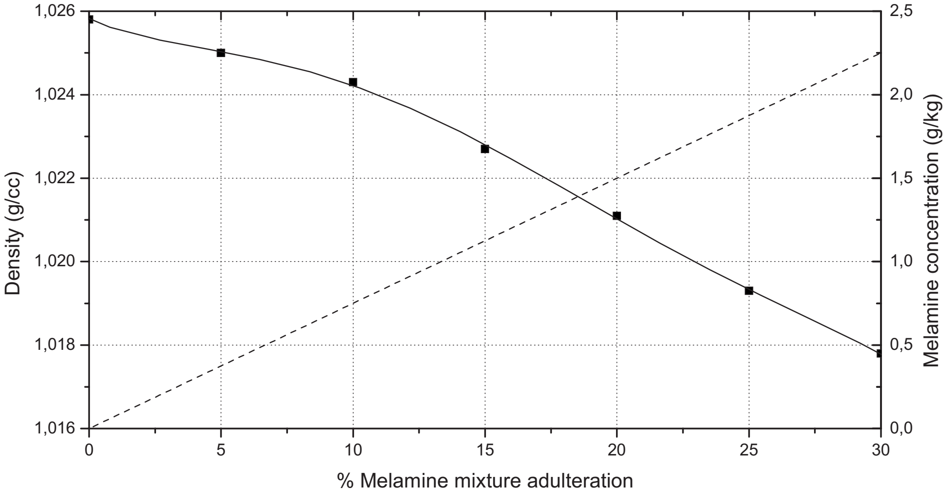


Fig. 1. Density and melamine concentration of the adulterated milk as a function of the melamine mixture concentration. Squares are density measurements and the solid line is a fifth order polynomial fit. Melamine concentration (dashed line) is obtained from Eq. (1).

in a 125 ml commercial glass bottle which was placed inside one of the ultrasonic measurement cells. The liquid path was 36 mm, traversed once. These cells are inside an environmental chamber. Each measurement cell consists of a thermostated housing with two ultrasonic transducers. Each consists of three PZT piezoceramic plates (PZ27 from Ferroperm). The transducers are placed on the opposite sides of the bottle and pressure-coupled through silicone layers when the housing door is closed. Temperature sensors, temperature actuators (Peltier cells and resistances), and control electronics based in PID (proportional-integral-derivative) controller devices are used to avoid specimen temperature variations greater than 0.02 °C. An efficient temperature control is necessary to obtain accurate measurements because sound speed is sensitive to temperature. For skim milk, near 40 °C, $dc/dT \sim 1.5$ m/s °C, close to the value for water.

The electronics to excite and receive signals is implemented in a PXI chassis (National Instruments). It is comprised of a ZTEC 530 Function Generator, a ZTEC 410 digitizer, a NI-2503 24 channel multiplexer, and a NI PXI-PCI-8336 MXI-4 module to communicate between the chassis, NI-1036, and a personal computer. The 4 MHz ultrasonic tone burst is captured after propagating through the liquid. From this through-transmitted pulse, variations in time of flight, resolved to ±1 ns (approximately 1 part in 23 000) were determined using fast Fourier transform signal processing. Sound speed c in skim milk at 40 °C was measured to be 1548 m/s. Therefore with this ultrasonic device, c changes as small as ±7 cm/s may be detected as a function of the melamine mixture added. In pure water at 40 °C, $c = 1528.863$ m/s.[6]

A peristaltic pump, Instech-720, was used to add the mixture to milk. Two tubes pass through the cap of the bottle for this purpose, one for mixture pumping and the other for pressure stabilization. The mixture was continuously added at a slow flow rate of 0.45 ± 0.01 ml/min to avoid temperature changes in the sample. The mixture, kept inside the climatic chamber, was maintained at the 40 °C working temperature.

3. Results and discussion

Density measurements were obtained at 5% mixture intervals. As can be seen in Fig. 1, a clear decrease in density was obtained. It is mainly due to the dilution of milk. True milk has a 32 g/kg concentration of proteins which is higher than its substitute (melamine) which was adjusted to 8 g/kg in the adulterating mixture. The density variation as a function of melamine is even more significant when the melamine concentration is more than 10%.

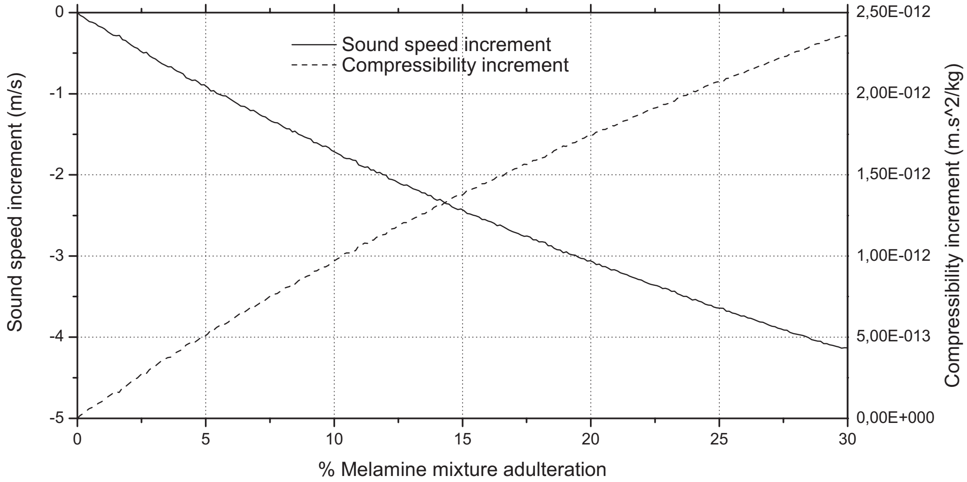


Fig. 2. The solid line shows the sound speed variations as a function of the melamine mixture concentration. Dashed line is the compressibility calculated from sound speed and density measurements.

When the milk was analyzed in the ultrasonic characterization device, an important *decreasing* of sound speed up to 4 m/s was registered (Fig. 2). The measured density decrease, at constant compressibility, would have produced a velocity increase, so the measured sound speed diminution means that the isentropic compressibility of the adulterated milk increases. This increase is shown in the dashed line which was computed from the measured density and the sound speed data (Fig. 2). As stated above, the liquid obtained when the mixture was added has a lower solute concentration. This lower concentration is responsible for both the compressibility increasing and the density decreasing. The c data for melamine in milk are believed to be new. It is reasonable to think that an adulteration made to obtain an economic profit by a fraudulent increasing of milk volume must incorporate a significant amount of melamine mixture, as occurred in this recent case. Ultrasonic screening, to the extent it is inexpensive, robust, and easy to implement, might prove to be a worthwhile quality control method for milk.

From data in Fig. 2 and the literature,[5,6] an equation relating c in melamine-adulterated skim milk to melamine mixture concentration M in percent and temperature T near 40 °C may be derived, for M up to 30% by weight:

$$\Delta c = AM + BM^2 + C\Delta T, \tag{2}$$

where $A=-0.1873$ m/s, $B=0.001\,65$ m/s, and $C=1.5$ m/s °C. Δc is in m/s and ΔT is in °C.

According to EFSA,[4] a harmful concentration for babies will be near 0.003 g of melamine/kg milk, which means $M=0.04\%$ from Eq. (1). This will increase transit time across a 36 mm liquid path by 0.1 ns and reduce c by 7 mm/s. Following Eq. (2), the same c reduction would occur for a T reduction of 0.005 °C. A small uncertainty in fat content or other composition factor is also likely to mask 3 mg of melamine/kg milk. However, if c uncertainties due to T, fat, or other variables amount to <1 m/s, and if c can be measured online to 0.1% accuracy (±1.5 m/s), then total uncertainty in sound speed will be <2.5 m/s. From Fig. 2, this speed detection limit corresponds to a minimum detectable value for M of 15%. It appears reasonable that 15% melamine adulteration can be detected on-line. Even though this level, $M=15\%$, is substantially above the stated TDI minimum harmful concentration ($M=0.04\%$), detecting at the $M=15\%$ level would have been sensitive enough to detect 30% gross adulteration as reached in the recent milk contamination.

Relationships among sound speed, density, and protein concentration have been noted previously in a liquid which is substantially different from milk, namely, blood at 37 °C.[7] Ultrasonic thickness gauges and ultrasonic transit time (contrapropagation) flowmeters operating at or above 1 MHz typically resolve transit times or transit time differences to subnanosecond precision. In some transit time *ultrasonic flowmeters*, the sound speed has been used to determine concentration of a binary mixture.[8] Sometimes *ultrasonic thickness gauges* are used to test specimens of known thickness, yielding sound speed in the specimen, which in turn may be related to elastic moduli, alloy type, or other compositional aspect. This industrial experience suggests that commercially-available NDT or flowmeter process control instrumentation might be adaptable to achieving reliable on-line resolution of transit time comparable to that obtained in the laboratory with the present equipment. However, limits on using c to determine m or M probably come from interfering variables, rather than from the precision of a c determination.

In the laboratory there is ample opportunity to determine the wall thickness and wall transit time of the vessel and eliminate its uncertainty from the measurement of sound speed. On-line, uncertainty in wall thickness might limit accuracy of clamp-on sound speed measurements. Differential path solutions[9] may obviate such on-line clamp-on potential problems. While not investigated in the present study, one might utilize scattering, attenuation, or received amplitude[5] as the basis of further ultrasonic methods to detect undissolved adulterants such as settled melamine powder in closed vessels. The practicality of an on-line ultrasonic detection system for screening or other purposes is limited in part by contributions to c by interfering variables. Quantifying the influence of fat content and other contributions to c as would be encountered during normal milk production appears to be the logical next step.

4. Conclusions

Melamine powder, dissolved in water and added to skim milk at 40 °C, to create an adulteration concentration of 30%, decreases the density of the mixture by about 0.78% and decreases the mixture's sound speed by about 0.25%. Sound speed c was determined in the present experiments with resolution of about 1 part in 23 000 using a through-transmission liquid path of 36 mm, traversed once, and timed to ±1 ns. Absent interfering uncertainties due to temperature T, fat content, or other factors, resolution of c to ± 1 m/s corresponds to 6% melamine mixture at minimum concentration. However, even with c measured on-line to 0.1% accuracy, if uncertainties in T or other factors lead to ± 1 m/s contributions to c, then melamine concentrations below 15% are likely to be masked. While the practicality of an on-line ultrasonic detection system for screening or other purposes is limited in part by contributions to c by interfering variables, an ultrasonic technique capable of detecting 15% melamine concentration would have been sufficiently sensitive to detect gross adulteration at the 30% level reached in the recent milk contamination (2.5 g of melamine per kilogram of milk). It is reasonable to think that an adulteration made to obtain an economic profit by a fraudulent increasing of milk volume must incorporate a significant amount of melamine mixture, as occurred in this recent case. Ultrasonic screening, if inexpensive, robust, and easy to implement, might prove to be a worthwhile quality control method for milk.

Acknowledgment

The authors acknowledge Dr. David M. Barbano's comments about practical limitations.

References and links

[1] J. M. Lynch and D. M. Barbano, "Kjeldahl nitrogen analysis as a reference method for protein determination in dairy products," J. AOAC Int. **82**, pp. 1389–1398 (1999).
[2] D. M. Barbano and J. M. Lynch, "Major advances in testing of dairy products: Milk component and dairy product attribute testing," J. Dairy Sci. **89**, 1189–1194 (2006).
[3] D. J. McClements, "Advances in the application of ultrasound in food analysis processing," Trends Food Sci. Technol. **6**, 293–299 (1995).
[4] "Statement of EFSA on risks for public health due to the presences of melamine in infant milk and other milk products in China," The EFSA Journal **807**, 1–10 (2008).
[5] L. Elvira, C. Durán, C. Sierra, P. Resa, and F. Montero de Espinosa, "Ultrasonic measurement device for the characterization of microbiological and biochemical processes in liquid media," Meas. Sci. Technol. **18**, 2189–2196 (2007).
[6] V. A. Del Grosso and C. W. Mader, "Speed of sound in pure water," J. Acoust. Soc. Am. **52**, 1442–1446 (1972).

[7] D. Schneditz, H. Pogglitsch, J. Horina, and U. Binswanger, "A blood protein monitor for the continuous measurement of blood volume changes during hemodialysis," Kidney Int. **38**, 342–346 (1990).
[8] S. A. Jacobson, "New developments in ultrasonic gas analysis and flowmetering," Proceedings of 2008 International Ultrasonics Symposium (IEEE, New York, 2008), pp. 508–516.
[9] L. C. Lynnworth, "Noninvasive measurement of fluid characteristics using reversibly deformed conduit," U.S. Patent No. 7,481,114 (27 January 2009).

Microbubble tunneling in gel phantoms

Charles F. Caskey and Shengping Qin
Department of Biomedical Engineering, University of California, Davis, 451 East Health Sciences Drive,
Davis, California 95616
cfcaskey@ucdavis.edu, spqin@ucdavis.edu

Paul A. Dayton
Joint Department of Biomedical Engineering, University of North Carolina–North Carolina State University,
Chapel Hill, North Carolina 27514
padayton@bme.unc.edu

Katherine W. Ferrara
Department of Biomedical Engineering, University of California, Davis, 451 East Health Sciences Drive,
Davis, California 95616
kwferrar@ucdavis.edu

Abstract: Insonified microbubbles were observed in vessels within a gel with a Young's modulus similar to that of tissue, demonstrating shape instabilities, liquid jets, and the formation of small tunnels. In this study, tunnel formulation occurred in the direction of the propagating ultrasound wave, where radiation pressure directed the contact of the bubble and gel, facilitating the activity of the liquid jets. Combinations of ultrasonic parameters and microbubble concentrations that are relevant for diagnostic imaging and drug delivery and that lead to tunnel formation were applied and the resulting tunnel formation was quantified.

© 2009 Acoustical Society of America

PACS numbers: 43.80.Gx [TM]
Date Received: December 5, 2008 **Date Accepted:** February 2, 2009

Our laboratory and others have previously shown that upon insonation of a region of interest, microbubble oscillations within vessels with diameters as large as 55 μm can substantially increase vascular permeability and cause local hemorrhage.[1–3] Recent studies have helped elucidate the parameter space for enhancing permeability and avoiding injury by indirect observation of biological effects (magnetic resonance imaging signal enhancement[4] and hemorrhage counts[5]). The current work examines a similar parameter space but focuses on direct observation of microbubble oscillations and the resulting disruption of a gel that has a Young's modulus similar to soft tissues. As an initial effort to identify the mechanisms of gel disruption, bubble oscillation characteristics are then compared to theoretical predictions of bubble oscillation and collapse using simplified models.

Microbubbles in a gel flow phantom were observed during a 20-s insonation with frequencies of 1, 2.25, and 5 MHz using one of two combinations for pulse repetition frequency (PRF) and pulse duration (PD), chosen to simulate either common imaging ($\text{PD}_{\text{image}}=10\ \mu\text{s}$, $\text{PRF}_{\text{image}}=10$ kHz) or drug delivery parameters ($\text{PD}_{\text{delivery}}=10$ ms, $\text{PRF}_{\text{delivery}}=10$ Hz) with matched time-averaged intensity. The center frequencies and peak negative pressure (PNP) examined were chosen to investigate the ranges used for drug delivery (0.26–2.5 MHz, PNP >0.64 MPa).[2,6]

The ultrasound system consists of an ultrasound source, spherically focused and aligned so that the acoustic field and observation area overlapped. The gel phantom was a small block ($30\times20\times2$ mm^3) of 0.75% (w/v) OmniPur agarose gel (EM Science, Gibbstown, NJ) with an embedded 230-μm channel created by a heating and cooling process.[7] The channel size is the smallest diameter that can be easily perfused with microbubbles within our gel phantom. At 0.75% (w/v), the phantom is estimated to have a Young's modulus in the range of 10–20 kPa, which is similar to soft tissues in biological systems, such as the kidneys, liver, and muscle.[7,8]

Lipid-shelled microbubbles were made using techniques described previously.[9] The concentration of microbubbles was approximately 1.5×10^{10} bubbles/ml, with a mean diameter of 1.7 ± 1.6 μm. The solution was diluted in distilled water so that the final solution ranged from the dosage used for diagnostic therapy ($\sim 1.6 \times 10^5$ bubbles/ml) to high dosages used in many drug delivery experiments ($\sim 2.5 \times 10^7$ bubbles/ml). A perfusion pump was set so that the flow rate within the vessel was approximately 36 mm/s. After insonation, a solution of blue 500-nm microbeads (Polysciences, Warrington, PA) was injected to delineate wall boundaries within the gel. High-speed strobe images were captured by synchronizing a pulsed copper vapor laser (30-ns pulse width) to a camera with a 10-kHz frame rate. The laser flash created an effective exposure time of 30 ns to capture microbubble oscillation activity during the time that the shutter was open (0.1 ms). The laser was pulsed at 9.997 kHz to achieve an incremental delay of 30 ns during the 10-ms acoustic pulse, which allowed us to visualize the bubble oscillation during the positive and negative cycles of the acoustic pulse.[10] Disruption of the vessel wall was quantified by measuring the width and depth of tunnels formed. Bubble diameter was measured after the onset of tunnel formation and used as an input to a model of bubble oscillation. In order to correlate the tunnel width to the bubble oscillation, a Rayleigh–Plesset-based model[11] was used to predict the maximum bubble expansion. A similar form of the Rayleigh–Plesset-based equation has been used to simulate a bubble surrounded by a soft gel during insonation.[12] For simplicity, the terms representing the elastic/plastic properties of the gel have been neglected in our simulation.

The apparent threshold for tunnel formation for a given combination of transmission frequency and bubble concentration was determined by incrementing the PNP by 200 kPa until tunnel formation began. Threshold experiments were repeated five times for each combination of frequency and concentration to ensure accurate measurement. The effect of angle of insonation was analyzed by measuring the average direction of tunnel formation relative to ultrasound propagation. All image measurements were performed with IMAGEJ (NIH, http://rsb.info.nih.gov/ij/). Statistical analysis of tunnel widths was performed with the Student's *t*-test ($p = 0.05$ indicates significance).

Figures 1(a)–1(c) show representative examples of tunnels formed in the gel during insonation at 1, 2.25, and 5 MHz, respectively, with a matched mechanical index (MI) of 1.5, PD of 10 ms, and a high bubble concentration (2.5×10^7 bubbles/ml). The axis of the ultrasound beam was vertical and directed upward. The average widths of tunnels created by the bubbles during insonation at 1, 2.25, and 5 MHz were 39.7 ± 6.8, 21.8 ± 2.3, and 7.4 ± 1.5 μm, respectively. The decrease in tunnel width with increasing frequency was expected due to the decreased maximum microbubble diameter during oscillation. The maximum depth of tunnels created by bubbles in these experiments increased with decreasing ultrasound center frequency; the depths of the tunnels were 1.2 ± 0.52, 0.36 ± 0.16, and 0.15 ± 0.09 mm with center frequencies of 1, 2.25, and 5 MHz, respectively.

Asymmetrical oscillation after tunnel formation has begun was visualized in sequential high-speed strobe images acquired with high magnification [Figs. 1(d)–1(g)], using a center frequency of 1 MHz, a PNP of 1.2 MPa, the beam-vessel axis as in Figs. 1(a)–1(c), and a high microbubble concentration (2.5×10^7 bubbles/ml). With these parameters, the PNP of 1.2 MPa is the threshold for tunnel creation in the agarose gel and is similar to the PNP previously determined during *in vivo* studies.[1] In the first image (1d), two microbubble clusters were shown just before coalescence. At a later point in time (1e), the recently formed microbubble has entered the tunnel. The bubble compressed to an undetectably small diameter during the positive phase of the ultrasonic pulse (1f). During a subsequent negative acoustic pressure phase (1g) the microbubble expanded to a diameter equal to the tunnel width. In this instance, the microbubble expanded to a diameter of 45 μm and translated a distance of 1.2 μm down the tunnel over the course of 10 ms (i.e., at a velocity of 1.2 mm/s). Since we observe that the microbubbles fully fill the tunnels at peak expansion, the tunnel width was representative of peak expansion, while the tunnel depth indicates repeated oscillation and persistence during a long ultrasonic pulse.

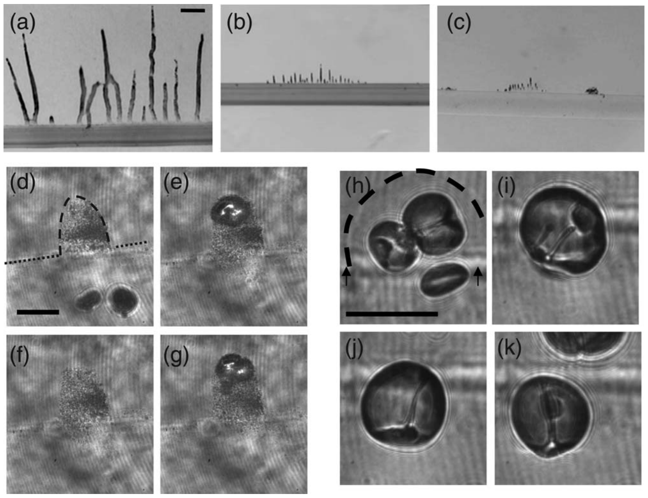


Fig. 1. Tunnel images and high-speed microscopy. Optical evidence of tunnel formation using a microbubble concentration of $\sim 2.5 \times 10^7$ bubbles/ml and pulse duration of 10 ms. Tunnels created from insonation at (a) 1-MHz, (b) 2.25-MHz, and (c) 5-MHz ultrasonic pulses at matched MI of 1.5. Scale bar indicates 250 μm. Sequential high-speed images (d)–(g) show a microbubble oscillating during a 1-MHz pulse creating a 45-μm-diameter tunnel. The frame rate is 10 kHz and the first image is acquired at 0.1 ms after the onset of insonation. Subsequent images were selected to show compressional and rarefactional half-cycles. The microbubble forms in (d) from the fusion of multiple microbubble fragments and then oscillates asymmetrically as it moves through the gel. Scale bar indicates 50 μm. High-speed images (h)–(k) show multi-bubble interactions and fluid jets during the formation of a 50-μm tunnel. The frame rate is 10 kHz and the images are selected for clear jet visualization. Scale bar indicates 50 μm.

While Figs. 1(d)–1(g) show significant events that occur during tunnel formation, including coalescence and expansion, higher magnification images were acquired to visualize fluid jets. Liquid jets are visible within each microbubble in Figs. 1(h)–1(k), which are a sequence of images acquired as a bubble enters a newly-formed tunnel. The arrows in Fig. 1(h) point to the boundary of the gel wall, which is evident in Figs. 1(h)–1(k) as a diffuse light horizontal line. The dotted line in Fig. 1(h) indicates the boundary of the tunnel that is being formed. The same tunnel is present but out of the frame in Figs. 1(i)–1(k). The images are similar to observations of fluid jets created by much larger bubbles (with diameters on the order of millimeters) that have been shown to disrupt gel boundaries.[13] The jets formed in various directions, as compared with the beam axis and surface normal, with jet neck diameters, r_j, ranging from 2 to 15 μm. Jets formed both in the presence [Figs. 1(h) and 1(k)] and absence [Figs. 1(i) and 1(j)] of surrounding bubbles and were observed to traverse the entire bubble diameter in some cases. As shown previously by other researchers, the wide range in jet diameter and direction is expected due to varied bubble size, the presence or absence of nearby bubbles, and the location of the bubble relative to the vessel wall during the acoustic pulse.[14] The pressure of a fluid jet impacting on a compressible solid boundary is

$$P_{WH} = \frac{\rho_1 c_1 \rho_2 c_2 v}{\rho_1 c_1 + \rho_1 c_1}, \qquad (1)$$

where v is the velocity of the jet and $\rho_{1,} c_1$ and $\rho_{2,} c_2$ are the products of the densities and speed of sound in the water and gel boundary, respectively.[15] The duration of the jet impact is estimated to be

$$\tau_j = \frac{r_j}{c}. \tag{2}$$

Assuming that the density and speed of sound are not significantly different between the gel and water, we take the densities to be 997.99 kg m^{-3} and the sound velocities as 1485.9 m s^{-1}. For jet velocities similar to predicted wall velocity at collapse, the jet impact pressure will be in the range of hundreds of megapascals, with impact lasting between 1.3 and 10 ns. This pressure far exceeds the tensile strength of the gel which is about 0.056 MPa for a similar gel,[12] and therefore the observed jets are expected to play a role in tunnel formation.

The effect of PNP was next evaluated for center frequencies of 1 and 2.25 MHz at two concentrations (PD=10 ms and PRF=10 Hz). Tunnels were not observed for a PNP of 1 MPa or below for a center frequency of 1 MHz at either concentration. The PNP threshold for tunnel digging at 1 MHz was 1.2 MPa which is comparable to the regime where Prentice *et al.*[16] observed microjet phenomenon. Tunnels were not observed with insonation at 2.25 MHz using the contrast agent concentration used for diagnostic imaging. At the higher concentration and with insonation at 2.25 MHz, a trend of increasing tunnel width with increasing PNP was observed. With this higher concentration, the PNP threshold for tunnel creation was 0.6 MPa for the 2.25-MHz center frequency.

The higher concentration employed in the 2.25-MHz studies resulted in microbubble coalescence during the long pulses. After insonation, the diameters of microbubbles responsible for tunnel formation were optically measured, yielding a mean diameter of 9.3 ± 3.4 μm (where the mean diameter was ~1.7 μm before coalescence). Smaller bubbles typically dissolved during the ultrasonic pulse. Larger microbubbles (>10 μm) were observed to form from coalescence but oscillated with a low amplitude within the tunnel and did not disrupt the gel.

Predictions for the maximum diameter achieved during insonation at 1 and 2.25 MHz, respectively, for a microbubble with a diameter of 9.3 μm are shown by the dashed lines [Fig. 2(a)]. The maximum diameter of the tunnels formed increases with PNP, peaking at 47 and 25 μm for the 1- and 2.25-MHz center frequencies, respectively. While the inception of tunnel formation was a result of jet formation, the width of the tunnels approaches the width associated with maximum expansion (as shown in Fig. 1).

Decreasing the ultrasound center frequency increased tunnel width for a constant MI, as shown in Fig. 2(b) where the tunnel width was averaged over five observations for each center frequency tested (MI=1.2, 1.5). Thus, tunnel width exhibits a stronger dependence on transmission center frequency than that predicted by MI, as previously noted.[1,17] The predicted maximum bubble diameters during peak expansion for MI values of 1.2 and 1.5 [solid and dotted lines in Fig. 2(b)] indicate that the tunnel width was similar to the expected diameter of oscillating microbubbles.

When the PD was decreased from 10 ms to 10 μs (while increasing the PRF from 10 Hz to 10 kHz so that the time-averaged acoustic intensity was matched), the threshold for disrupting the gel wall increased from 1.2 to greater than 2.5 MPa at both low and high bubble concentrations. With the shorter pulse, diffusion of the gas bubble into the surrounding liquid may occur prior to the next pulse. A thorough investigation of the effect of pulse duration on gel disruption is underway.

Increasing microbubble concentration decreases the PNP threshold for gel disruption at transmission frequencies greater than or equal to 2.25 MHz (PRF of 10 Hz and PD of 10 ms) [Fig. 2(c)]. At 1 and 1.5 MHz, the threshold for gel disruption was similar across all concentrations investigated in our study. The effect on the gel wall became undetectable at diagnostic bubble concentrations (1.5×10^5 bubbles/ml) using transmission frequencies above 1.5 MHz for PNPs as high as 5 MPa.

In all cases examined, tunnels were formed on the distal side of the vessel relative to the ultrasound transducer; vessel wall damage was never observed on the side of the vessel proximal to the transducer. While in contact with the wall, the bubbles moved into the gel in the direction of the ultrasound propagation. The direction of the tunnel formation was the same as

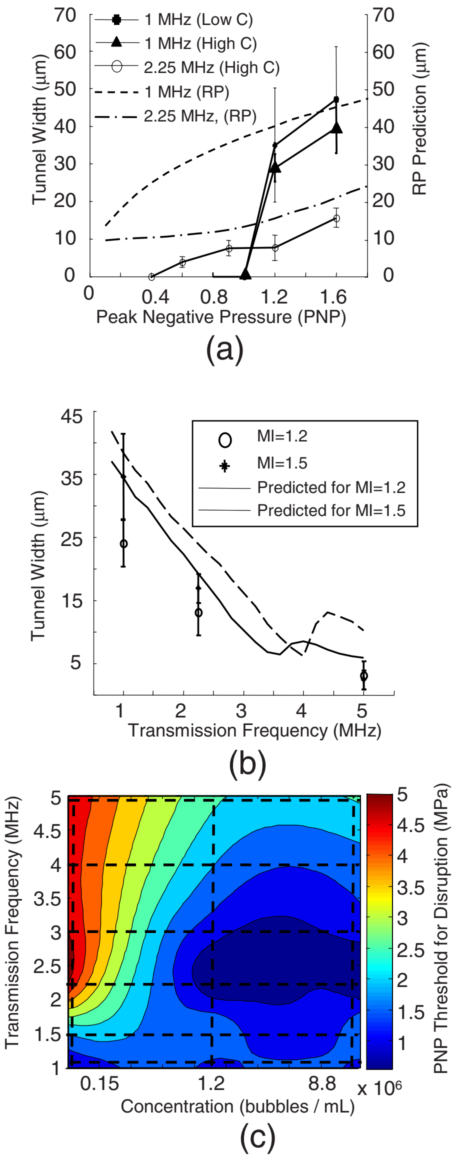


Fig. 2. (Color online) Tunnel formation and parameters. (a) Tunnel width for two transmission frequencies (1 and 2.25 MHz) and two concentrations (1.5×10^5 and 2.5×10^7 bubbles/ml) with a pulse duration of 10 ms; Rayleigh–Plesset predictions for maximum diameter of a microbubble in an infinite fluid. Tunnel width increases with PNP for 1 and 2.25 MHz at high bubble concentration, while tunnel formation at low concentration was only possible at 1 MHz. The lower threshold is indicated by the point where tunnel width is zero. At the higher pressures, tunnel width approaches the maximum diameter predicted by a 9.3-μm bubble (the diameter observed to interact with the wall). (b) Average tunnel width for three center frequencies using a matched MI, pulse duration of 10 ms, and microbubble concentration of 2.5×10^7 bubbles/ml. The solid and dashed lines in the plot indicate the predicted size of a 9.3-μm bubble according to simulation. (c) Surface fit indicating thresholds for tunnel formation as a function of frequency and bubble concentration. Intersection of dotted lines indicates a frequency and concentration combination tested.

that of ultrasound propagation, regardless of the angle of the vessel [Fig. 3(a)]. At vessel angles of 90, 75, and 60 degrees relative to the beam axis, the orientation of the tunnels formed was consistent with the direction of ultrasound propagation [Fig. 3(b)]. Primary radiation force maintains the bubble's contact with the wall and translates the bubble in the direction of ultra-

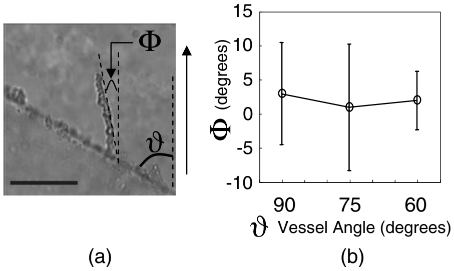


Fig. 3. Varying the beam-vessel angle. (a) With 2.25-MHz insonation (PNP of 1.2 MPa), pulse duration of 10 ms, and concentration of 2.5×10^7 bubbles/ml, representative image of tunnel formation vs beam direction indicated by arrow and vessel wall orientation shown as diagonal wall. (b) Summary of tunnel angle, Φ, and vessel angle, ϑ, each relative to the ultrasound beam for conditions summarized in (a). For vessel angles as small as 60° relative to the beam axis, the tunnel consistently forms in the same direction as the ultrasonic pulse.

sonic beam. We hypothesize that tunnel creation was likely due to a combination of fluid jets, bubble expansion, and primary radiation force during bubble oscillation.

Dayton *et al.*[18] calculated the magnitude of primary radiation force to be approximately 1×10^{-5} N for a bubble with an initial radius of 1.63 μm during a 2.25-MHz pulse (PNP =100 kPa). In the present study, the PNP and PD are both substantially larger and the local stress due to radiation force is higher and the radiation force deflects bubbles along the beam axis (Fig. 3). After tunnel formation has begun, the microbubble is constrained by two nearby boundaries and its oscillation could expand against the surrounding tunnel, which is similar to a scenario previously examined,[17] where bubble expansion results in circumferential stress that can contribute to breakdown of the gel boundary.

The studies reported here demonstrate multiple factors that must be considered when designing safe diagnostic and drug delivery procedures using ultrasound contrast agents, including microbubble concentration, center frequency, acoustic pressure, and pulse duration. Observations reported here are similar to those reported previously for an *ex vivo* preparation, while allowing for easier quantification of bubble activity and associated acoustic parameters.[10] In the present study, a set of parameters was observed that does not result in tunnel formation, including a center frequency of 2.25 MHz (the threshold frequency was between 1 and 2.25 MHz) with a diagnostic concentration of small-lipid shelled microbubbles and a pressure up to 5 MPa. Alternatively, a high concentration of microbubbles and long pulse can produce an effect even with the 5-MHz center frequency. When the frequency was reduced to 1 MHz, tunnel formation was observed with both the diagnostic and therapeutic concentrations of microbubbles for either insonation with a long pulse (high duty cycle) and PNP above 1 MPa or a short (imaging pulse) and PNP above 2.5 MPa.

As in all phantom studies, this study has limitations that must be recognized in the interpretation of our results. First, while we seek to investigate a phantom capillary using a well-known concept (a tunnel within a gel), due to the logistics of the experiments, the vessel was larger than a typical capillary. Further, the basement membrane of a vessel may limit the depth of tunneling within a vessel. Alternatively, the studies facilitate an examination of the nucleation of small pores or tunnels and the interaction of the primary radiation pressure and microbubble collapse resulting in the translation of the microbubble in the direction of the radiation pressure.

Acknowledgments

The authors would like to acknowledge funding from NIH R01CA103828 and R01CA112356.

References and links

[1] S. M. Stieger, C. F. Caskey, R. H. Adamson, S. Qin, F. R. Curry, E. R. Wisner, and K. W. Ferrara, "Enhancement of vascular permeability with low-frequency contrast-enhanced ultrasound in the chorioallantoic membrane model," Radiology **243**, 112–121 (2007).

[2] D. L. Miller and J. Quddus, “Diagnostic ultrasound activation of contrast agent gas bodies induces capillary rupture in mice,” Proc. Natl. Acad. Sci. U.S.A. **97**, 10179–10184 (2000).
[3] R. J. Price, D. M. Skyba, S. Kaul, and T. C. Skalak, “Delivery of colloidal particles and red blood cells to tissue through microvessel ruptures created by targeted microbubble destruction with ultrasound,” Circulation **98**, 1264–1267 (1998).
[4] N. McDannold, N. Vykhodtseva, and K. Hynynen, “Effects of acoustic parameters and ultrasound contrast agent dose on focused-ultrasound induced blood-brain barrier disruption,” Ultrasound Med. Biol. **34**, 930–937 (2008).
[5] D. L. Miller, C. Dou, and R. C. Wiggins, “Frequency dependence of kidney injury induced by contrast-aided diagnostic ultrasound in rats,” Ultrasound Med Biol. **34**, 1678–1687 (2008).
[6] K. Hynynen, N. McDannold, N. Vykhodtseva, S. Raymond, R. Weissleder, F. A. Jolesz, and N. Sheikov, “Focal disruption of the blood-brain barrier due to 260-kHz ultrasound bursts: A method for molecular imaging and targeted drug delivery,” J. Neurosurg. **105**, 445–454 (2006).
[7] V. Normand, D. L. Lootens, E. Amici, K. P. Plucknett, and P. Aymard, “New insight into agarose gel mechanical properties,” Biomacromolecules **1**, 730–738 (2000).
[8] V. Egorov, S. Tsyuryupa, S. Kanilo, M. Kogit, and A. Sarvazyan, “Soft tissue elastometer,” Med. Eng. Phys. **30**, 206–212 (2008).
[9] M. A. Borden, D. E. Kruse, C. F. Caskey, S. K. Zhao, P. A. Dayton, and K. W. Ferrarra, “Influence of lipid shell physicochemical properties on ultrasound-induced microbubble destruction,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1992–2002 (2005).
[10] C. Caskey, S. Stieger, S. Qin, P. Dayton, and K. W. Ferrarra, “Direct observation of microbubble interaction with the endothelium,” J. Acoust. Soc. Am. **122**(2), 1191–1200 (2007).
[11] K. E. Morgan, J. S. Allen, P. A. Dayton, J. E. Chomas, A. L. Klibaov, and K. W. Ferrarra, “Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1494–1509 (2000).
[12] E. A. Brujan, and A. Vogel, “Stress wave emission and cavitation bubble dynamics by nanosecond optical breakdown in a tissue phantom,” J. Fluid Mech. **558**, 281–308 (2006).
[13] T. Kodama and Y. Tomita, “Cavitation bubble behavior and bubble-shock wave interaction near a gelatin surface as a study of in vivo bubble dynamics,” Appl. Phys. B: Lasers Opt. **70**, 139–149 (2000).
[14] E. A. Brujan, G. S. Keen, A. Vogel, and J. R. Blake, “The final stage of the collapse of a cavitation bubble close to a rigid boundary,” Phys. Fluids **14**, 85–92 (2002).
[15] J. Brunton, “High speed liquid impact,” Philos. Trans. R. Soc. London, Ser. A **260**, 79–85 (1966).
[16] P. Prentice, A. Cuschieri, K. Dholakia, M. Prausnitz, and P. Campbell, “Membrane disruption by optically controlled microbubble cavitation,” Nat. Phys. **1**, 107–110 (2005).
[17] S. Qin and K. W. Ferrara, “Acoustic response of compliable microvessels containing ultrasound contrast agents,” Phys. Med. Biol. **51**, 5065–5088 (2006).
[18] P. A. Dayton, J. S. Allen, and K. W. Ferrara, “The magnitude of radiation force on ultrasound contrast agents,” J. Acoust. Soc. Am. **112**, 2183–2192 (2002).

Real-time algorithm for acoustic imaging with a microphone array

Xun Huang
Department of Mechanical and Aerospace Engineering, College of Engineering, Peking University,
Beijing, 100871, China
huangxun@pku.edu.cn

Abstract: Acoustic phased array has become an important testing tool in aeroacoustic research, where the conventional beamforming algorithm has been adopted as a classical processing technique. The computation however has to be performed off-line due to the expensive cost. An innovative algorithm with real-time capability is proposed in this work. The algorithm is similar to a classical observer in the time domain while extended for the array processing to the frequency domain. The observer-based algorithm is beneficial mainly for its capability of operating over sampling blocks recursively. The expensive experimental time can therefore be reduced extensively since any defect in a testing can be corrected instantaneously.

© 2009 Acoustical Society of America

PACS numbers: 43.60.Fg, 43.60.Lq, 43.60.Mn [JC]
Date Received: January 27, 2009 **Date Accepted:** February 19, 2009

1. Introduction

Acoustic phased array[1] has become an important tool in wind tunnel tests[2] to identify the main noise source of an aircraft model[3,4] with the objective of developing a quieter design[5] that can reduce environmental impact. The conventional beamforming algorithm[6] performed in the frequency domain[1] was normally adopted as the fundamental processing method. The more advanced techniques, which include multiple signal classification[7] and robust adaptive beamforming,[8] worked unsatisfactorily in a noisy wind tunnel.[9]

The conventional beamforming algorithm has to be operated off-line due to the expensive cost in the computation of cross power matrix, which was computed by averaging over many sampling blocks. To achieve real-time computation a new recursive algorithm is presented in this paper. The derivation of the algorithm is based on the classical state observer in linear control theory.[10] The observer-based algorithm performed in real-time will be quite helpful for wind tunnel tests since any defect in a test setup and experiments can be found and restored instantaneously. The expensive experimental time in a wind tunnel can therefore be reduced extensively.

The rest of this paper is organized as follows. Section 2 briefly introduces the conventional beamforming algorithm in wind tunnel tests. The classical observer method is summarized in Sec. 3, followed by the development of the real-time algorithm that works recursively over each block of noisy wind tunnel signals. Section 4 presents a summary of this work.

2. Fundamentals of beamforming

The beamforming algorithm for aeroacoustic testing is proceeded by firstly sampling the time-domain signal of acoustic pressure $y_m(t)$ from the mth microphone of a phased array through a data acquisition system. The sampling frequency is typically 40 kHz or higher. The time series of the digital samples is subsequently cut into blocks. The discrete Fourier transform (DFT) is performed over each block of $y_m(t)$ to produce the counterpart in the frequency domain, that is, $Y_m(\omega)|_k$ for kth block at the angular frequency of ω. The size of each block is typically a multiple of 4096 for the efficient computation of the DFT. In case of no confusion ω will be omitted in all following equations for conciseness.

For a single noise source X at the position of ξ, the measurement satisfies

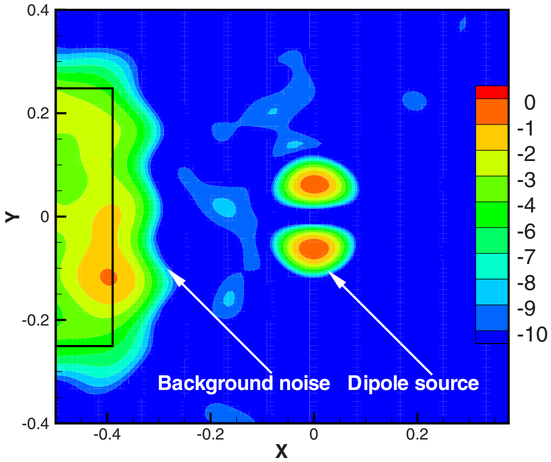


Fig. 1. (Color online) An acoustic source in a noisy testing facility.

$$\mathbf{Y}|_k = \mathbf{G}X|_k, \quad (1)$$

where $\mathbf{Y}=[Y_1,\ldots,Y_n]'$, $\mathbf{G}$ steering vector matrix, $\mathbf{G}=[G_1,\ldots,G_n]' \in \mathbb{C}^{n\times 1}$, the symbol of $'$ denotes transpose, G_m is the steering vector that depends on the relative position between the noise source X and the mth microphone,[5] and the value of n is the overall number of microphones in the phased array. In this work $n=56$. The energy spectral density (E) of X can be approximated by the conventional beamforming[5]

$$E = \mathbf{G}^*\mathbf{C}\mathbf{G}/\|\mathbf{G}\|^4, \quad (2)$$

where the symbol of * denotes complex transpose, $\mathbf{C}$ is the cross-spectral matrix, $\mathbf{C} = \langle \mathbf{Y}|_k \mathbf{Y}^*|_k\rangle$, and $\langle\,\rangle$ means the operation of averaging over a number of blocks. The sound pressure level (SPL) of the noise source X at the position of ξ in decibels is $P=10\log_{10}(E/(4\times 10^{-10}))$. An acoustic image for an aircraft model can thereafter be produced by performing the beamforming method over every ξ of an area.

One of the salient features in aerospace tests is the severe background noise from a wind tunnel facility. Figure 1 is an acoustic image of P at 5 kHz. The image includes a background noise that was measured from a tunnel with an open test section. The value of P is nondimensionalized to the maximal SPL in the domain. The rectangle in Fig. 1 represents the exit of the tunnel nozzle, where severe background noise can be found. The conventional beamforming algorithm in aeroacoustic research follows Welch's method[11] and can be proceeded in the following steps to remedy the effect of background noise.

Step 1. A measurement is performed at the intended flow speed without the installation of aircraft model. The Fourier transformed outcome $\mathbf{Y}_B$ represents the sole result from the background noise.

Step 2. A measurement is performed again at the same flow speed after an aircraft model is installed. The Fourier transformed outcome $\mathbf{Y}_{BS}$ is the collective results from the acoustic source of the model and the background noise.

Step 3. Compute the cross-spectral matrix for $\mathbf{Y}_B$ and $\mathbf{Y}_{BS}$, respectively, producing $\mathbf{C}_B$ and $\mathbf{C}_{BS}$. With the assumption of little coherence[12] between the acoustic source of the model and the background noise, the cross-spectral matrix for the model acoustic source can be approximated by $\mathbf{Y}_S=\mathbf{Y}_{BS}-\mathbf{Y}_B$. As a result the interference from the background noise source X_B can be minimized. Equation (2) is thereafter employed to obtain the amplitude of the acoustic source X_S.

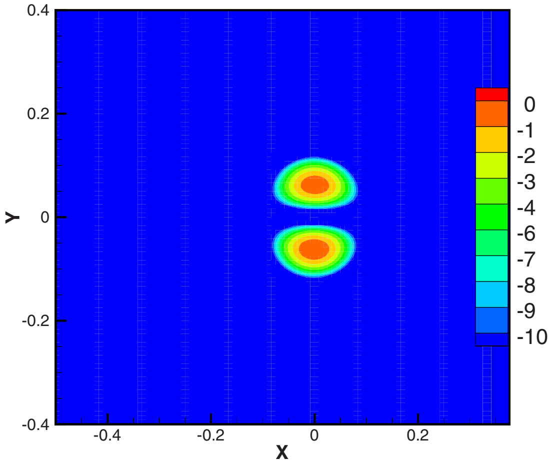


Fig. 2. (Color online) Acoustic image by using the beamforming algorithm.

Step 4. Operate Eq. (2) continuously over a plane to find X_S at different positions ξ to generate an acoustic image.

A simulated dipole source produced from an analytical solution was used as a benchmark in this work for the testing of the algorithms. The dipole source is superimposed onto the background noise which was measured in a wind tunnel. Figure 1 shows the dipole source that represents typical acoustic sources from an aircraft model, along with the severe background noise mainly from a noisy fan that produces the testing flow of a free stream at 30 m/s.

Figure 2 is the acoustic imaging results obtained by performing the beamforming algorithm over $K=100$ blocks. It can be seen that the interference from the background noise (mainly from the fan) is taken off satisfactorily. The beamforming algorithm is not real-time since the computation of the cross-spectral matrix has to be operated over many blocks. Excluding the cost of DFT, the rest cost of matrix multiplications over K blocks from total n microphones at single ξ and single ω is proportional to Kn^2.

3. Innovative real-time algorithm

An innovative algorithm approximating the SPL of an acoustic source recursively is proposed in this work to achieve real-time performance. The origin of the algorithm is from the method of state observer in the linear control theory.[10] It is worth mentioning that the original observer was presented for signals in the time domain.[10] The idea was modified in this work to approximate the energy spectral density from data in the frequency domain. As a result, the observer-based algorithm proposed below is different from the existing recursive algorithms of adaptive beamforming that were normally operated in the time domain.[13]

From the perspective of the linear system theory, the state equation and the measurement equation of the acoustic testing in the frequency domain can be written as

$$X|_{k+1} = AX|_k, \tag{3}$$

$$\mathbf{Y}|_k = \mathbf{G}X|_k. \tag{4}$$

Subsequently a new matrix can be defined as

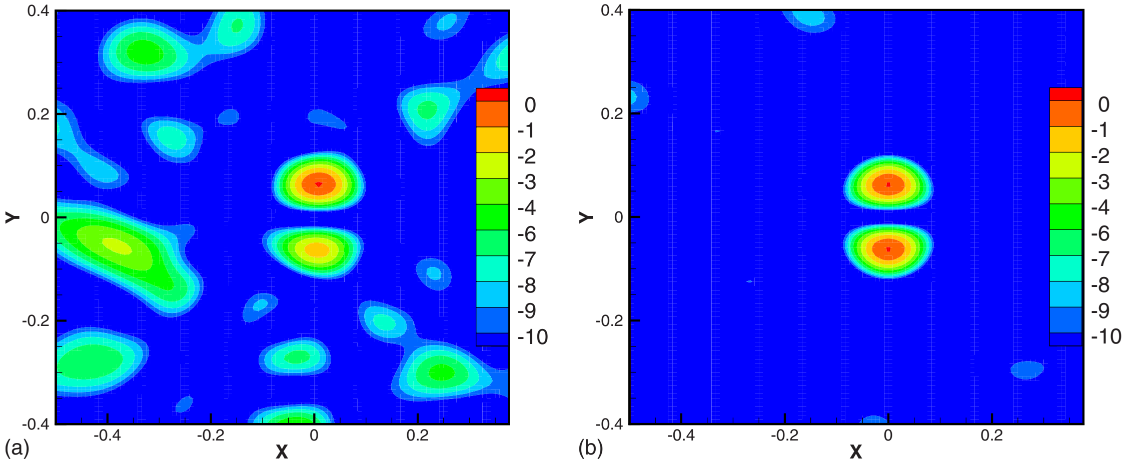


Fig. 3. (Color online) Acoustic image by using state observer, where the results are from (a) $\hat{X}_{BS}|_k$ and (b) $\hat{X}_S|_k$, $k=10$.

$$\mathbf{O} = \begin{bmatrix} \mathbf{G} \\ \mathbf{G}A \\ \vdots \\ \mathbf{G}A^{n-1} \end{bmatrix}, \tag{5}$$

where A is the state matrix that equals an identity matrix as long as the measured signal is stationary. The rest of the variables have been defined previously. As the rank of $\mathbf{O}$ equals n, all states X in Eq. (3) are observable[10] from the microphone measurements. An observer can consequently be designed to approximate X from the measurements $\mathbf{Y}$. The observer[10] has the form of

$$\hat{X}|_{k+1} = A\hat{X}|_k + \mathbf{L}(\mathbf{Y}|_k - \hat{\mathbf{Y}}|_k), \tag{6}$$

$$\hat{\mathbf{Y}}|_k = \mathbf{G}\hat{X}|_k, \tag{7}$$

where the symbol of ˆ denotes the estimation of the states by the observer, and $\mathbf{L}$ is the observer gain. By subtracting Eq. (6) to Eq. (3), the error between the observation and the real signal at block k, $e|_{k+1} = \hat{X}|_{k+1} - X|_{k+1}$, satisfies

$$e|_{k+1} = (A - \mathbf{LG})\, e|_k. \tag{8}$$

As a result, the error e converges to zero when $k \to \infty$, as far as all eigenvalues of the matrix $(A - \mathbf{LG})$ are within a unit circle.[10] Normally proper eigenvalues are firstly assigned and the observer gain $\mathbf{L}$ is solved accordingly. The computation of $\mathbf{L}$ can be performed off-line and the results can be stored in a table for the real-time computation of Eqs. (6) and (7).

Figure 3(a) shows the result of $\hat{X}_{BS}|_k$, which is obtained by performing Eqs. (6) and (7) recursively from the first block to the tenth block of the Fourier transformed outcome, $\mathbf{Y}_{BS}$. In this work the eigenvalues of $(A - \mathbf{LG})$ were set to -0.5, for instance. Figure 3(a) suggests that the observer has the ability to reconstruct the dipole source. However, the acoustic image is still interfered by the background noise. To eliminate the effect of the background noise, the original state equations [Eqs. (3) and (4)] are modified to describe background noise and the acoustic source, respectively. The new state equations are

$$\begin{bmatrix} X_B|_{k+1} \\ X_S|_{k+1} \end{bmatrix} = \begin{bmatrix} A & \\ & A \end{bmatrix} \begin{bmatrix} X_B|_k \\ X_S|_k \end{bmatrix}, \tag{9}$$

$$\begin{bmatrix} \mathbf{Y}_B|_k \\ \mathbf{Y}_{BS}|_k \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \\ \mathbf{G}e^{i\phi} & \mathbf{G} \end{bmatrix} \begin{bmatrix} X_B|_k \\ X_S|_k \end{bmatrix}. \tag{10}$$

Basically two sets of experiments are required to obtain $\mathbf{Y}_B$ and $\mathbf{Y}_{BS}$. The first experiment measured the background noise when an aircraft model is not installed in the test section. The second experiments measured the background noise plus the acoustic noise of the model. The symbol of ϕ denotes the phase difference between the background noise in the two experiments. The exact value of ϕ can be estimated by placing an extra sensor beside the dominant generator of the background noise.

The corresponding equations of the state observer are

$$\begin{bmatrix} \widehat{X_B}|_{k+1} \\ \widehat{X_S}|_{k+1} \end{bmatrix} = \begin{bmatrix} A & \\ & A \end{bmatrix} \begin{bmatrix} \widehat{X_B}|_k \\ \widehat{X_S}|_k \end{bmatrix} + \mathbf{L}\left(\begin{bmatrix} \mathbf{Y}_B|_k \\ \mathbf{Y}_{BS}|_k \end{bmatrix} - \begin{bmatrix} \mathbf{G} & \\ \mathbf{G}e^{i\phi} & \mathbf{G} \end{bmatrix} \begin{bmatrix} \widehat{X_B}|_k \\ \widehat{X_S}|_k \end{bmatrix} \right), \tag{11}$$

$$\begin{bmatrix} \widehat{\mathbf{Y}_B}|_k \\ \widehat{\mathbf{Y}_{BS}}|_k \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \\ \mathbf{G}e^{i\phi} & \mathbf{G} \end{bmatrix} \begin{bmatrix} \widehat{X_B}|_k \\ \widehat{X_S}|_k \end{bmatrix}. \tag{12}$$

The computational cost of Eqs. (11) and (12) is proportional to n^2 for each block. After finishing the acquisition of one block of data, Eqs. (11) and (12) can be applied to the block. In the meantime of the computation, the data acquisition system starts to obtain the next block of data. Although its overall cost is comparable to the cost of the conventional beamforming, the observer-based algorithm can be performed recursively in real-time. Figure 3(b) shows the outcome of acoustic image at tenth block. Compared to Fig. 3(a), the interference from the background noise is minimized clearly. The outcome also suggests a satisfactory convergence rate.

4. Summary

A new algorithm with the real-time capability was presented for phased microphone arrays in this work. The algorithm was proposed from the perspective of the linear system theory and is similar to a classical observer in the form. It is worthwhile mentioning that the idealized assumptions such as free space of sound propagation and little sensor noise were hold for both the conventional beamforming and the observer-based algorithm. The method of Kalman filter, which can be regarded as an extended observer with the constraint of Gaussian noise, can be used to include practical imperfections such as multi-path, reflection, and sensor noise in the state model. As an extension of the present work the related research is ongoing and beyond the scope of this paper.

In summary the present observer-based algorithm is able to be performed recursively over each sampling block. The outcome is still comparable to the corresponding beamforming result. The finding from the numerical experiment confirmed the working of the observer-based algorithm. Therefore, the expensive experimental time in a wind tunnel could be reduced extensively since any defect in a testing can be revealed and corrected instantaneously. The proposed algorithm should also be applicable to other areas, such as communications and ultrasonics.

References and links

[1] D. E. Dudgeon, “Fundamentals of digital array processing,” Proc. IEEE **65**, 898–904 (1977).

[2] H.-C. Shin, W. R. Graham, P. Sijtsma, C. Andreou, and A. C. Faszer, “Implementation of a phased microphone array in a closed-section wind tunnel,” AIAA J. **45**, 2897–2909 (2007).

[3] Y. W. Wang, J. Li, P. Stoica, M. Sheplak, and T. Nishida, “Wideband relax and wideband clean for aeroacoustic imaging,” J. Acoust. Soc. Am. **115**, 757–767 (2004).

[4] 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).

[5]T. J. E. Mueller, *Aeroacoustic Measurements* (Springer, Germany, 2002).
[6]B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**, 4–24 (1988).
[7]R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**, 276–280 (1986).
[8]J. Li and P. Stoica, *Robust Adaptive Beamforming* (Wiley-Interscience, New York, 2005).
[9]K. D. Donohue, J. Hannemann, and H. G. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments," Signal Process. **87**, 1677–1691 (2007).
[10]S. Eduardo, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 2nd ed. (Springer, Germany, 1998).
[11]P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms," IEEE Trans. Audio Electroacoust. **15**, 70–73 (1967).
[12]D. R. Morgan and T. M. Smith, "Coherence effects on the detection performance of quadratic array processors, with applications to large-array matched-field beamforming," J. Acoust. Soc. Am. **87**, 737–747 (1990).
[13]S. N. Jagadeesha, S. N. Sinha, and D. K. Mehra, "A recursive modified Gram-Schmidt algorithm based adaptive beamformer," Signal Process. **39**, 69–78 (1994).

Vowel-pitch matching in Wagner's operas: Implications for intelligibility and ease of singing

John Smith and Joe Wolfe
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia
john.smith@unsw.edu.au

Abstract: European vowels are mainly distinguished by the two lowest resonance frequencies ($R1$ and $R2$) of the vocal tract. Once the pitch frequency f_0 exceeds the value of $R1$ in normal speech, sopranos can deliberately "tune" $R1$ to match f_0. This increases loudness, uniformity of tone, and ease of singing, at some cost to intelligibility. Resonance tuning would be assisted if the pitch of the note written for a vowel corresponded with its usual range of $R1$. Analysis of several soprano roles indicates that Wagner aided the acoustics of the soprano voice at high pitch when setting text to music.

© 2009 Acoustical Society of America

PACS numbers: 43.75.Rs [DD]
Date Received: December 24, 2008 **Date Accepted:** February 26, 2009

1. Introduction

In normal speech, the vibrating vocal folds generate a harmonically rich signal with pitch frequency f_0, which interacts with resonances of frequency Ri in the vocal tract to produce sound with a spectral envelope that exhibits broad peaks, called formants,[1] with frequency Fi (Fant, 1970). These resonances have bandwidths of 100 Hz or so, and can be controlled independently of f_0 by varying the position and shape of the tongue, jaw, lips, and larynx (Lindblom and Sundberg, 1971). Vowels in European languages are largely identified by the frequencies of the first two formants ($F1,F2$) and thus of the resonances ($R1,R2$) that produce them. In adult conversational speech, f_0 is typically in the range 100–300 Hz, whereas the resonances $R1$ and $R2$ lie in the approximate ranges 300–800 and 800–2000 Hz, respectively. Consequently, f_0 and/or some of its harmonics (i.e., $2f_0$, $3f_0$, etc.) usually fall close enough to each resonance to receive a useful power boost and to produce identifiable formants. The situation is similar for most singing ranges when f_0 is less than about 500 Hz (near the "high C" of tenors). However, f_0 for sopranos can range from 250 to 1000 Hz or even higher.

Four problems can arise when f_0 exceeds significantly the normal range of $R1$ for a vowel. First, the acoustic load of the tract on the vocal folds and glottis changes from inertive ("mass-like") when $f_0 < R1$ to compliant ("spring-like") when $f_0 > R1$. Vocal fold vibration can then become less efficient and less stable (Titze 1988, 2008; Titze *et al.*, 2008).

Second, the sound level is usually reduced, as little acoustic energy will be radiated at frequency f_0. Third, a strong variation in the amplitudes of the fundamental and/or second harmonic may occur as the pitch changes, producing possibly undesirable discontinuities in timbre. Finally, the effective absence of $F1$ means that vowels with a similar value of $F2$ become indistinguishable. Indeed, as f_0 increases, the spacing of harmonics can become so large that even $F2$ may also effectively disappear.

For a soprano, a solution to all but the last of these problems is to tune $R1$ close to, but slightly above, f_0 (Sundberg, 1975; Joliveau *et al.*, 2004a, 2004b; Titze *et al.*, 2008). This should increase power, increase the ease of singing, and help maintain timbral homogeneity. Accordingly, sopranos often "modify" their vowels when singing at high pitch (e.g., Coffin, 1974, 1976).

This solution, called resonance tuning, has no disadvantages when singing *vocalise* because there is no textual information. However, when singing text, $F1$ will consequently be similar to $R1$ and no longer have an appropriate value for many phoneme-pitch combinations. This increases the probability that vowels are confused as the difference between f_0 and the value of $R1$ in normal speech increases (Morozov, 1965; Scotto di Carlo and Germain, 1985; Hollien *et al.*, 2000). In an earlier study (Dowd *et al.*, 1997), we found that correct recognition in speech decreases exponentially with displacement on the $(R1,R2)$ plane. (For French, the characteristic length is 0.36 times the average distance between vowels.) Thus, with resonance tuning, a vowel would be likely to be confused with another of higher $R1$ once f_0 exceeded its normal $R1$ by about 100–200 Hz. This confusion could be minimized if a composer took advantage of vowel-pitch matching, i.e., if each vowel were sung with a fundamental frequency f_0 that was consistent with its usual range of $R1$. Further, the singer would no longer need to "tune" the resonance significantly (Carlsson-Berndtsson and Sundberg, 1991): the libretto would partially do it for her, making it easier to sing with a high ratio of output power to input effort.

Composers have long been aware of the problem of reduced intelligibility (Berlioz, 1844). But have they used vowel-pitch matching (whether consciously or unconsciously) to reduce it and/or to make the music easier to sing? If so, the vowel distribution would vary systematically with pitch: vowels associated with a low $R1$ would be sung less often at high pitch and vice versa. If not, we should expect the distribution of vowels to be independent of pitch, simply reflecting that of the libretto. A test of this hypothesis would require libretti with a large number of vowels in different words to ensure that any observed distribution was not a statistical accident. Compositions wherein the intelligibility of the text is important should also be more likely to show a non-uniform distribution of vowels with pitch. Not all operas are thus suitable: in some, the high pitch writing for soprano is aimed at technical display, rather than conveying text. In others, a relatively simple plot is adequately conveyed by repetition or by actions on stage. A composer who was also the librettist might be most likely to match vowels with pitch. For these reasons, this study focused initially on the two great Wagnerian soprano roles; Brünnhilde and Isolde. Each singer has to communicate lengthy, subtle aspects of plot via text alone. Further, the libretti were written by Wagner, published prior to composing the scores, and considered by him and others to be significant literary works in their own right. Four operas by other composers were then studied for comparison.

2. Methods

Using published scores, the phoneme and fundamental (pitch) frequency f_0 were recorded for each note sung by the chosen soprano(s) in each operas studied—see Table 1. Any obvious ornamentation, grace notes, trills, and mordants were not included as they carry no textual information. (Because resonances and formants are broad, high precision in fundamental frequency is not required, so $A4=440$ Hz was assumed throughout).

Wagner provided little ornamentation and generally writes only one note per syllable, (sometimes two for Isolde). To simplify presentation, the 12 vowels of German were grouped into the four standard categories according to their jaw height in the vocoid or Cardinal Vowel space. The range of $R1$ we associated with each category were as follows: *closed* 250–400 Hz, *close-mid* (or *half-close*) 400–550 Hz, *open-mid* (or *half-open*) 550–750 Hz, and *open* 750–1000 Hz. Of course the values of $R1$ will vary with language, dialect, accent, etc. It is also possible that listeners may learn to use a different formant "map" for sopranos (i.e., a different categorization of the vowel plane), in much the same way that we use different maps for men, women, and children. Ultimately, we can only make an informed guess at the vowel sound imagined by the composer-librettist. Although there might be uncertainty about the range of resonance frequencies in each category, the important feature for this study is that their order with increasing frequency is known. The soprano roles of Fiordiligi in *Così fan tutte* and Sophie in *Der Rosenkavalier* were analyzed in a similar fashion.

A slightly abbreviated analysis was made of the soprano roles in *Don Giovanni* by Mozart and *The Barber of Seville* by Rossini, which included only notes longer than a crochet (quarter note) or a quaver (eighth note) in slow sections, in the range from G4 to C6, a range below which resonance tuning by sopranos is not observed (Joliveau *et al.* 2004a, 2004b). There are thus no data on vowel-pitch matching for the closed vowels in these two operas.

Table 1. Details of the operas.

Opera	Composer	Librettist	Year	Language	Soprano role
Der Rosenkavalier[a]	Strauss	von Hofmannsthal	1909–1910	German	Sophie
Der Ring des Nibelungen[b,c]	Wagner	Wagner	1854–1874	German	Brünnhilde
Götterdämmerung[b]	Wagner	Wagner	1869–1874	German	Brünnhilde
Siegfried[c]	Wagner	Wagner	1856–1869	German	Brünnhilde
Tristan und Isolde[d]	Wagner	Wagner	1857–1859	German	Isolde
Die Walküre[c]	Wagner	Wagner	1854–1856	German	Brünnhilde
The Barber of Seville[e]	Rossini	Sterbini	1816	Italian	All
Così fan tutte[f]	Mozart	da Ponte	1790	Italian	Fiordiligi
Don Giovanni[g]	Mozart	da Ponte	1787	Italian	All

[a] The edition of the score used was from Boosey & Hawkes, London, 1943.
[b] The edition of the score used was from Klavierauszug-VEB Breitkopf und Härtel Musikverlag, Leipzig, 1980.
[c] The edition of the score used was from Klavierauszug-VEB Breitkopf und Härtel Musikverlag, Leipzig, 1984.
[d] The edition of the score used was from Ernst Eulenberg, London, 1900.
[e] The edition of the score used was from Kalmus 368, New York.
[f] The edition of the score used was from Neue Ausgabe sämtlicher Werke, Bärenreiter Kassel, Basel, 1991.
[g] The edition of the score used was from Kassel Bärenreiter, 1975.

3. Results and discussion

The upper part of Fig. 1(A) shows, for each pitch, the total number of notes sung for the combined soprano roles of Brünnhilde and Isolde. Each note has a particular pitch and an associated vowel. There are sufficient of these vowel-pitch combinations (more than 10 000) to provide useful statistics and a reasonably smooth distribution with pitch, although some preference for keys harmonically close to C major is apparent. Fewer notes are written at the extremes of the soprano range, presumably because very high notes are physically demanding and dramatically effective if used sparingly, and because sopranos' low notes often have reduced dynamic range.

The lower part of Fig. 1(A) shows the distribution of vowels grouped into the four categories for jaw height and thus $R1$. The vowel distribution does vary with frequency, and in a fashion that helps match $R1$ to f_0. Thus the closed and close-mid vowels with low $R1$ become less common as f_0 rises above 500–600 Hz. Conversely, the fraction of open-mid and open vowels increases significantly above 600 Hz. The open vowels are, of course, preferred at high pitch, whereas other vowels would be seriously distorted by resonance tuning. However, open vowels are also used across the whole pitch range. The unmodified $R1$ for these vowels is sufficiently high that, for almost the whole of the soprano range, the vocal tract load at the glottis is inertive. Further, for notes in the low soprano range, $R1$ may be excited by a second or even third harmonic. So the open vowels are least likely to be distorted by resonance tuning.

Figure 1(B) shows the data for the vowel-pitch combinations for the role of Fiordilgi in *Così fan tutte*, an opera buffa by Mozart. There is no significant variation of vowels with pitch, except surprisingly at low notes where the closed vowels (u and i) are less likely than at the highest notes; this could reduce intelligibility.

Figure 2 shows the extent of pitch-resonance matching for the eight operas studied in terms of γ, a parameter we define as follows. For all notes lying in a frequency band corresponding to a particular jaw height, let g be the average fraction of notes whose vowel corresponds to that jaw height. Let h be the average fraction of notes at all other frequencies having those vowels. The parameter $\gamma = g/h - 1$ then indicates the preference for the appropriate vowel-pitch combinations. Positive and negative values indicate favorable and unfavorable pitch-resonance matching, respectively. The vowel distribution is seen to change systematically with frequency

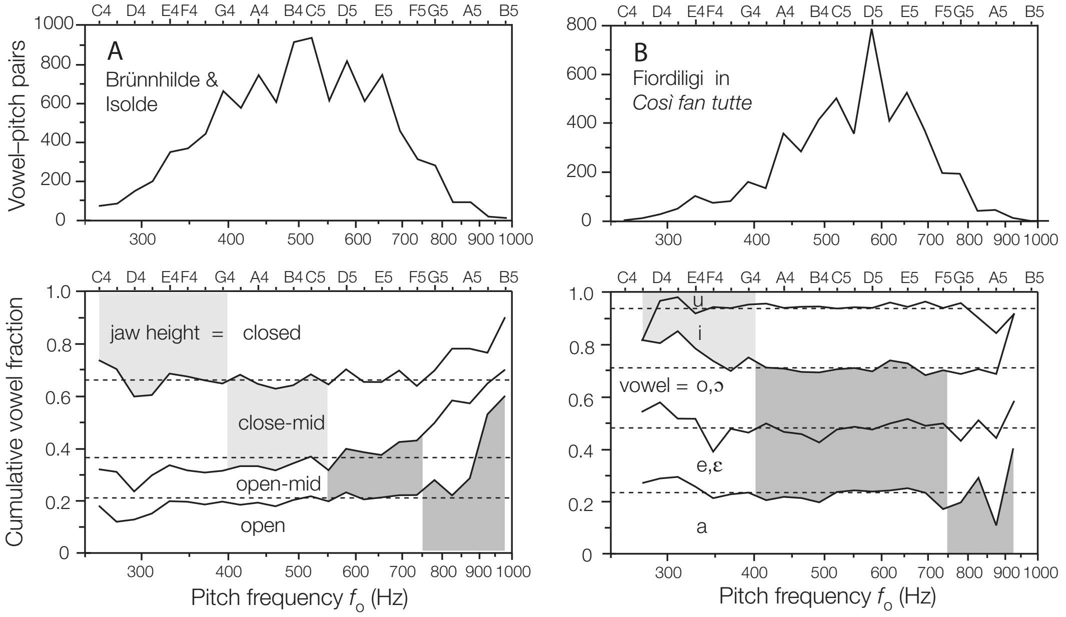


Fig. 1. Semi-logarithmic plots of the number of vowel-pitch combinations (upper) and the cumulative vowel fraction (lower) as function of written pitch frequency f_0 for soprano roles in operas by Wagner and Mozart. On the lower figures, the open vowels lie between the axis and the lowest continuous line, open-mid between the two lowest continuous lines, etc. The dashed lines show the null hypothesis: equal distribution of vowels across pitch. Shaded areas indicate regions where $R1$ and pitch might be favorably matched. Heavily shaded areas indicate approximate regions where resonance tuning would be particularly favorable.

in all four Wagner operas studied. The direction of the changes strongly supports the hypothesis of non-accidental matching of vowels with pitch, but it is worth considering non-acoustic causes. For example, high notes are often used for dramatic emphasis, so important declarative text might alter the distribution via preferential repetition at high pitch. Isolde often sings "Tristan" and "Liebe," usually with descending pitch. Brünnhilde is also heavily involved with "Sieglinde," "Siegmund," "Siegfried," and "Liebe." However, removal of these words from the analysis did not significantly alter the overall distribution of vowels with frequency. The data presented in the figures do not include the war-whoops of Brünnhilde in *Die Walküre* because, although her repeated cries of "Hojotoho Heiaha" occur at high pitch, they convey negligible textual information. Their inclusion does not significantly alter the vowel distribution.

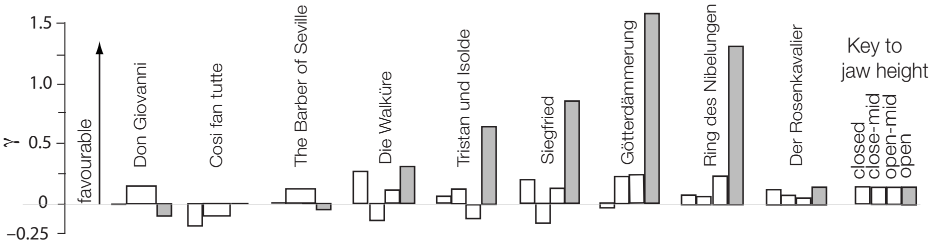


Fig. 2. The extent of pitch-resonance matching for soprano roles in the eight operas studied. The degree of matching is indicated by γ, a parameter that indicates the preference for the appropriate vowel-pitch combinations (see text). Positive and negative values of γ indicate favorable and unfavorable pitch-resonance matching, respectively. Values associated with open vowels (high $R1$) are indicated by shading. The operas are shown in historical order from left to right.

Do other composers or librettists match vowels with pitch? Figure 2 indicates that, for the limited sample of works studied, a significant increase in γ for open vowels occurs only in the operas by Wagner. Interestingly, γ for open vowels increased systematically as Wagner's experience as a composer increased. For comparison, the role of Sophie in the later opera *Der Rosenkavalier* by Richard Strauss shows no significant effect.

Wagner's idea of opera was a continuous music drama. Earlier operas often linked separate arias and choruses with explanatory recitative and thus had less need for intelligibility at high pitch. Furthermore Wagner wrote for much larger orchestras than those available to Mozart or Rossini, and wrote vocal parts that severely test the stamina and capabilities of singers. Thus the employment of vowel-pitch matching could have helped satisfy the concomitant requirements of intelligibility, vocal power, and easier singing of difficult parts.

A number of complications and differences are also relevant: in some operas, important phrases can be repeated several times at different pitches—rarely in Wagner and Strauss, but often in Mozart. One might also consider the time available to composers to "polish" the operas; Rossini wrote some 26 operas in 7 years whereas Wagner wrote 14 in over 50 years.

It is also possible that a composer-librettist might alter the overall vowel distribution from that of normal speech to favor more open vowels, particularly when writing text for orchestral accompaniment. First, singers can produce higher sound pressure levels using open vowels at any pitch (Gramming and Sundberg, 1988). Furthermore, the time-averaged power spectrum produced by an orchestra peaks around 500–600 Hz and then decreases with increasing frequency (Sundberg, 1977). Consequently, the voice of a singer using more open vowels would be expected to have more power above 500 Hz through reinforced harmonics, even when singing at lower pitch. Thus a libretto to be sung with orchestra might advantageously include a higher proportion of open vowels than written text. This was not investigated here, in part because of the difficulty in choosing appropriate texts to use as the non-operatic controls.

Many factors, apart from vowels, are involved in intelligibility. For instance, if only one vowel in a string of phonemes produces a real word, then the vowel need not be recognized. Similarly, context in a sentence can mean that a weird can be understood, independently of the vowel. The results of this study also suggest that those translating a libretto for performance in another language might occasionally consider vowel-pitch matching—among all of the other constraints.

4. Conclusions

The authors are unaware of any written evidence about the composers' intentions nor of whether they were advised on this issue by sopranos, with whom they sometimes had quite close relations. However, it appears that Wagner, either consciously or unconsciously, did take the acoustics of the soprano voice at high pitch into account when setting text to music. This is consistent with the increased importance of textual information in his operas, the increasing size of his orchestras, and the more complex vocal parts.

Acknowledgment

The support of the Australian Research Council is gratefully acknowledged.

References and links

[1] We follow the original definitions of Fant (1970) and reserve the term "formant" for broad peaks in the spectral envelope and "resonance" for the acoustic resonances of the vocal tract that produce them—see Wolfe *et al.* (2009) for further discussion.

Berlioz, H. (**1844**). *Grand traité d'instrumentation et d'orchestration modernes (A treatise on modern instrumentation and orchestration)*, translated by M. C. Clarke (Novello, London).

Carlsson-Berndtsson, G., and Sundberg, J. (**1991**). "Formant frequency tuning in singing," Speech Transm. Lab. Q. Prog. Status Rep. **32**, 29–35.

Coffin, B. (**1974**). "On hearing, feeling and using the instrumental resonance of the singing voice," NATS Bulletin **31**, 26–30.

Coffin, B. (**1976**). *Coffin's Sounds of Singing* (The Scarecrow, Lanham).

Dowd, A., Smith, J., and Wolfe, J. (**1997**). “Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time,” Lang Speech **41**, 1–20.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Gramming, P., and Sundberg, J. (**1988**). “Spectrum factors relevant to phonetogram measurement,” J. Acoust. Soc. Am. **83**, 2352–2360.

Hollien, H., Mendes-Schwartz, A. P., and Nielsen, K. (**2000**). “Perceptual confusions of high-pitched sung vowels,” J. Voice **14**, 287–298.

Joliveau, E., Smith, J., and Wolfe, J. (**2004a**). “Tuning of vocal tract resonances by sopranos,” Nature (London) **427**, 116.

Joliveau, E., Smith, J., and Wolfe, J. (**2004b**). “Vocal tract resonances in singing: The soprano voice,” J. Acoust. Soc. Am. **116**, 2434–2439.

Lindblom, B. E. F., and Sundberg, J. E. F. (**1971**). “Acoustical consequences of lip, tongue, jaw, and larynx movement,” J. Acoust. Soc. Am. **50**, 1166–1179.

Morozov, V. P. (**1965**). “Intelligibility in singing as a function of fundamental voice pitch,” Sov. Phys. Acoust. **10**, 279–283.

Scotto di Carlo, N., and Germain, A. (**1985**). “A perceptual study of the influence of pitch on the intelligibility of sung vowels,” Phonetica **42**, 188–197.

Sundberg, J. (**1975**). “Formant technique in a professional female singer,” Acustica **32**, 89–96.

Sundberg, J. (**1977**). “The acoustics of the singing voice,” Sci. Am., 82–91.

Titze, I. R. (**1988**). “The physics of small-amplitude oscillations of the vocal folds,” J. Acoust. Soc. Am. **83**, 1536–1552.

Titze, I. R. (**2008**). “Nonlinear source-filter coupling in phonation: Theory,” J. Acoust. Soc. Am. **123**, 2733–2749.

Titze, I. R., Tobias, R., and Popolo, P. (**2008**). “Nonlinear source-filter coupling in phonation: Vocal exercises,” J. Acoust. Soc. Am. **123**, 1902–1915.

Wolfe, J., Garnier, M., and Smith, J. (**2009**). “Vocal tract resonances in speech, singing and playing musical instruments,” HFSP J., **3**(1), 6–23.

Time-domain simulations of sound propagation in a stratified atmosphere over an impedance ground

Benjamin Cotté and Philippe Blanc-Benon
LMFA, UMR CNRS 5509, École Centrale de Lyon, 69134 Ecully Cedex, France
benjamin.cotte@ec-lyon.fr, philippe.blanc-benon@ec-lyon.fr

Abstract: Finite-difference time-domain simulations of broadband sound propagation in a stratified atmosphere are presented. A method recently proposed to obtain an impedance time-domain boundary condition is implemented in a linearized Euler equations solver, which enables to study long range sound propagation over an impedance ground. Some features of the pressure pulse evolution with time are analyzed in both upward-and downward-refracting conditions, and the time-domain simulations are compared to parabolic equation calculations in the frequency domain to show the effectiveness of the proposed impedance boundary condition.

© 2009 Acoustical Society of America

PACS numbers: 43.28.Js, 43.28.En [VO]
Date Received: January 19, 2009 **Date Accepted:** February 18, 2009

1. Introduction

Time-domain numerical solutions of the linearized Euler equations are becoming increasingly popular to study broadband noise propagation outdoors,[1–3] since they can accurately take into account the interactions of the acoustic waves with local wind and temperature fluctuations in the atmospheric boundary layer. They are also well suited to study the sound field close to the acoustic sources, which can be complex in the context of transportation noise.

When performing finite-difference time-domain (FDTD) simulations, one of the main difficulties is to account for the reflection of acoustic waves over an impedance ground. Indeed, impedance models classically used for outdoor grounds have been obtained in the frequency domain, and most of them do not meet the necessary conditions for an impedance model to be physically possible, which means that they cannot be directly translated into the time domain.[4,5] Furthermore, when the translation into the time domain is possible, a convolution needs to be solved in the time-domain boundary condition (TDBC), which is computationally expensive. Different methods have been proposed to derive impedance TDBC.[6–8] Recently, Cotté *et al.*[9] presented a TDBC for the Miki impedance model,[4] which is a physically possible extension of the Delany–Bazley impedance model.[10] This TDBC is based on an approximation of the impedance in the frequency domain, which allows the use of the recursive convolution method,[11] a very efficient technique to calculate a discrete convolution. It can be noticed that another way to account for the interaction of sound waves with an impedance ground is to add an explicit porous layer to the computational domain, as done by Salomons *et al.*[1] and by Van Renterghem and Botteldooren.[2] This method is useful to model certain types of grounds, such as extended-reaction grounds but requires additional calculations to be performed in the porous medium.

In this paper, the method proposed to obtain the impedance TDBC is first summarized. Then, after a brief description of the linearized Euler equations solver, two-dimensional simulations of broadband noise propagation in a stratified atmosphere are presented. The pressure pulses obtained in downward and upward-refracting conditions are compared to the homogeneous case. Finally, the spectra of the time-domain simulations are calculated and compared to parabolic equation simulations in the frequency domain.

2. Derivation of time-domain impedance boundary conditions

Let $p(t)$ be the acoustic pressure and $v(t)$ the component of particle velocity normal to the interface between the ground and the air, with $P(\omega)$ and $V(\omega)$ their respective Fourier transform, t the time, and ω the angular frequency; the $\exp(-j\omega t)$ convention is assumed throughout this paper. Classically, the characteristic impedance $Z(\omega)$ is defined in the frequency domain using $P(\omega)=Z(\omega)V(\omega)$. The direct translation of this boundary condition in the time domain involves a convolution operator that is computationally expensive to solve. Thus, following a method proposed by Fung and Ju[6] and Reymen *et al.*,[7] the characteristic impedance $Z(\omega)$ is approximated as a sum of first-order systems:

$$Z(\omega) \approx \sum_{k=1}^{S} \frac{A_k}{\lambda_k - j\omega}, \tag{1}$$

with λ_k the real poles in the approximation, A_k the corresponding coefficients, and S the number of poles. Using these first-order systems, the impedance is guaranteed to be physically possible if $\lambda_k \geq 0$ and if the passivity condition is met (positive real part of the impedance). Furthermore, the form of these functions allows the use of the recursive convolution method, introduced by Luebbers and Hunsberger[11] in the context of electromagnetic propagation through dispersive media. Considering the discretized variables $p^{(n)}=p(n\Delta t)$ and $v^{(n)}=v(n\Delta t)$, with Δt the time step, the following TDBC is obtained:[9]

$$p^{(n)} = \sum_{k=1}^{S} A_k \phi_k^{(n)}, \tag{2}$$

where the accumulators ϕ_k are given by the recursive formula:

$$\phi_k^{(n)} = v^{(n)} \frac{1 - e^{-\lambda_k \Delta t}}{\lambda_k} + \phi_k^{(n-1)} e^{-\lambda_k \Delta t}. \tag{3}$$

It can be seen that S accumulators ϕ_k are needed in the TDBC, with only two storage locations per accumulator.

Cotté *et al.*[9] compared different methods to identify the coefficients A_k and λ_k of Eq. (1), and showed that it is desirable to constrain the values of the poles λ_k to obtain accurate numerical results. They propose an optimization method in the frequency domain, which guarantees that the impedance model is physically possible and that the values of the poles are sufficiently small. This method is applied to the Miki impedance model of a semi-infinite ground layer:[4]

$$Z/\rho_0 c_0 = 1 + 0.0699(f/\sigma_e)^{-0.632} + j0.107(f/\sigma_e)^{-0.632}, \tag{4}$$

with ρ_0 the air density, c_0 the sound speed in the air, f the frequency, and σ_e an effective flow resistivity [the coefficients in Eq. (4) are in SI units]. The coefficient identification method is performed on the frequency band [50 Hz, 1200 Hz], which includes the spectrum of the pulse used in the simulations presented in Sec. 3. The coefficients A_k and λ_k corresponding to the Miki impedance model with an effective flow resistivity of 100 kPa s m^{-2} are given in Table 1. It can be seen in Fig. 1 that the real and imaginary parts of the impedance are very well approximated using only five real poles over the frequency band of interest. Note that this method has also been applied to the Miki model of a rigidly backed layer in Ref. 9.

3. Time-domain simulations in a stratified atmosphere

3.1 Linearized Euler equation solver

The linearized Euler equations are solved using FDTD methods developed in the computational aeroacoustics community.[12–14] Optimized finite-difference schemes and selective filters over 11 points are used for spatial derivation and grid-to-grid oscillations removal, respectively. These

Table 1. Coefficients A_k and λ_k for the Miki impedance model of a semi-infinite ground layer of effective flow resistivity 100 kPa s m^{-2}. This set of coefficients was referred to as OF v1 in Ref. 9.

k	A_k	λ_k
1	$1.414\,390\,450\,609 \times 10^6$	$5.233\,002\,301\,836 \times 10^1$
2	$1.001\,354\,674\,975 \times 10^6$	$4.946\,064\,975\,401 \times 10^2$
3	$-3.336\,020\,206\,713 \times 10^6$	$1.702\,517\,657\,290 \times 10^3$
4	$5.254\,549\,668\,250 \times 10^6$	$1.832\,727\,486\,745 \times 10^3$
5	$3.031\,704\,943\,714 \times 10^7$	$3.400\,000\,000\,000 \times 10^4$

numerical schemes are optimized in the wave number space by minimizing the dispersion and dissipation errors, so that acoustic wavelengths down to five or six times the spatial mesh size are accurately calculated.[9,15] For the interior points, which are the points separated from the boundary by at least five points, the centered fourth-order finite-difference scheme of Bogey and Bailly[12] and the centered sixth-order selective filter recently proposed by Bogey *et al.*[14] are chosen. For the boundary points, that are the five extreme points in each direction, the 11-point non-centered finite-difference schemes and selective filters of Berland *et al.*[13] are used. Applying a selective filter to a variable U on a uniform mesh of size Δx provides

$$U^f(x_0) = U(x_0) - s_f \sum_{m=-P}^{Q} d_m U(x_0 + m\Delta x). \tag{5}$$

In the following, a filtering coefficient s_f of 0.2 is taken for all selective filters except at the extreme points where a filtering coefficient of $s_f/20$ is chosen. A smaller coefficient is indeed taken for the completely off-centered filter because this filter is much more dissipative than the other selective filters. The completely off-centered filter is needed when a large number of time iterations is performed[15] as it is the case in this paper.

The optimized six-stage Runge–Kutta algorithm of Bogey and Bailly[12] is used for time integration. The TDBC presented in Sec. 2 can easily be adapted to this algorithm, as shown in Ref. 9. The simulations presented in this paper are performed with a CFL$=c_0\Delta t/\Delta x$ of 1, with

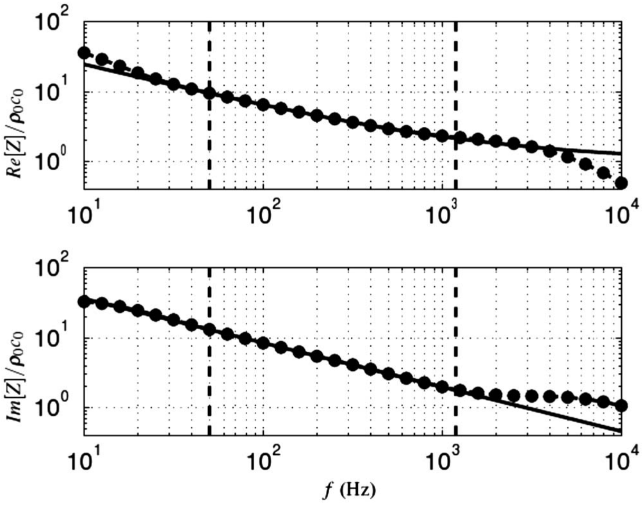


Fig. 1. Real and imaginary parts of the normalized impedance $Z/\rho_0 c_0$. The solid line corresponds to the Miki model of a semi-infinite ground layer with an effective flow resistivity of 100 kPa s m^{-2}, and the dots correspond to the fit obtained using the frequency-domain approximation. The frequency band [50 Hz, 1200 Hz] is represented by vertical lines.

Δt the time step. An orthogonal non-staggered grid is used, and the TDBC given by Eqs. (2) and (3) is only applied at the ground boundary ($z=0$); more details on the numerical implementation of the TDBC can be found in Ref. 9. At all other boundaries, the radiation boundary conditions of Tam and Dong[16] are considered. The initial pressure distribution used in the simulations has the following Gaussian form: $p(r,t=0)=A\exp(-\ln 2r^2/B^2)$, with $B=5\Delta x=0.25$ m the Gaussian half-width, $r=\sqrt{x^2+(z-z_S)^2}$ the source-receiver separation, and $z_S=2$ m the source height. This starter has a broadband spectrum with significant frequency content up to 800 Hz approximately.[9,15]

3.2 Numerical results

In the rest of the paper, a two-dimensional propagation configuration is studied, with a size of approximately 500 m in the x-direction and 100 m in the z-direction. The mesh size is 0.05 m, and there are about 22×10^6 points in the computational domain. The calculation is run over 11 000 time iterations to enable the pulse to leave the computational domain. This calculation is performed on a NEC SX-8 vector machine and takes about 8 h to run (more details can be found in Ref. 15). A logarithmic sound speed profile is considered, which has the following form: $c(z)=c_0+a_c\ln(1+z/z_0)$, with $c_0=340$ m/s and $z_0=0.1$ m. Three values of the coefficient a_c are considered, corresponding to downward-refraction ($a_c=+1$ m/s), upward-refraction ($a_c=-1$ m/s), and homogeneous conditions ($a_c=0$).

Two movies of the pulse evolution with time are given in Mm. 1 for a receiver height of 2 m and Mm. 2 for a receiver height of 10 m. The pressure amplitude is corrected by $\sqrt{r}$ to take into account geometrical spreading in two dimensions. At each distance x, the pressure waveforms are centered at the time x/c_0. The movies show that the pressure pulse arrives earlier in the downward-refracting case ($a_c=+1$ m/s) and later in the upward-refracting case ($a_c=-1$ m/s) compared to the homogeneous case ($a_c=0$). There is evidence of multiple arrivals in the downward-refracting case, with relatively strong acoustic pressure amplitude. In the upward-refracting case, the pulse amplitude is attenuated compared to the homogeneous case, except at short range where an amplification of the acoustic pressure can be observed. This amplification can be attributed to a modification of the ground effect due to the vertical sound speed gradients close to the ground. This effect is particularly clear for a receiver height of 10 m and distances between 50 and 100 m approximately in Mm. 2, and the modification of the ground effect will appear on the spectra plotted in Fig. 3. Another feature of the pressure waveforms shown in the movies is the presence of a long trail at the end of the pulses. This long trail component has been shown to be a surface wave in Ref. 17. It must be noted that in the atmosphere, the insonification of the refractive shadow zone is mainly caused by sound scattering by turbulence; however, this effect is not considered in this paper.

Mm. 1. Normalized pressure $p(t)\sqrt{r}/A$ with respect to time at a height of 2 m and distances between 10 and 500 m. The blue curve corresponds to the downward-refracting case ($a_c=+1$ m/s), the black curve to the homogeneous case ($a_c=0$), and the red curve to the upward-refracting case ($a_c=-1$ m/s). This a file of type “avi” (3.7 Mbytes).

Mm. 2. The same as in Mm. 1, but for a height of 10 m. This a file of type “avi” (3.9 Mbytes).

4. Comparison with parabolic equation simulations in the frequency domain

The time-domain simulations are now compared to parabolic equation (PE) calculations in the frequency domain to show the effectiveness of the impedance TDBC described in Sec. 2. The spectra of the sound pressure level relative to the free field, noted ΔL, are calculated using a fast Fourier transform. The PE calculations are obtained using a wide-angle parabolic equation (WAPE) code described in Refs. 18 and 15, with a frequency-domain impedance boundary condition based on Eq. (4). The spectra of ΔL for the FDTD and PE calculations are compared in Fig. 2 for a receiver height of 2 m, and in Fig. 3 for a receiver height of 10 m. Both solutions agree very well between 50 and 800 Hz. Above 800 Hz, there is no significant energy in the

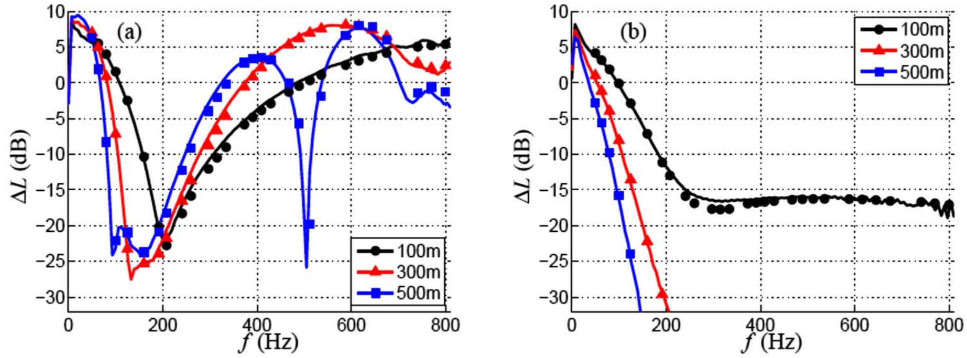


Fig. 2. (Color online) Spectra of the sound pressure level relative to the free field ΔL at a height of 2 m and propagation distances of 100, 300, and 500 m (a) in downward-refracting conditions ($a_c = +1$ m/s) and (b) in upward-refracting conditions ($a_c = -1$ m/s). The solid lines correspond to the FDTD calculations, and the symbols correspond to the PE calculations.

initial pulse of the FDTD calculation. Propagation distances are thus greater than 1000 acoustic wavelengths for the largest range and frequency considered in these simulations, which shows that the proposed FDTD model is well suited to study long range sound propagation.

These spectra also explain the pressure pulse evolution with time shown in Mm. 1 and Mm. 2. In downward-refracting conditions [see Figs. 2(a) and 3(a)], the sound pressure levels are quite large and the interference pattern is complex, especially at a range of 500 m, which can be linked to the multiple arrivals that are observed in the movies. In upward-refracting conditions [see Figs. 2(b) and 3(b)], the frequency components above 200 Hz are strongly attenuated at ranges greater or equal to 300 m. Thus, the pressure pulse contains only very low-frequency components at these ranges. At a receiver height of 10 m and a range of 100 m, however, the sound pressure level is quite large over the whole frequency band of interest, and even larger than the sound pressure level in downward-refracting conditions. The interference dip located at about 200 Hz in the downward-refracting case is shifted to higher frequencies in the upward-refracting case, which explains the pulse amplification observed in the movie Mm. 2.

5. Conclusion

In this paper, long range sound propagation over an impedance ground is studied using FDTD methods. In particular, a method recently proposed to obtain an impedance TDBC has been applied to the Miki impedance model. This TDBC has been implemented in a linearized Euler

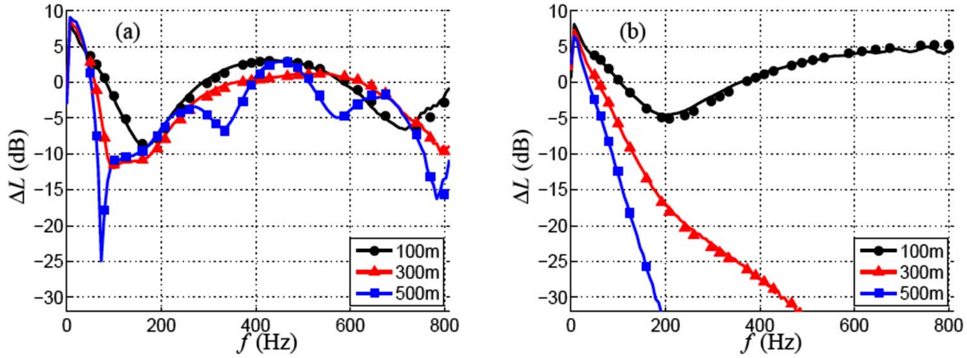


Fig. 3. (Color online) The same as in Fig. 2, but for a height of 10 m.

equations solver and two-dimensional simulations in a stratified atmosphere have been presented. One type of ground has been tested in the simulations on the frequency range [50 Hz, 800 Hz]. The solver used in this work enables to study sound propagation over distances greater than 1000 acoustic wavelengths, as the comparison with PE calculations in the frequency domain has shown. Interesting features of the pressure pulse evolution with time in both upward- and downward-refracting conditions have been analyzed. In the future, the surface wave component that appears in the time-domain solution will be studied in more detail.

Acknowledgments

Support by CNRS and SNCF is gratefully acknowledged. Computing time was supplied by the Institut du Développement et des Ressources en Informatique Scientifique, Centre National de la Recherche Scientifique. The authors would like to thank the reviewers for their constructive comments and suggestions.

References and links

[1] E. M. Salomons, R. Blumrich, and D. Heimann, “Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain models,” Acta. Acust. Acust. **88**, 483–492 (2002).

[2] T. Van Renterghem and D. Botteldooren, “Numerical simulation of the effect of trees on downwind noise barrier performance,” Acta. Acust. Acust. **89**, 764–778 (2003).

[3] D. K. Wilson, S. L. Collier, V. E. Ostashev, D. F. Aldridge, N. P. Symons, and D. H. Marlin, “Time-domain modeling of the acoustic impedance of porous surface,” Acta. Acust. Acust. **92**, 965–975 (2006).

[4] Y. Miki, “Acoustical properties of porous materials—Modifications of Delany-Bazley models,” J. Acoust. Soc. Jpn. **11**, 19–24 (1990).

[5] Y. H. Berthelot, “Surface acoustic impedance and causality,” J. Acoust. Soc. Am. **109**, 1736–1739 (2001).

[6] K.-Y. Fung and H. Ju, “Broadband time-domain impedance models,” AIAA J. **39**, 1449–1454 (2001).

[7] Y. Reymen, M. Baelmans, and W. Desmet, “Time-domain impedance formulation suited for broadband simulations,” in 13th AIAA/CEAS Aeroacoustics Conference, Rome, Italy, (2007), AIAA Paper No. 2007-3519.

[8] V. E. Ostashev, S. L. Collier, D. K. Wilson, D. F. Aldridge, N. P. Symons, and D. Marlin, “Padé approximation in time-domain boundary conditions of porous surfaces,” J. Acoust. Soc. Am. **122**, 107–112 (2007).

[9] B. Cotté, P. Blanc-Benon, C. Bogey, and F. Poisson, “Time-domain impedance boundary conditions for simulations of outdoor sound propagation,” AIAA J. (in press). See also AIAA Paper 2008-3021.

[10] M. E. Delany and E. N. Bazley, “Acoustical properties of fibrous absorbent materials,” Appl. Acoust. **3**, 105–116 (1970).

[11] R. J. Luebbers and F. Hunsberger, “FDTD for Nth-order dispersive media,” IEEE Trans. Antennas Propag. **40**, 1297–1301 (1992).

[12] C. Bogey and C. Bailly, “A family of low dispersive and low dissipative explicit schemes for flow and noise computations,” J. Comput. Phys. **194**, 194–214 (2004).

[13] J. Berland, C. Bogey, O. Marsden, and C. Bailly, “High order, low dispersive and low dissipative explicit schemes for multiple-scale and boundary problems,” J. Comput. Phys. **224**, 637–662 (2007).

[14] C. Bogey, N. Cacqueray, and C. Bailly, “A shock-capturing methodology based on adaptative spatial filtering for high-order non-linear computations,” J. Comput. Phys. **228**, 1447–1465 (2009).

[15] B. Cotté, “Propagation acoustique en milieu extérieur complexe: Problèmes spécifiques au ferroviaire dans le contexte des trains à grande vitesse (Outdoor sound propagation in complex environments: Specific problems in the context of high speed trains),” Ph.D. thesis 2008-19, École Centrale de Lyon, Ecully Cedex (2008).

[16] C. K. W. Tam and Z. Dong, “Radiation and outflow boundary conditions for direct computation of acoustic and flow disturbances in a nonuniform mean flow,” J. Comput. Acoust. **4**, 175–201 (1996).

[17] B. Cotté and P. Blanc-Benon, “Outdoor sound propagation simulations in the time domain using linearized Euler equations with suitable impedance boundary conditions,” in Proceedings of the 13th International Symposium on Long Range Sound Propagation, Lyon, France, (2008).

[18] B. Cotté and P. Blanc-Benon, “Estimates of the relevant turbulent scales for acoustic propagation in an upward refracting atmosphere,” Acta. Acust. Acust. **93**, 944–958 (2007).

The pitch levels of female speech in two Chinese villages

Diana Deutsch[a)]
Department of Psychology, University of California, San Diego, La Jolla, California 92093
ddeutsch@ucsd.edu

Jinghong Le
School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China
jhle@psy.ecnu.edu.cn

Jing Shen and Trevor Henthorn
Department of Psychology, University of California, San Diego, La Jolla, California 92093
jshen@psy.ucsd.edu, trevor@music.ucsd.edu

Abstract: The pitch levels of female speech in two villages situated in a relatively remote area of China were compared. The dialects spoken in the two villages are similar to Standard Mandarin, and all subjects had learned to read and speak Standard Mandarin at school. Subjects read out a passage of roughly 3.25 min in Standard Mandarin, and pitch values were obtained at 5-ms intervals. The overall pitch levels in the two villages differed significantly, supporting the conjecture that pitch levels of speech are influenced by a mental representation acquired through long-term exposure to the speech of others.

© 2009 Acoustical Society of America

PACS numbers: 43.71.Bp, 43.71.Es, 43.71.An [JH]
Date Received: January 24, 2009 **Date Accepted:** March 12, 2009

1. Introduction

While a substantial literature exists concerning the features of particular languages and dialects, overall pitch level as a feature has so far received little attention. This is due in part to the assumption frequently made that the pitch level of speech is physiologically determined and that it serves as a reflection of body size (Kunzel, 1989; Van Dommelen and Moxness, 1995). However, taking male and female speech separately, a convincing absence of correlate has been obtained between overall pitch level and the speaker's body dimensions such as height, weight, size of larynx, and so on (Hollien and Jackson, 1973; Kunzel, 1989; Van Dommelen and Moxness, 1995; Collins, 2000; Gonzales, 2004; Lass and Brown, 1978; Majewski *et al.*, 1972). In contrast, various studies have found that the pitch level of speech varies with the speaker's language (Hollien and Jackson, 1973; Majewski *et al.*, 1972; Hanley *et al.*, 1966; Yamazawa and Hollien, 1992), indicating that it is subject to a cultural influence (Dolson, 1994; Honorof and Whalen, 2005; Xue *et al.*, 2002), though the precise nature of this influence has not received much consideration.

Deutsch and co-workers (cf. Deutsch, 1992) proposed that the pitch level of an individual's speaking voice is strongly influenced by the pitch levels of speech in his or her linguistic community. More specifically, it was hypothesized that through long-term exposure to the speech of others, the individual acquires a mental representation of the expected pitch range and pitch level of speech (for male and female speech taken separately), and that this representation includes a delimitation of the octave band in which the largest proportion of pitch values occurs. It should be noted that the pitch range of speech has frequently been determined to be roughly

[a)]Author to whom correspondence should be addressed.

an octave, for both male and female speakers, and across a diversity of languages and dialects (Dolson, 1994; Hudson and Holbrook, 1982; Kunzel, 1989; Majewski *et al.*, 1972; Xue *et al.*, 2002; Yamazawa and Hollien, 1992; Hanley *et al.*, 1966; Hollien and Jackson, 1973). A detailed account and appraisal of the proposed model can be found in Dolson (1994). It leads to the further assumption that, taking two communities, each of which is linguistically homogeneous, the overall pitch levels of speech should cluster within each community, but might differ across communities. It is further hypothesized that when acquiring a second language or dialect, the individual imports the mental representation of the pitch levels that he or she had originally acquired (Deutsch *et al.*, 2004).

The present study was designed to test the above hypothesis by comparing the overall pitch levels of female speech in two communities. The subject populations were located in two villages situated in a relatively remote area of China, which is considered to be stable and homogeneous in terms of ethnicity, culture, and lifestyle (Blunden and Elvin, 1998). The villages are less than 40 miles apart, though travel time between them by automobile takes several hours. The dialects spoken in these villages are quite similar, being in the same general family as Standard Mandarin, and communication between residents of the two villages using their native dialects occurs without difficulty. Further, the speech of residents of both villages can be readily understood by speakers of Standard Mandarin. All subjects in the study had learned to speak and read Standard Mandarin in school.

Each subject was given a passage of roughly 3.25 min in duration to read out in Standard Mandarin, and from this reading, pitch values were obtained at 5-ms intervals. Two types of analysis were then performed. First, for each subject an average fundamental frequency (F0) was obtained, and statistical comparison was made between the average F0s obtained from subjects in the two villages. Second, for each subject the F0s were allocated to semitone bins, a histogram was created of the percentage occurrence of F0 values in each bin, and from this histogram the octave band containing the largest number of F0 values was derived. Statistical comparison was then made between the positions of the octave bands derived from subjects in the two villages.

2. Method

2.1 Subjects

Thirty-three female subjects participated in the experiment. They were tested in two locations: 17 subjects in Taoyuan Village, near Guandu, Zhushan County, in Hubei Province, and 16 subjects in Jiuying Village, near Bailu, Wuxi County, in the municipality of Chongqing. Those tested in Taoyuan Village were of average age 33.7 years (18–48 years) and had received an average of 7.6 years (4–12 years) of school education where they had learned to speak and read Standard Mandarin. They had all been born in or near Guandu and had not lived outside Zhushan County for more than 5 years. Those tested in Jiuying Village were of average age 37.6 years (25–52 years) and had received an average of 7.1 years (3–12 years) of school education where they had learned to speak and read Standard Mandarin. They had all been born in or near Bailu and had not lived outside Wuxi County for more than 5 years. All subjects except two were married, and their husbands were all locally born. With two exceptions, the subjects' parents had been born in the same county as the subjects and had lived in the same county for most of their lives. All subjects reported that they had normal hearing and were free of respiratory illness at the time of testing.

2.2 Apparatus and procedure

The subjects in both locations were tested individually in a quiet environment. They were first interviewed to inquire into their state of health and hearing and to determine that they had an adequate level of competence in speaking Standard Mandarin. They were also administered a questionnaire that inquired into their linguistic background and life history. Then they were given a short, emotionally neutral article to read out in Standard Mandarin for practice. Following this, they were given the test article to read out in Standard Mandarin, and their speech was

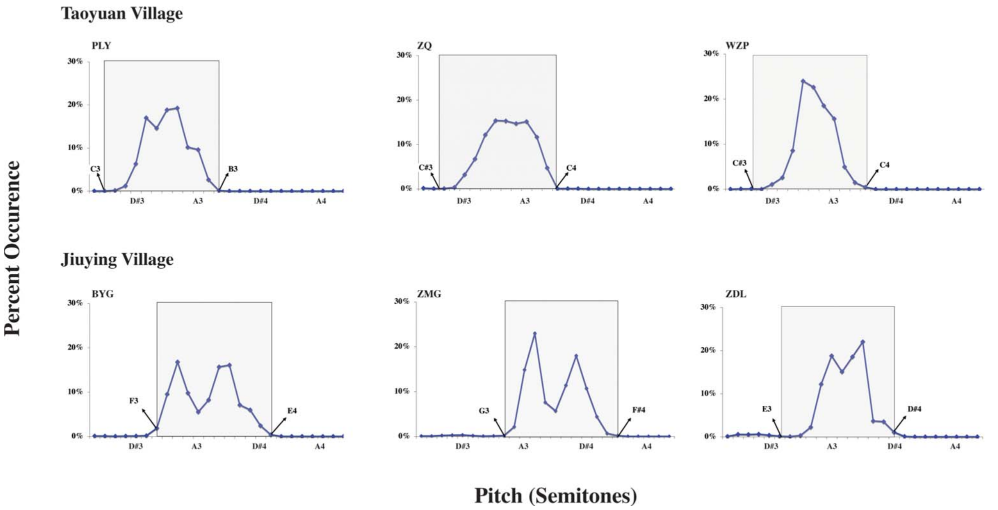


Fig. 1. (Color online) The percentage occurrence of F0 values in a 3-min segment of speech plotted in semitone bins. The data from three subjects in each village are displayed. The center of each bin is displayed on the abscissa: D#3=155.6 Hz; A3=220 Hz; D#4=331.1 Hz; A4=440 Hz. The gray area on each histogram shows the octave band in which the largest number of F0 values occurred.

recorded. The test article was also emotionally neutral, contained 480 Chinese characters, and took an average of roughly 3.25 min to read out. The subjects were paid for their services.

For each subjects' reading, the speech samples were recorded via a SONY ECM-CS10 lavalier microphone onto an Edirol R-1 digital recorder as 16-bit, 44.1-kHz WAV files. The sound files were transferred to an iMac running OSX 10.5, converted to a sampling rate of 11.025 kHz, and the first 20 s of each file was deleted. The soundfiles were then converted to NeXT format and transferred to a NeXT computer (NeXTstation Turbo Color).

The sound files were lowpass filtered with a cutoff frequency of 1300 Hz. F0 estimates were then obtained at 5-ms intervals, using a procedure derived from Rabiner and Schafer (1978). The low and high boundaries for the F0 estimates were set at 107 and 639 Hz, respectively. In addition, for each subject's recording, the time-varying energy level of the signal was obtained, and only those F0 estimates that were associated with levels no lower than 25 dB below the peak level were saved for further analysis. (This procedure was employed so as to eliminate spurious F0 estimates, such as obtained during pauses in the subject's speech.) Then for each subject's reading, the F0 estimates were averaged along the musical scale; that is, along a log frequency continuum, so producing an average F0 for each subject. Furthermore, as a separate procedure, the raw F0 estimates were allocated to semitone bins, with the center frequency of each bin determined by the equal-tempered scale (A=440 Hz). Histograms were then generated for each subject showing the percentage occurrence of F0 values in each semitone bin.

3. Results

Figure 1 presents, as examples, the histograms showing the percentage occurrence of F0 values in each semitone bin derived from the readings of six subjects taken individually—three from Jiuying Village and three from Taoyuan Village. Also indicated on each histogram are the semitone bins delimiting the octave band containing the largest number of F0 values in the subject's speech. (We note that some of the histograms are bimodal and hypothesize that this reflects the characteristics of the tones in the subjects' speech.) Taking all those from Taoyuan Village, the F0 values included in the octave bands comprised 98.91% of the total, and taking all those from

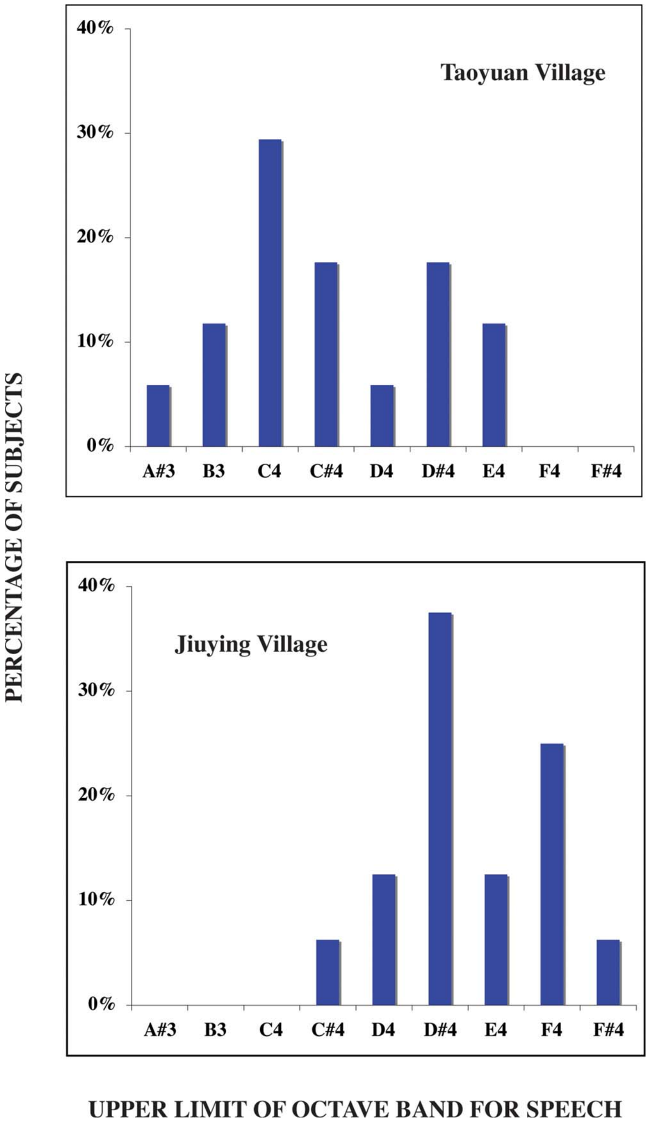


Fig. 2. (Color online) The upper limits of the octave band for speech in the two villages plotted in semitone bins. The center of each bin is displayed on the abscissa: A#3=233.1 Hz; B3=246.9 Hz; C4=261.6 Hz; C#4=277.2 Hz; D4=293.7 Hz; D#4=311.1 Hz; E4=329.6 Hz; F4=349.2 Hz; F#4=370 Hz.

Jiuying Village, these values comprised 97.24% of the total. Therefore, as had been found in the earlier studies referred to above, the F0 values in the speech of these subjects, taken individually, spanned close to an octave.

The F0s were found to be higher overall for subjects in Jiuying Village than those in Taoyuan Village; specifically, the F0s averaged over a log scale were 231.4 Hz for Jiuying Village and 200.6 Hz for Taoyuan Village. On a one-way analysis of variance (ANOVA), the difference in average F0s between the two villages was found to be highly significant [$F(1,31) = 19.106; p < 0.001$].

A further analysis was performed to test the hypothesis that the octave bands for speech would cluster within each village, but would differ significantly across villages. Figure 2 presents the percentages of subjects for whom the upper limit of the octave band fell in each semitone bin, plotted for each village separately. As can be seen, the values indeed clustered within each village, but differed overall across villages by roughly 3 semitones. On a one-way ANOVA, this difference between the two villages was found to be highly significant [$F(1,31) = 19.803$; $p < 0.001$].

4. Discussion

The present findings are in accordance with the conjecture that the overall pitch level of an individual's speaking voice varies as a function of his or her linguistic community and so reflects an influence of long-term exposure to the speech of others (Deutsch, 1992). In most previous work on this issue, comparison was made between the pitch levels of speech derived from passages that were read out in different languages (see, for example, Hollien and Jackson, 1973; Majewski *et al.*, 1972; Hanley *et al.*, 1966; Xue *et al.*, 2002). Since readings of different words were compared, this procedure introduced possible confounds. As an exception, Yamazawa and Hollien (1992) tested two groups of female speakers—one speaking primarily Japanese and the other speaking primarily American English—with all subjects reading out passages in both Japanese and English. The authors found that the Japanese speakers exhibited higher F0s than did the English speakers, though the differences between the two groups were more pronounced for passages read out in the speakers' native languages. The authors concluded that these F0 differences could be due to a number of factors, including ethnicity, culture, and the substantial differences in the structural characteristics of the Japanese and English languages.

In the present study, the subjects in the two communities are considered to be homogeneous ethnically and culturally, and their dialects are quite similar. They also read out the identical passage in Standard Mandarin so that no confound could have been introduced by differences in the words that were spoken. Our present findings are therefore in accordance with the hypothesis that the overall pitch level of a speaker's voice is influenced by a mental representation that is acquired through exposure to the speech of others. Assuming a relatively homogeneous linguistic community, such a representation would be particularly useful for tone languages. Here individual words assume different lexical meanings depending on the tones in which they are enunciated. For example, the first tone in Mandarin is high in pitch, and the word "ma" when spoken in this tone means "mother." In contrast the overall pitch level of the third tone is low, and the word "ma" spoken in this tone means "horse." An agreed-upon pitch level (taking male and female speech separately) would therefore facilitate the identification of individual tones and so the comprehension of individual words. Such a pitch representation would also be useful for speakers of nontone languages, for example, in facilitating speaker identification and evaluating the emotional tone of the speaker's voice. However, it remains to be determined whether effects similar to those found here occur in speakers of nontone languages also.

Acknowledgments

Diana Deutsch and Jinghong Le contributed equally to this paper and should both be considered first authors. The authors are grateful to Yu Chen and Rong Zhou for research assistance. They are also grateful to James Hillenbrand and two anonymous reviewers for helpful comments on an earlier version.

References and links

Blunden, C., and Elvin, M. (**1998**). *Cultural Atlas of China*, 2nd ed. (Checkmark Books, New York).
Collins, S. A. (**2000**). "Men's voices and women's choices," Anim. Behav. **60**, 773–780.
Deutsch, D. (**1992**). "Some new pitch paradoxes and their implications," Philos. Trans. R. Soc. London, Ser. B **336**, 391–397.
Deutsch, D., Henthorn, T., and Dolson, M. (**2004**). "Speech patterns heard early in life influence later perception of the tritone paradox," Music Percept. **21**, 357–372.
Dolson, M. (**1994**). "The pitch of speech as a function of linguistic community," Music Percept. **11**, 321–331.
Gonzalez, J. (**2004**). "Formant frequencies and body size of speaker: A weak relationship in adult humans," J. Phonetics **32**, 277–287.
Hanley, T. D., Snidecor, J. C., and Ringel, R. L. (**1966**). "Some acoustic difference among languages," Phonetica **14**, 97–107.

Hollien, H., and Jackson, B. (**1973**). "Normative data on the speaking fundamental frequency characteristics of young adult males," J. Phonetics **1**, 117–120.

Honorof, D. N., and Whalen, D. H. (**2005**). "Perception of pitch location within a speaker's F0," J. Acoust. Soc. Am. **117**, 2193–2200.

Hudson, A. I., and Holbrook, A. (**1982**). "Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading," J. Speech Hear. Res. **25**, 25–28.

Kunzel, H. J. (**1989**). "How well does average fundamental frequency correlate with speaker height and weight?," Phonetica **46**, 117–125.

Lass, N. J., and Brown, W. S. (**1978**). "Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies," J. Acoust. Soc. Am. **63**, 1218–1220.

Majewski, W., Hollien, H., and Zalewski, J. (**1972**). "Speaking fundamental frequency of Polish adult males," Phonetica **25**, 119–125.

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).

Van Dommelen, W. A., and Moxness, B. H. (**1995**). "Acoustic parameters in speaker height and weight identification: Sex-specific behavior," Lang Speech **38**, 267–287.

Xue, S. A., Hagstrom, F., and Hao, J. (**2002**). "Speaking fundamental frequency characteristics of young and elderly bilingual Chinese-English speakers: A functional system approach," Asia Pac. J. Speech, Lang. Hear. **7**, 55–62.

Yamazawa, H., and Hollien, H. (**1992**). "Speaking fundamental frequency pattern of Japanese women," Phonetica **49**, 128–140.

Fluid loading effects for acoustical sensors in the atmospheres of Mars, Venus, Titan, and Jupiter

T. G. Leighton
Institute of Sound and Vibration Research, University of Southampton, Highfield,
Southampton SO17 1BJ, United Kingdom
tgl@soton.ac.uk

Abstract: This paper shows that corrections for fluid loading must be undertaken to Earth-based calibrations for planetary probe sensors, which rely on accurate and precise predictions of mechanical vibrations. These sensors include acoustical instrumentation, and sensors for the mass change resulting from species accumulation upon oscillating plates. Some published designs are particularly susceptible (an example leading to around an octave error in the frequency calibration for Venus is shown). Because such corrections have not previously been raised, and would be almost impossible to incorporate into drop tests of probes, this paper demonstrates the surprising results of applying well-established formulations.

© 2009 Acoustical Society of America

PACS numbers: 43.40.Rj, 43.28.Vd, 43.28.Tc, 43.40.Fz, 43.30.Jx [JM]
Date Received: February 5, 2009 **Date Accepted:** March 2, 2009

1. Introduction

Only a handful of probes sent to other worlds have been equipped with instrumentation for recording extraterrestrial soundscapes.[1] Of those, only two[2–7] have returned data, and in both cases the passive acoustic data were dominated by wind noise, the pressure fluctuations caused by flow over the microphone surface (although active acoustics worked very well on the *Huygens* mission[4–7]). Greater chances of success in recording an alien soundscape require that probe designers take full account of acoustical capabilities, which, though established on Earth, require care in transposition to other worlds. Some solutions are so well-established that their basis can be found in commercial products (for example, the use of windshields or multiple microphones to distinguish acoustic signals from aerodynamic noise). In other cases, it is the esoteric nature of the alien world itself, which will bring into play unexpected acoustical phenomena. An example of the latter (fluid loading of structures) is explained in this paper, with specific attention to the effect of other worlds on acoustical sensors. Fluid loading is a well-established phenomenon, and this paper does not introduce new physics. Rather it undertakes calculations using established formulas to demonstrate that fluid loading on some planets will cause significant deviation in the performance of acoustical sensors of certain geometries, if that performance is calibrated on Earth and not corrected for fluid loading.

It is well-known that whenever a solid structure vibrates in an atmosphere, the gas, which is set into motion by the structural vibration, contributes to the inertia and damping associated with that oscillation. This usually increases both compared to the *in vacuo* characteristics of the vibrating component, but since Earth's atmosphere is denser than that of Mars, both are usually reduced when a structure is transposed from ground level on Earth to Mars. Both are, however, increased when structures are transposed from ground level on Earth to ground level in the denser atmospheres of Venus and Titan. From the viewpoint of the design of acoustical sensors for planetary probes, the importance of this lies in the fact that (i) the characteristics of a vibrating structure on Earth are sometimes taken to be intrinsic, which is to confuse them with the *in vacuo* characteristics; (ii) inclusion of the effect of dense atmospheres is commonplace in some areas of planetary probe design (such as parachutes or dirigibles, the effect of turbulence, etc.), but the effect of fluid loading in changing the natural frequency and damping of structures is less common; (iii) such effects will not be included in many Earth-based calibrations and tests, e.g., a drop test of a probe through Earth's atmosphere; and (iv) the effects are greatest on light, stiff structures, such as are common on planetary probes. These issues are particularly germane for acoustical sources and sensors (e.g., anemometers, sensors for atmospheric sound speed and dissipation, and microphones for probes and suits). This is especially the case when the sensors are not in free space, but embedded within tubes, a feature which has appeared in several designs for acoustical instrumentation for planetary probes.

There is a range of acoustical measurements, which may be undertaken to determine the properties of planetary atmospheres.[8] Currently there is considerable interest in acoustic anemometry, and the measurement of atmospheric dissipation, sound speed, and soundscapes.[8–11] Since the earliest proposals,[12,13] many designs mount acoustic transmitters and receivers in tubes or sample vessels in order to infer the properties of planetary gases through acoustical measurements[14] or acousto-optical methods[15] or flow excitation of the natural frequency of a gas-filled vessel.[8] Encapsulating the gas in an enclosure for measurement has many attractions: For example, it allows the gas temperature to be controlled, so that the sound speed can, given other constraints, be inverted to obtain the gas composition.[8] Perhaps the earliest reference[12] states the following: "The velocity of sound in a gas c is a function of T [temperature], M [molecular weight], and γ [specific heat ratio], and is given by $c^2=\gamma RT/M$, where R is the gas constant. This well-known relation has been used in the past in various techniques to measure the temperature $[T]$ of Earth's atmosphere, where M and γ were accurately known. It is proposed to reverse this method and bring a volume of the atmosphere of Venus into a thermostatically controlled tube where the temperature is known accurately and to determine M/γ by measuring the velocity of sound through the medium in the tube." Many of those proposing such acoustical instruments emphasize the need for the use of accurately preset frequencies and amplitudes, and while such geometries would not incur significant fluid loading in low-density atmospheres, they have been proposed for deployment on Venus, or Jupiter, and the planets beyond. For example, for a Venus probe Hanel and Strange[13] wrote the following: "The wave propagation in the unknown gas mixture is contained in a narrow channel that was cut in spiral form in a solid aluminum disk as shown in [their] Fig. 3. A small sonic transducer at the center of the disk generates a sound wave of constant amplitude, with a constant and precisely known audio frequency. Two identical condenser microphones, separated by several wavelengths, form part of the tube wall." This design, established in the 1960s, is still in current proposals in various forms, for example, in the design of a probe to the Jovian planets (Jupiter, Saturn, Uranus, and Neptune).[14] When sampling low-density fluids, radiation mass is not an issue, but when sampling dense gases or liquids,[16] then a range of acoustical phenomena, such as fluid loading and the coupling between the fluid and the walls, can become important. In addition to acoustical sensors, this can affect others that rely on mechanical vibration, for example, those which respond with high sensitivity to changes in the inertia or stiffness associated with vibrating surfaces as, for example, species accumulate upon an oscillating plate.[17]

This paper calculates the inertial effect of fluid loading at ground level on Venus, Mars, Titan, and at two locations in the atmosphere of Jupiter, on a range of structures, showing that the above designs, which mount sensors in tubes, will incur significant fluid loading.

2. Method

It is well-known that the force that drives a structure to vibrate in a gas encounters a mechanical impedance (Z_m+Z_r), which differs from the *in vacuo* input mechanical impedance (Z_m) because of the contribution of the radiation impedance $(Z_r=R_r+jX_r)$. The real component of Z_r is well-known as the radiation resistance, R_r, a positive value of which indicates an additional power dissipation by the source as a result of the presence of the fluid. The imaginary component of Z_r is the radiation reactance, X_r, a positive value of which indicates increased mass loading caused by the presence of the fluid. The inertia $(m+m_r)$ associated with the immersed oscillation differs from the *in vacuo* value m by the "radiation mass" $m_r=X_r/\omega$. If $m_r>0$ then for an oscillator of stiffness s the resonance frequency of the source is reduced from the *in vacuo* value f_0

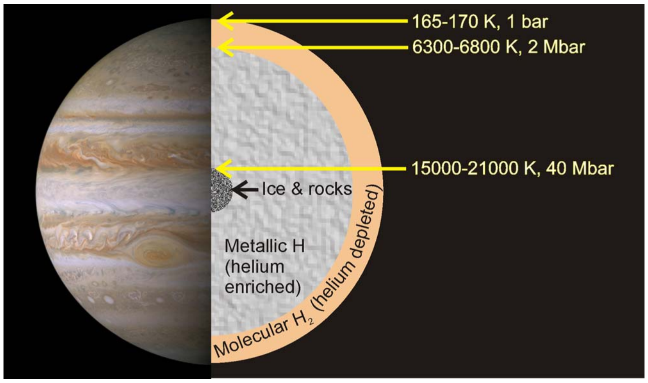


Fig. 1. (Color online) A schematic of the interior of Jupiter. The photographic elements of the image are credited to NASA, ESA, and the Hubble Heritage Team (AURA/STScI). The diagram indicates the pressures, temperatures, and composition of layers. The two points for which calculations are undertaken in the paper (close to the 1 bar level) are indistinguishable from one another on this scale (Ref. 28).

$=(1/2\pi)\sqrt{s/m}$ to the fluid loaded value $f_{\mathrm{atm}}=(1/2\pi)\sqrt{s/(m+m_r)}$, a change in $\Delta f/f_0=(f_{\mathrm{atm}}-f_0)/f_0=(\sqrt{m/(m+m_r)}-1)$, which tends to $-m_r/2m$ if $|m_r|\ll|m|$.

The effect of fluid loading on the natural frequencies of two versions of each of the following objects will be estimated for extraterrestrial locations: (i) a hollow steel sphere of outer radius $a=10$ cm ringing with spherical symmetry; (ii) a piston in a short pipe of length l, which opens out to the atmosphere through a hole in a plate (the hole having the same radius d as the pipe); and (iii) a length l of an infinite uniformly vibrating wire (of radius b). Assuming acoustically rigid structures and in the long wavelength limit ($ka\ll 1$, $kd<kl\ll 1$, and $kb\ll 1$ for wavenumber k), the radiation impedances of all these structure are primarily imaginary, such that the reactance dominates and the respective radiation masses equal: (i) $m_{r,\mathrm{sphere}}\approx 4\pi a^3\rho$, (ii) $m_{r,\mathrm{pipe}}\approx\rho\pi d^2 l+8\rho\pi d^3/(3\pi)$, and (iii) $m_{r,\mathrm{wire}}\approx\rho\pi b^2 l$.[18–20] The same principle can be applied to other geometries[21–23] and when the wavelength is not much larger than the structure (the Huygens sound speed sensor used 1 MHz), although the expressions are more complicated.

Recognizing that conditions (e.g., temperature T and static pressure P) can vary at a given altitude on a given world, the added mass depends primarily on the atmospheric density, ρ. The following nominal values are used for ρ for the atmospheres of Earth ($\rho=1.3$ kg m^{-3}), Mars ($\rho=0.02$ kg m^{-3}, with $T=220$ K, $P=0.007$ bar, and compositions of 95% CO_2, 2.7% N_2, and 300 ppm H_2O), Venus ($\rho=65$ kg m^{-3}, with $T=730$ K, $P=90$ bar, and compositions of 3.5% N_2 and 96% CO_2), and Titan ($\rho=5.5$ kg m^{-3}, with $T=95$ K, $P=1.6$ bar, and compositions of 95% N_2 and 5% CH_4).[24] These correspond to ground level averages since exploration by a lander is not atypical for these worlds.

Other candidate planets will require atmospheric probes. The upper atmospheres of Jupiter and Saturn are 90% hydrogen with ~10% helium and trace amounts of other compounds. Uranus and Neptune, in comparison, contain less hydrogen and helium, and more oxygen, carbon, nitrogen, and sulfur. Jupiter has a dense core of uncertain composition but probable existence,[25] surrounded by a 40 000-km-thick layer of liquid metallic hydrogen and some helium, which extends out to about 78% of the radius of the planet[26,27] (Fig. 1). At the top of this layer of liquid metallic hydrogen the temperature is 10 000 K and the pressure is 200 GPa. Droplets resembling rain of helium and neon precipitate down through the metallic hydrogen layer, depleting the abundance of helium and neon in the upper atmosphere.[28] Above the metallic hydrogen is a 21 000-km-thick layer of liquid hydrogen and gaseous hydrogen, with no sharp boundary between the two, called the interior atmosphere. The clouds (primarily of crystalline ammonia, ammonia hydrosulfide, and water) exist at the top of the atmosphere, in a layer that is around 50 km thick, where the atmospheric pressure is 20–200 kPa.

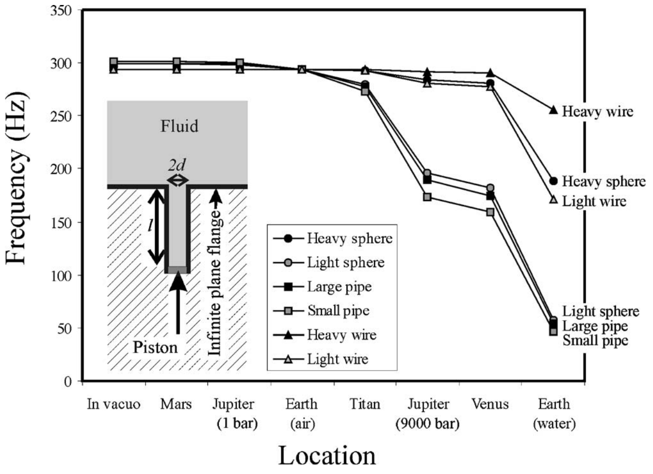


Fig. 2. The resulting natural frequency when added mass is included, for two hollow steel spheres with outer radii $a=10$ cm and wall thicknesses of 1 cm (a "heavy sphere" containing 8.74 kg of steel) and 0.5 mm (a "light sphere" of 0.48 kg steel). Ground-level atmospheres on Earth, Venus, Mars, and Titan, plus two locations on Jupiter, are used, with immersion in water at Earth's surface shown for comparison (a location that could be used for ground-truthing predictions). Predictions are also shown for two short pipes, open to the atmosphere, containing pistons (the large pipe has a length of 18 cm, diameter of 2 cm, and, at its base, contains a piston of mass 2 g; the short pipe has a length of 10 cm, diameter of 1.5 cm, and at its base contains a piston of mass 0.5 g). Predictions are shown for two wires, a heavy wire (10^{-2} kg/m, 1 mm radius) and a light wire (10^{-4} kg/m, 0.25 mm radius). The inset shows the geometry for the piston-driven pipe discussed in the text [which could represent a biomimetic sound source (e.g., vocal tract) or sensor (ear canal), or a component of a sampling device].

Calculations are made for two locations of possible interest for future probes to Jupiter: (i) at an equatorial radius of 71 492 km from Jupiter's center, where $P=1$ bar (105 Pa), $\rho=0.1$ kg m^{-3}, and $T\sim 165$ K;[29] and (ii) the estimated maximum operational penetration depth of some future very robust probe. Extrapolating from current terrestrial seismic sensors,[30] $P=0.9$ GPa would set a limit beyond the capability of such a sensor, and provide a useful point for comparison (see later). This pressure occurs 6.96×10^7 m from Jupiter's center, where $T\sim 2000$ K and $\rho\sim 50$ kg m^{-3}.[31]

3. Results

For comparative purposes, the artificial assumption is made that all the objects are tuned to vibrate at ground level on Earth at 293.66 Hz (the pitch of the note D) which, because the fluid loading on Earth differs across the range of objects tested, gives them different *in vacuo* frequencies (Fig. 2). Figure 2 shows to what extent fluid loading, taken in isolation to other effects (e.g., sound speed variations), changes the pitch for two versions of the different shapes (i)–(iii) described in Sec. 2 (details are given in the caption).

The choice of the altitude on Jupiter where the static pressure equals 0.9 GPa can now be seen in context, since although it represents an environment beyond the likely reach of near-future probes, from Fig. 2 the fluid loading there is less than that which instruments of the same geometry on Venus would experience. The effect is significant for some geometries: For example, the natural frequency of the piston-driven pipe on Earth is around twice the value it would have on Venus (a location to which acoustical sensors have already been sent[2,3]). The effect is smaller on Mars and of the opposite sign.

The magnitude of the effect depends primarily on two key parameters. For each structure, the one with the smaller volume-averaged density is more affected by fluid loading. This is because the first key parameter is the "buoyancy," the ratio of the mass of the structure to the mass of fluid it displaces when stationary [since the radiation mass is related to the latter, and (m/m_r) determines the value of $\Delta f/f_0$]. The second key parameter concerns the geometrical constraints, which determine how much of the atmosphere is set into motion when the oscillation occurs. Of the three structures, the wire causes primarily local fluid motion in a dipolelike manner, whereas the sphere is assumed to act as a monopole source of atmospheric motion, such the fluid velocity falls off to first order as an inverse square law from the fluid wall. The pipe, however, entrains a body of air to move in a one-dimensional manner, the fluid velocity not falling off with distance from the piston until outside of the pipe, giving it the greatest fluid loading of the three.

4. Discussion

The above calculations illustrate the effect of fluid loading, which can affect both the inertia and dissipation associated with the motion of a structure. It is important to stress that the fluid loading has been treated in isolation to other effects, which can influence the frequency responses of structures, such as the effect of thermal expansion on stiffness. In the specific case of sound sources, some sources are particularly affected by the atmospheric sound speed (such as organ pipes). It is interesting to note that the pitch of an organ pipe will increase on Venus compared to Earth, because the sound speed in the atmosphere there is greater than that on Earth, but that the pitch of the human voice (were it to be made operable on Venus) will decrease because there fluid loading, not the sound speed, is the dominant effect (see Ref. 32; the pipes used above roughly model the human adult and child vocal tract, and the two wires correspond to the dimensions and masses of heavy and light guitar strings).

Probe structures are lightweight where possible, so that (depending on the geometry) fluid loading could have significant effects on Venus and the outer planets, and make Earth-based calibrations incorrect. While the effect is smaller on other worlds, sensitive instrumentation based on monitoring of mechanical resonances[17] should account for the alien fluid loading when transposing Earth-based calibrations. Given the constraints for conserving power yet obtaining good signal-to-noise ratios, active acoustic sensors are often narrowband. Where fluid structure interaction is significant (for example, in coupling between fluid and container walls[33]), the frequency characteristics and dissipation of the fluid in the sample tube will differ from those found in the bulk atmosphere. Account must be taken of these effects when acoustic sensors are used in the dense atmospheres of other worlds.

5. Conclusion

This paper has outlined how fluid loading can affect the dissipation and inertia associated with the motion of a structure, and shown that, at least for Venus, the effects can be considerable, depending on the geometry and density of the structure. Fluid loading calculations should be undertaken for future acoustic systems on Venus and Titan.

Acknowledgments

I am greatly indebted to Chris Morfey for the stimulation he has provided on this topic and to Andi Petculescu for assisting in finding the gas densities for the worlds studied in this report. I am grateful to James Jiang and Kyungmin Baik for reading through the manuscript.

References and links

[1] T. G. Leighton and A. Petculescu, "Sounds in space: The potential uses for acoustics in the exploration of other worlds," Hydroacoustics **11**, 225–238 (2008).

[2] L. Ksanfomality, N. V. Goroschkova, and V. Khondryev, "Wind velocity near the surface of Venus from acoustic measurements," Cosmic Res. **21**, 161–167 (1983).

[3] L. V. Ksanfomality, F. L. Scarf, and W. W. L. Taylor, in *Venus*, edited by D. M. Hunten, L. Colin, T. M. Donahue, and V. I. Moroz (University of Arizona Press, Tucson, AZ, 1983), pp. 565–603.

[4] M. Fulchignoni, F. Ferri, F. Angrilli, A. J. Ball, A. Bar-Nun, M. A. Barucci, C. Bettanini, G. Bianchini, W. Borucki, G. Colombatti, M. Coradini, A. Coustenis, S. Debei, P. Falkner, G. Fanti, E. Flamini, V. Gaborit, R. Grard, M. Hamelin, A. M. Harri, B. Hathi, I. Jernej, M. R. Leese, A. Lehto, P. F. Lion Stoppato, J. J. López-Moreno, T. Mäkinen, J. A. M. McDonnell, C. P. McKay, G. Molina-Cuberos, F. M. Neubauer, V. Pirronello, R. Rodrigo, B. Saggin, K. Schwingenschuh, A. Seiff, F. Simões, H. Svedhem, T. Tokano, M. C. Towner, R. Trautner, P. Withers, and J. C. Zarnecki, "In situ measurements of the physical characteristics of Titan's environment," Nature (London) **438**, 785–791 (2005).

[5] J. C. Zarnecki, M. R. Leese, B. Hathi, A. J. Ball, A. Hagermann, M. C. Towner, R. D. Lorenz, J. Anthony, M.

McDonnell, S. F. Green, M. R. Patel, T. J. Ringrose, P. D. Rosenberg, K. R. Atkinson, M. D. Paton, M. Banaszkiewicz, B. C. Clark, F. Ferri, M. Fulchignoni, N. A. L. Ghafoor, G. Kargl, H. Svedhem, J. Delderfield, M. Grande, D. J. Parker, P. G. Challenor, and J. E. Geake, "A soft solid surface on Titan as revealed by the Huygens Surface Science Package," Nature (London) **438**, 792–795 (2005).

[6] A. Hagermann, P. D. Rosenberg, M. C. Towner, J. R. C. Garry, H. Svedhem, M. R. Leese, B. Hathi, R. D. Lorenz, and J. C. Zarnecki, "Speed of sound measurements and the methane abundance in Titan's atmosphere," Icarus **189**, 538–543 (2007).

[7] M. C. Towner, J. R. C. Garry, R. D. Lorenz, A. Hagermann, B. Hathi, H. Svedhem, B. C. Clark, M. R. Leese, and J. C. Zarnecki, "Physical properties of the Huygens landing site from the surface science package acoustic properties sensor (API S)," Icarus **185**, 457–465 (2006).

[8] R. D. Lorenz, "Speed of sound in outer planet atmospheres," Planet. Space Sci. **47**, 67–77 (1999).

[9] H. E. Bass and J. P. Chambers, "Absorption of sound in the Martian atmosphere," J. Acoust. Soc. Am. **109**, 3069–3071 (2001).

[10] H. U. Eichelberger, K. Schwingenschuh, G. Prattes, B. Besser, Ö. Aydogar, I. Jernej, H. I. M. Lichtenegger, R. Hofe, P. Falkner, and T. Tokano, "Acoustics of planetary atmospheres with active multi-microphone techniques, probing Titan," Geophys. Res. Abstr. **10**, EGU2008-A-09049 (2008).

[11] J.-P. Williams, "Acoustic environment of the Martian surface," J. Geophys. Res. **106**, 5033–5041 (2001).

[12] R. A. Hanel, "Exploration of the atmosphere of Venus by a simple capsule," NASA Technical Note TN D-1909, 1964.

[13] R. A. Hanel and M. G. Strange, "Acoustic experiment to determine the composition of an unknown planetary atmosphere," J. Acoust. Soc. Am. **40**, 896–905 (1966).

[14] D. Banfield, P. Gierasch, and R. Dissly, "Planetary descent probes: Polarization nephelometer and hydrogen ortho/para instruments," Proceedings of IEEE Aerospace Conference (2005), pp. 691–697.

[15] H. O. Edwards and J. P. Dakin, "Gas sensors using sensors and actuators, correlation spectroscopy compatible with fibre-optic operation," Sens. Actuators B **11**, 9–19 (1993).

[16] J. Powell, J. Powell, G. Maise and J. Paniagua, "NEMO: A mission to search for and return to Earth possible life forms on Europa," Acta Astronaut. **57**, 579–593 (2005).

[17] A. P. Zent, R. C. Quinn, and M. Madou, "A thermo-acoustic gas sensor array for photochemically critical species in the Martian atmosphere," Planet. Space Sci. **46**, 795–803 (1998).

[18] L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, 2nd ed. (Wiley, New York, 1962).

[19] F. J. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, London, UK, 1985), pp. 118–125.

[20] D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000), pp. 144–163.

[21] M. G. Junger and D. Feit, *Sound, Structures and Their Interaction*, 2nd ed. (MIT, Cambridge, MA, 1993).

[22] S. V. Sorokin and S. G. Kadyrov, "Modelling of nonlinear oscillations of elastic structures in heavy fluid loading conditions," J. Sound Vib. **222**, 425–451 (1999).

[23] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates. I. Theory and experiments," J. Acoust. Soc. Am. **109**, 1422–1432 (2001).

[24] A. Petculescu and R. M. Lueptow, "Fine-tuning molecular acoustic models: sensitivity of the predicted attenuation to the Lennard-Jones parameters," J. Acoust. Soc. Am. **117**, 175–184 (2005).

[25] T. Guillot, D. J. Stevenson, W. B. Hubbard, and D. Saumon, in *Jupiter: The Planet, Satellites and Magnetosphere*, edited by F. Bagenal, T. E. Dowling, and W. B. McKinnon (Cambridge University Press, Cambridge, 2004), Chap. 3.

[26] T. Guillot, "The interiors of giant planets: Models and outstanding questions," Annu. Rev. Earth Planet Sci. **33**, 493–530 (2005).

[27] L. T. Elkins-Tanton, *Jupiter and Saturn* (Chelsea, New York, 2006).

[28] J. J. Fortney, "The structure of Jupiter, Saturn, and Exoplanets: Key questions for high-pressure experiments," Astrophys. Space Sci. **307**, 279–283 (2007).

[29] A. Seiff, D. B. Kirk, T. C. D. Knight, L. A. Young, F. S. Milos, E. Venkatapathy, J. D. Mihalov, R. C. Blanchard, R. E. Young, and G. Schubert, "Thermal structure of Jupiter's upper atmosphere derived from the Galileo probe," Science **276**, 102–104 (1997).

[30] W. Bosum and J. H. Scott, "Interpretation of magnetic logs in Basalt, hole 418A," Proceedings of the Ocean Drilling Programme, Scientific Results (1988), Vol. **102**, pp. 77–94.

[31] W. B. Hubbard, "Thermal models of Jupiter and Saturn," Astrophys. J. **155**, 333–344 (1969).

[32] T. G. Leighton and A. Petculescu, "The sound of music and voices in space," Acoust. Today (in press).

[33] V. A. Del Grosso, "Analysis of multimode acoustic propagation in liquid cylinders with realistic boundary conditions—Application to sound speed and absorption measurements," Acustica **24**, 299–311 (1971).

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.

Radiation force calculation for oblique ultrasonic beams (L)

K. Beissner
Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany

(Received 21 January 2009; revised 12 February 2009; accepted 20 February 2009)

The acoustic radiation force exerted on a perfect absorber in a lossless fluid has recently been calculated for the case of a rectangular transducer emitting a static (i.e., "frozen") ultrasonic field in the forward direction. The calculation is extended here, at least approximately, to the case of an oblique beam. This is important for measuring the ultrasonic power of scanning diagnostic devices. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097839]

PACS number(s): 43.25.Qp, 43.80.Vj [CCC] Pages: 2827–2829

I. INTRODUCTION

The acoustic radiation force exerted on a large perfect absorber in a lossless fluid has recently been calculated[1] for the case of a rectangular transducer emitting a static (i.e., arrested or "frozen") ultrasonic field in the forward direction. The result is important for measuring the ultrasonic power emitted by medical ultrasonic devices. IEC 61161 (Ref. 2) recommends the radiation force balance technique and the use of water as the sound-propagating medium. The relation between F, the component of the radiation force in the direction of the force-measuring device, and the time-averaged ultrasonic power P is to be provided by theory. This relation is

$$P = cF \quad \text{or} \quad cF/P = 1 \tag{1}$$

for a plane-progressive wave that is collinear with the force-measuring device; c is the speed of sound in the sound-propagating fluid (water). Deviations from Eq. (1) due to diffraction and focusing have been calculated in Ref. 1. This will be briefly extended here to angular beam scanning.

Diagnostic ultrasound devices are widely used all over the world. They are usually equipped with rectangular linear-array transducers, whose beam is automatically scanned over the region of interest in order to obtain as much information as possible. So far there has not been a generally accepted standard as to specifying how the emitted power can be measured directly in the scanning mode. At present there is an effort in IEC TC 87 "Ultrasonics" to accomplish this task. The present paper is intended to briefly provide the theoretical background.

The radiation force exerted by an oblique beam has been investigated in literature.[3,4] Reference 3 deals with a reflecting target; the corresponding absorber result is not mentioned explicitly but can be inferred by additional reasoning. Reference 4 deals with the reflector and the absorber case. The treatments in both references are restricted to plane waves that exhibit only one uniform direction of the energy and momentum transport. Real ultrasonic fields, however, have a three-dimensional structure due to diffraction and/or focusing (this applies particularly to the fields from diagnostic devices), and the present paper briefly investigates the result of the integration over the various directions in the case of an oblique beam. Reflecting targets are not dealt with here, as already explained in Ref. 1.

The same terminology as in Ref. 1 will be used. A plane-rectangular transducer with dimensions of $2a$ and $2b$ in the x and y directions, respectively, is assumed. The origin of the x-y-z coordinate system is at the transducer center. A hemispherical absorber (with an inner radius of r) in the geometric far-field range is considered (see Fig. 1 in Ref. 1). If $I(r,\theta,\varphi)$ is the radial component of the time-averaged intensity in spherical coordinates, the ultrasonic power is given by the surface integral of the far-field intensity over the absorber front face, i.e., by

$$P = \int\int r^2 I(r,\theta,\varphi) \sin\,\theta d\theta d\varphi. \tag{2}$$

The radiation force F is obtained as the surface integral of the intensity divided by c, but now weighted with cos τ, i.e., as

$$cF = \int\int r^2 I(r,\theta,\varphi) \sin\,\theta \cos\,\tau d\theta d\varphi, \tag{3}$$

where τ is the angle between the local direction of incidence and the direction of the force-measuring device. If, as in Ref. 1, the direction of the force-measuring device is assumed to coincide with the z axis, τ is simply given by $\tau=\theta$. The final result is the characteristic ratio cF/P.

It should be noted that the particular target configuration is chosen here for calculational purposes, and it is different

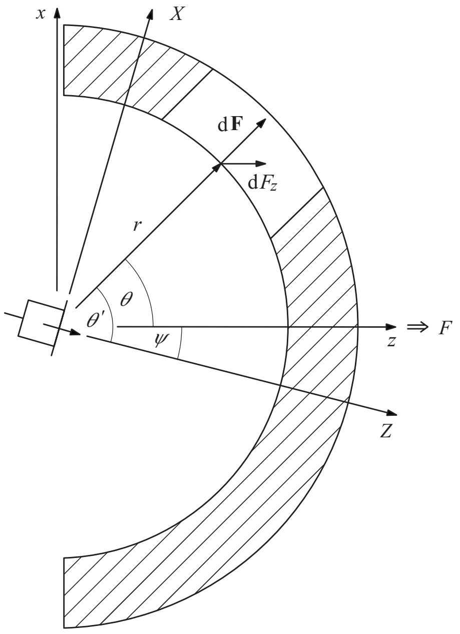


FIG. 1. Schematic section through the arrangement considered with ultrasonic transducer (left), absorbing target (hatched), old coordinates x and z (case A), and new coordinates X and Z (case B). Note that this is the azimuth plane ($\varphi=\varphi'=0$) and that in this special case, the relation between the angle values is $\tau=\theta=\theta'-\psi$, whereas in general these angles do not lie in one plane and the relation is as in Eq. (4).

from those used in practice. The result, however, is the same, as discussed to some extent in Ref. 1, along with a number of theoretical assumptions involved.

II. OBLIQUE BEAM

The treatment so far is independent of the beam direction. In Ref. 1 it was assumed that the beam axis coincides with the z axis. This will be referred to here as case A and the intensity in Eqs. (2) and (3) as I_A. We now consider case B where the beam again is arrested, but the beam axis is tilted from the z axis by an angle ψ toward the negative x axis. This can be treated within the algorithm of Ref. 1 by including an additional phase function in the expression for v_n, which is the normal velocity of the transducer surface.

However, it is not intended to repeat the entire algorithm here. A simpler procedure is chosen as follows. A new coordinate system X-Y-Z is considered (see Fig. 1 of the present paper). The Y axis is the same as the previous y axis. The Z axis coincides with the new beam axis; i.e., it is tilted from the z axis by an angle ψ toward the negative x axis. And the X axis is tilted accordingly. Spherical coordinates in the new system are R, θ', and φ', with $R=r$. The radial far-field intensity is now $I_B(r,\theta',\varphi')$.

The formula for the power P_B is the same as in Eq. (2); only I is replaced with I_B, θ is replaced with θ', and φ is replaced with φ'. The radiation force F_B is now different as the direction of the force-measuring device (in the z direction) does not coincide with the Z axis but is tilted from the Z axis by an angle ψ toward the positive X axis. Three symbols have to be replaced in Eq. (3) as before. The angle τ is no longer equal to θ' but can be found from simple geometric relations to follow

$$\cos\tau=\sin\theta'\cos\varphi'\sin\psi+\cos\theta'\cos\psi. \tag{4}$$

When this is inserted into an equation like Eq. (3) but with variables θ' and φ', the integral is the sum of two integrals. We shall initially consider the first one. The integration with respect to φ' runs from 0 to 2π or from $-\pi$ to π. The integrand (regarding φ') is a function $I_B(r,\theta',\varphi')\cos\varphi'$. We assume that the far-field intensity $I_B(r,\theta',\varphi')$ is symmetric with respect to the Y-Z plane. In this case the first integral vanishes for symmetry reasons. The remaining integral is then

$$cF_B=\cos\psi\int\int r^2 I_B(r,\theta',\varphi')\sin\theta'\cos\theta' d\theta' d\varphi'. \tag{5}$$

We assume that irrespective of the propagation direction, the ultrasonic beam in itself is identical in case A and case B, which means that $I_B(r,\theta',\varphi')$ is equal to $I_A(r,\theta,\varphi)$. The final result is then

$$P_B=P_A \quad \text{and} \quad F_B=F_A\cos\psi. \tag{6}$$

The integration limits need some additional consideration. In Ref. 1 and in case A above, the θ-φ integration was over the entire positive half-space, i.e., for θ from 0 to $\pi/2$. Applying this also to the θ'-φ' integration in case B (i.e., θ' from 0 to $\pi/2$) would require that the (hemispherical) target be tilted with the beam. Tilting the target together with the scanning beam, however, is almost impossible in practice; therefore the following additional condition should apply. The field distribution in the x-z plane or the X-Z plane, i.e., in the scan or azimuth plane (see Fig. 1), should be characterized by a limiting angle value θ_{lim} (or θ'_{lim}) so that the far-field intensity is only relevant inside $\pm\theta_{\text{lim}}$ with respect to the beam axis and can be neglected outside $\pm\theta_{\text{lim}}$. And this θ_{lim} should be equal to or less than $\pi/2-|\psi|$.

The above-mentioned theoretical condition corresponds to the practical aspect that the relevant part of the field in its lateral extent should be limited so that it is entirely intercepted by the target used, even under conditions of an inclined beam. This and other practical aspects, however, are not discussed in detail here.

Summarizing Ref. 1 and the present paper, it can be stated that the radiation force formula for an absorbing target is influenced by three effects, namely, diffraction, focusing, and scanning. Each of them leads to a decrease in the ratio cF/P from the plane-wave value of 1. The scanning effect can simply be taken into account by a factor of $1/\cos\psi$ to be applied to the measured radiation force F_B. Applying this factor, however, does not in general lead directly to the ultrasonic power but to a radiation force F_A that has still to be corrected for the two other effects, namely, diffraction and/or focusing.[1]

III. SCANNING

So far, only one forward beam (case A) and one oblique beam (case B) have been considered and compared. In practice, scanning means that the system produces a number of beams, say, n beams under n tilt angles ψ_i. Let us assume (a) that all beams and their power outputs are equal, irrespective of their direction, (b) that each beam is activated for the same time interval before the system switches to the next beam, and (c) that this time interval is much smaller than the reaction time of the radiation force balance so that the balance measures the temporal average radiation force F. If F_A is the radiation force produced by the same beam if arrested and emitting all the time in the forward direction ($\psi=0$), then

$$F = F_A \frac{1}{n}\sum_{i=1}^{n} \cos \psi_i = F_A \overline{\cos \psi}, \tag{7}$$

with

$$\overline{\cos \psi} = \frac{1}{n}\sum_{i=1}^{n} \cos \psi_i . \tag{8}$$

Finally, the case is dealt with that the scan is over a large number of equidistant ψ values so that it can be considered a quasi-continuous scan from ψ_1 to ψ_2. Then

$$\overline{\cos \psi} = \int_{\psi_1}^{\psi_2} \cos \psi d\psi \Bigg/ \int_{\psi_1}^{\psi_2} d\psi = (\sin \psi_2 - \sin \psi_1)/(\psi_2 - \psi_1). \tag{9}$$

If the scan is in a symmetric way from $\psi_1 = -\psi_0$ to $\psi_2 = \psi_0$, then

$$\overline{\cos \psi} = \sin \psi_0 / \psi_0 = \text{sinc } \psi_0, \tag{10}$$

using the sinc function that has already been explained in Ref. 1. Values of ψ appearing in the denominator of Eq. (9) or Eq. (10) are to be understood in radians.

Applying the factor $1/\overline{\cos \psi}$ either in accordance with Eq. (8) or with Eq. (9) or Eq. (10), the measured radiation force F can thus be transferred into an equivalent radiation force F_A (of a forward beam), and this can be transferred into an equivalent ultrasonic power, not simply using the plane-wave relation (1) but taking into account the influences of diffraction and/or focusing as described in Ref. 1. This final power value is then the desired result.

IV. CONCLUDING REMARKS

If the scan is in the y direction, some formal details may be different from the above considerations, but the final result is equivalent to Eq. (6). Moreover, the above treatment is not restricted to rectangular transducers but, in principle, applies to transducers of any shape.

The result expressed by Eq. (6) turns out to be equivalent to the plane-wave result in Refs. 3 and 4. The meaning, however, is different. Whereas the field in Refs. 3 and 4 is characterized by just one direction of incidence, namely, that of a plane wave, ψ here characterizes the direction of the beam axis, quasi the "average" of a number of various directions of incidence. It could have been assumed that taking the average is sufficient and that all the other directions of incidence due to diffraction and/or focusing are included in the average effect. But this is not the case. Equation (5) shows that the spread of incidence directions due to diffraction and/or focusing has to be taken into account in addition to the inclination effect that is represented by the factor $\cos \psi$.

ACKNOWLEDGMENTS

The author would like to thank R. T. Hekkenberg, Miwa Naito, K. Sandstrom, and Y. T. Kim of Working Group 8 of IEC TC 87 for discussions regarding formulas similar to Eqs. (8) and (10). It is shown in the present paper how these formulas can be derived from Ref. 1.

[1] K. Beissner, "Radiation force calculations for ultrasonic fields from rectangular weakly focusing transducers," J. Acoust. Soc. Am. **124**, 1941–1949 (2008).

[2] *IEC 61161: Ultrasonics—Power Measurement—Radiation Force Balances and Performance Requirements*, 2nd ed. (International Electrotechnical Commission, Geneva, 2006).

[3] F. E. Borgnis, "Acoustic radiation pressure of plane-compressional waves at oblique incidence," J. Acoust. Soc. Am. **24**, 468–469 (1952).

[4] E. M. J. Herrey, "Experimental studies on acoustic radiation pressure," J. Acoust. Soc. Am. **27**, 891–896 (1955). Please note that the conclusion on p. 895 that "beam spreading does not affect results since the observed radiation force is independent of angle of incidence" is not correct. The radiation force exerted on a perfect absorber by a certain part of the field lies indeed in the propagation direction of that part of the wave and is independent of the absorber orientation, but its contribution to the measured radiation force depends on the angle with the direction of the force-measuring device, and, therefore, the total radiation force depends on the spread of directions. It is just this effect that was investigated in Ref. 1 and in other papers cited there.

Sound absorption by Menger sponge fractal (L)

Tetsuji Kawabe,[a] Takatsuna Miyazaki,[b] Daisuke Oka, Sin'ichiro Koyanagi,[c] and Atsushi Hinokidani

Department of Physics and Department of Acoustic Design, Kyushu University, Shiobaru, Fukuoka 815-8540, Japan

(Received 30 October 2007; revised 6 January 2009; accepted 9 February 2009)

For the purpose of investigation on acoustic properties of fractals, the sound absorption coefficients are experimentally measured by using the Menger sponge which is one of typical three-dimensional fractals. From the two-microphone measurement, the frequency range of effectively absorbing sound waves is shown to broaden with degree of fractality, which comes from the fractal property of the homothetic character. It is shown that experimental features are qualitatively explained by an electrical equivalent circuit model for the Menger sponge.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3095807]

PACS number(s): 43.55.Ev, 43.20.Bi, 43.55.Rg, 43.20.Mv [LLT] Pages: 2830–2833

It is well-known that the fractal geometry provides a precise model of complicated structures in nature.[1] For acoustics, with the aim of investigating the acoustic properties of strongly irregular systems, there are numerical and theoretical studies on the sound propagation in the prefractal waveguide,[2] the sound vibration in fractal drum, and fractal cavities.[3]

For electromagnetics, it has recently reported that the electromagnetic waves can be confined by a three-dimensional fractal called the Menger sponge.[4] The localization of electromagnetic waves is clearly observed in the Menger sponge that is constructed up to stage 3 of the self-similar structure. Furthermore, the attenuation of the wave is also observed both for reflection and transmission intensity.

The purpose of this paper is to experimentally study the acoustic properties of this Menger sponge. From the measurement of frequency response such as the sound absorption coefficient, we show that the frequency range corresponding to the absorption peaks becomes broader owing to the fractal structure of the Menger sponge. We also estimate the resonance frequencies by analytically using an equivalent electrical circuit model for the Menger sponge.

The Menger sponge is a three-dimensional extension of the Cantor bar and Sierpinski carpet fractals.[1] Figure 1(a) shows the Menger sponge of stage 1 which is made by first dividing into $27(=3\times 3\times 3)$ identical cubic pieces, and then by extracting $7(=1+6)$ pieces at the body (1) and face centers (6). Similarly, the Menger sponge of stage 2 is obtained by repeating the same extraction process for the 20 remaining pieces, as shown in Fig. 1(b). The fractal dimension D_f is defined by $D_f=\log n/\log r$ based on the relation $n=r^{D_f}$, where n is the number of the self-similar units newly created when the size of the initial unit decreases to $1/r$.[1] The Menger sponge is characterized by $n=20$ and $r=3$ so that $D_f=\log 20/\log 3=2.7268\cdots$. For the present investigation, we made the Menger sponge from an acrylic material cube. The initiator in the construction is a cube with a length $l=90$ mm so that the outer dimension of the Menger sponge is $90\times 90\times 90$ mm^3. We constructed the Menger sponges of stages 1 and 2 by gluing acrylic cubes ($10\times 10\times 10$ mm^3), as shown in Figs. 1(c) and 1(d), respectively.

We measured a perpendicular sound absorption coefficient α of the Menger sponge by using the two-microphone method,[5] as shown schematically in Fig. 2. In this method, the sample to be measured is placed at D mm away from the one end of a straight tube. A sound source is connected at the opposite end of the tube. A pair of microphones is mounted flush with the inner wall of the tube near the sample end of the tube. The α value can be determined from the complex pressure reflection coefficient R at microphone location 1 as follows:

$$\alpha = 1-|R|^2 = 1-\left|\frac{H_{12}-e^{-jks}}{e^{jks}-H_{12}}\right|^2. \tag{1}$$

Here, H_{12} is a transfer function between acoustic pressures at two microphones with a distance of s, $j=\sqrt{-1}$, and $k=2\pi f/c$ is the wave number, where f and c are the frequency and the speed of sound, respectively.

The α values of the Menger sponge for stages 1 and 2 were measured in the frequency range [85, 1500] Hz for the three cases of D (0, 30, 60 mm) in Fig. 2. From the measurement done at 24.5 °C, we have found the feature that stage 2 exhibits a broader resonance but a smaller absorption than stage 1. Furthermore, we have found that this feature qualitatively remains for three values of D. Thus, we present and analyze only the case of $D=0$ in the present paper. Figure 3 shows the dependence of α on f, from which the resonance frequency is $f_0\approx 627$ Hz for stage 1, and $f_0'\approx 691$ Hz for stage 2.

In order to see whether these resonance values are analytically derived from the Menger sponge or not, we try to estimate them by using an electrical equivalent circuit for the Menger sponge. As the electrical equivalent circuit suitable for the Menger sponge, we adopt a lossy cylindrical tube

[a] Electronic mail: kawabe@design.kyushu-u.ac.jp

[b] Present address: FOR-A Company Limited, RD Center, 2-3-3 Osaku, Chiba 285-8580, Japan.

[c] Present address: Takenaka R&D Institute, 1-5-1 Otsuka, Inzai, Chiba 270-1395, Japan.

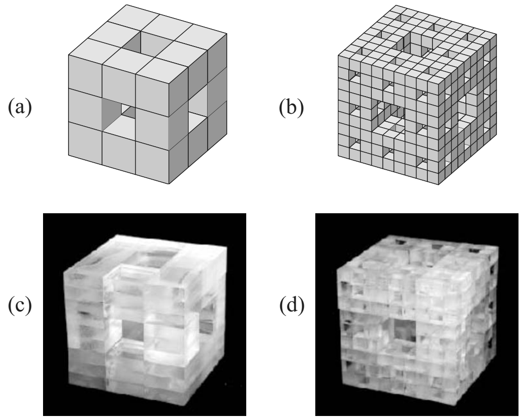


FIG. 1. Illustration of the Menger sponge of (a) stage 1 and (b) stage 2, and the model of the Menger sponge made by acrylic material: (c) stage 1 and (d) stage 2.

model,[6] as shown in Fig. 4(a). In this model, the sound is assumed to propagate in the non-uniform tube composed of N lossy tubes with length l_i $(i=1,2,\ldots,N)$. The energy losses are assumed to occur at the wall through viscous friction, which is almost the same as the assumption used in Ref. 3 that the walls present a small but finite specific admittance. As shown in Fig. 4(b), the characteristics of its propagation are described by an elementary electric circuit with a resistance R_i, an inductance L_i, a capacitance C_i, and a shunt conductance G_i, which are the quantities per unit length. For a piece of line l_i in length, the relation between the sending-end voltage E_i and current I_i and the receiving-end E_{i+1} and I_{i+1} is given by the transfer matrices as follows:

$$\begin{pmatrix} E_1 \\ I_1 \end{pmatrix} = T\begin{pmatrix} E_{N+1} \\ I_{N+1} \end{pmatrix} = \left[\prod_{i=1}^{N}\begin{pmatrix} \cosh \gamma_i l_i & Z_0(i)\sinh \gamma_i l_i \\ Y_0(i)\sinh \gamma_i l_i & \cosh \gamma_i l_i \end{pmatrix}\right]\begin{pmatrix} E_{N+1} \\ I_{N+1} \end{pmatrix}, \tag{2}$$

where the characteristic impedance $Z_0(i)=\sqrt{z_i/y_i}$, the admittance $Y_0(i)=\sqrt{y_i/z_i}$, and $\gamma_i=\sqrt{z_i y_i}$ are functions of the angular frequency $\omega=2\pi f$ through $z_i=R_i+j\omega L_i$ and $y_i=G_i+j\omega C_i$. Here we have used the transfer matrix of a one-dimensional acoustical system[2] because the sound absorption was measured by using plane waves in tubes which the Menger sponge consists of.

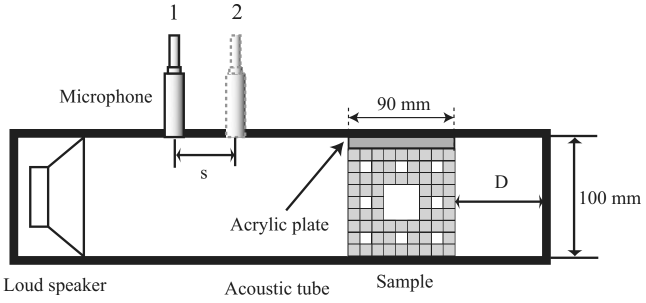


FIG. 2. Schematic illustration of the two-microphone measurement system for observing the sound absorption property of the Menger sponge. The acrylic plate is to fill a gap between the waveguide and the Menger sponge.

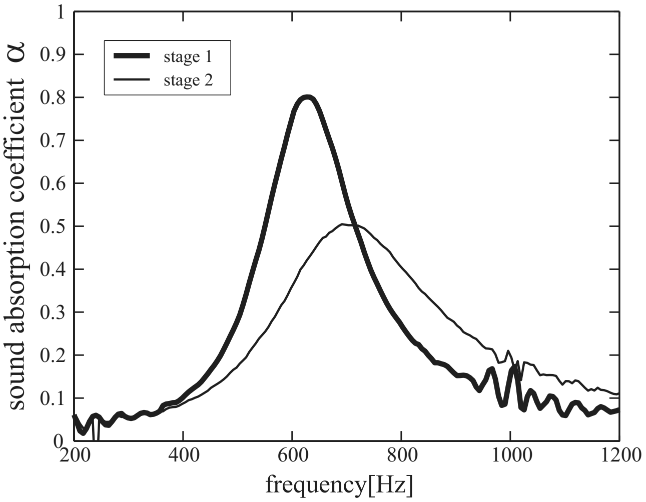


FIG. 3. Perpendicular sound absorption coefficient α for stages 1 and 2, which are measured in the frequency range [85, 1500] Hz.

The sound pressure P_i is analogous to the voltage E_i and the volume velocity U_i is to the current I_i in an electrical line, and the correspondence between electric elements and acoustic elements are as follows:

$$L_i=\frac{\rho}{A_i(N)}, \quad C_i=\frac{A_i(N)}{\rho c^2}, \quad R_i=\frac{S_i(N)}{A_i^2(N)}\sqrt{\frac{\omega\rho\mu}{2}},$$
$$G_i=S_i(N)\frac{\eta-1}{\rho c^2}\sqrt{\frac{\kappa\omega}{2c_p\rho}}, \tag{3}$$

where $A_i(N)$ is the tube area, $S_i(N)$ the tube circumference, ρ the air density, μ the viscosity coefficient, κ the coefficient of heat conduction, η the adiabatic constant, and c_p the specific heat of air at constant pressure. For the Menger sponge of stage 1 $(N=3)$, the three lossy cylindrical tubes have the

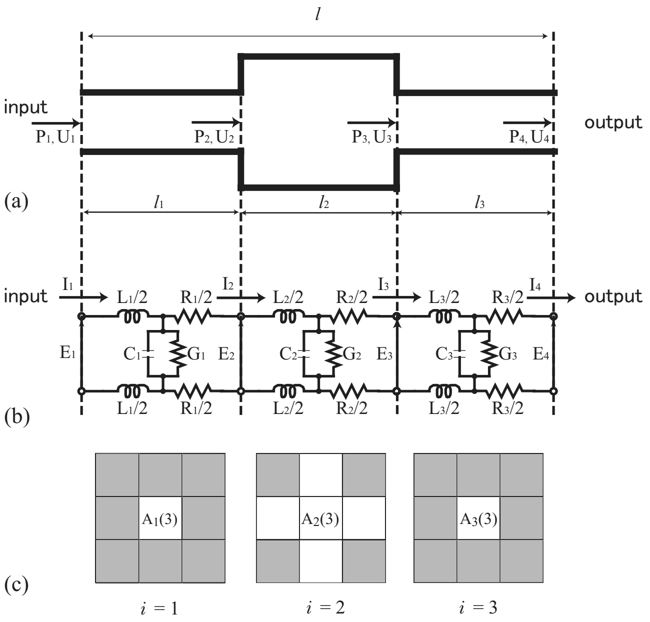


FIG. 4. Model for the Menger sponge of stage 1 $(N=3)$: (a) the lossy cylindrical tube model, (b) the electrical equivalent circuit, and (c) three cross sections $A_i(N)$ $(i=1,2,3)$. All notations used here are explained in the text.

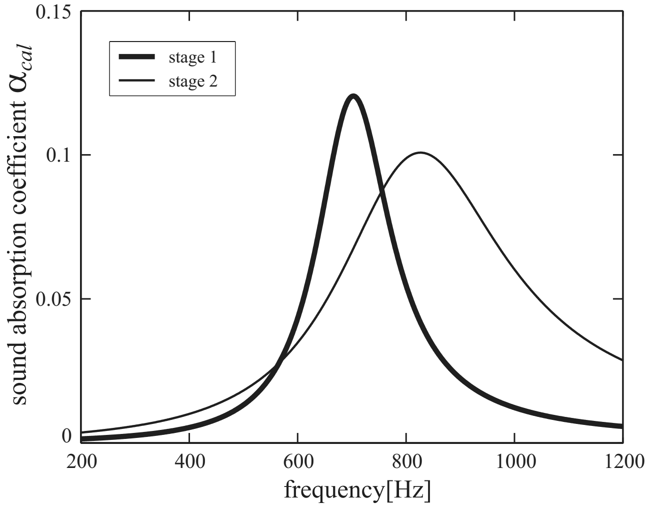


FIG. 5. Absorption coefficient α_{cal} for stages 1 and 2, which are calculated from Eq. (4) with Eqs. (2) and (3).

cross section $A_i(N) \equiv n_i(N) l_i^2 = n_i(N) l^2 / N^2$ and the circumference $S_i(N) \equiv m_i(N) l_i = m_i(N) l / N$, as shown in Fig. 4(c), where $n_1(3) = n_3 = 1$, $n_2(3) = 5$, and $m_1(3) = m_3 = 4$, $m_2(3) = 12$. Similarly, the Menger sponge of stage 2 ($N=9$) is modeled by the nine lossy cylindrical tubes with different $n_i(9)$ and $m_i(9)$ as $n_1(9) = n_3 = n_7 = n_9 = 17$, $n_2(9) = n_4 = n_6 = n_8 = 49$, $n_5(9) = 65$, and $m_1(9) = m_3 = m_7 = m_9 = 44$, $m_2(9) = m_8 = 84$, $m_4(9) = m_6 = 52$, $m_5(9) = 68$.

For the calculation of the absorption coefficient, we use the acoustic impedance $Z = l^2 P_1 / U_1 = l^2 E_1 / I_1$ at input section of the tube with area l^2. From the reflection coefficient $R = (Z - \rho c)/(Z + \rho c)$, the absorption coefficient $\alpha_{cal} = 1 - |R|^2$ is given by

$$\alpha_{cal} = 1 - \left| \frac{l^2 E_1 - \rho c I_1}{l^2 E_1 + \rho c I_1} \right|^2 . \tag{4}$$

In the present experiment, the end of the waveguide behind the Menger sponge is closed, as shown in Fig. 2. This condition is satisfied by imposing a zero volume velocity ($U_{N+1} = 0$) at output section so that we put the output current $I_{N+1} = 0$ as a boundary condition in Eq. (2).

Figure 5 shows the results of α_{cal} for stages 1 and 2. Although the peak values of α_{cal} are smaller than those of α in Fig. 3, Fig. 5 qualitatively agrees with the experimental feature. From Fig. 5, the resonance frequency is $f_0 = 703$ Hz for stage 1, and $f_0' = 827$ Hz for stage 2. By taking the open end correction into consideration, we can obtain the improved values as $f_0 = 609$ Hz and $f_0' = 658$ Hz for stages 1 and 2, respectively, so that the difference between the theoretical values and the measured ones is sizably reduced. Here, we have used the parameters for air at 24.5 °C in Eq. (3) as follows: $\rho = 1.171$ kg/m^3, $c = 346.40$ m/s, $\mu = 18.4 \times 10^{-6}$ Pa s, $\kappa = 0.0259$ W/(m K), $\eta = 1.4$, and $c_p = 1.0083$ J/(kg K).

In conclusion, we have experimentally studied the property of the sound absorption by the Menger sponge with the two-microphone method in order to investigate the acoustic characteristics of fractal objects. Then we have shown that the experimental features of α can be approximately explained by the lossy cylindrical tube model, in spite of the simple model. From these experimental studies and analytical ones, we will deduce the conclusion that the Menger sponge can absorb the sounds in broad band of frequency due to the self-similarity of fractals and the hierarchical scales of length. This conclusion is consistent with the existence of the forbidden frequencies in a one-dimensional Cantor waveguide with self-similarity[2] and a wave trapping phenomenon which means a strong localization introducing stop-band in such structure.[7]

Let us briefly comment on our results. First, we would like to show that the resonance curve of Fig. 5 is approximated by the Lorentzian[8] whose analytic expression is helpful to understand the role of the electric elements in Eq. (3). Under the approximations that $\cosh \gamma_i l_i \approx 1$ and $\sinh \gamma_i l_i \approx \gamma_i l_i$ and the small loss conditions that $R_i \ll \omega L_i$ and $G_i \ll \omega C_i$, the components T_{11} and T_{21} of the transfer matrix T, which are necessary to estimate $E_1 = T_{11} E_4$ and $I_1 = T_{21} E_4$, are given by $T_{11} = 1 + P_2 Q_1 + P_1 Q_2 + P_1 Q_1$ and $T_{21} = 2Q_1 + Q_1^2 P_2 + Q_2 \approx 2Q_1 + Q_2$, where $P_i \equiv Z_0(i) \gamma_i l_i = (R_i + j\omega L_i) l_i$ and $Q_i \equiv Y_0(i) \gamma_i l_i = (G_i + j\omega C_i) l_i$. In the derivation, we neglect all higher terms with $\mathrm{O}(l_1^3)$. As $P_i Q_k \approx -L_i C_k \omega^2 l_i l_k + j(L_i G_k + R_i C_k) \omega l_i l_k$, the real part of the denominator in Eq. (4) is written by $\Re T_{11} + (\rho c / l^2) \Re T_{21} \approx \Re T_{11} = 1 - (L_1 C_1 + L_1 C_2 + L_2 C_1) l_1^2 \omega^2 \equiv 1 - \omega^2 / \omega_0^2$. Note that $\Re T_{11} \approx -2(\omega - \omega_0)/\omega_0$ at $\omega \approx \omega_0$. Thus, by putting $\Re T_{11}|_{\omega=\omega_0} = 0$ in the numerator of Eq. (4), the absorption coefficient α_{cal} can be approximated by the Lorentzian as follows:

$$\alpha_{cal}(\omega) = \frac{\alpha_M \Gamma^2}{(\omega - \omega_0)^2 + \Gamma^2}, \tag{5}$$

where $\alpha_M = (\omega_0^2 / \Gamma^2)(\rho c / l^2)(\Im T_{11}|_{\omega=\omega_0})(\Im T_{21}|_{\omega=\omega_0})$ and $\Gamma = (\omega_0 / 2)(\Im T_{11}|_{\omega=\omega_0} + (\rho c / l^2) \Im T_{21}|_{\omega=\omega_0})$. For stage 1, we have $\Im T_{11}|_{\omega=\omega_0} = (L_1 G_1 + L_1 G_2 + L_2 G_1 + C_1 R_1 + C_1 R_2 + C_2 R_1) \times \omega_0 l_1^2$ and $\Im T_{21}|_{\omega=\omega_0} = (2C_1 + C_2) \omega_0 l_1$. From Eq. (5), we can understand how the resonance curve depends on the quantities as L_i, C_i, R_i, and G_i. It should be noticed that ω_0 is determined by the summation of the products $L_i C_k$ as $\omega_0 = (l_1)^{-1} (L_1 C_1 + L_1 C_2 + L_2 C_1)^{-1/2}$, which is regarded as a generalization of the resonance frequency $(LC)^{-1/2}$ in the ordinary LCR circuit.

Second, we would like to analytically estimate the resonance frequencies. For stage 1 with $l_1 = l/3 = 0.03$, we obtain $f_0 = 738$ Hz. On the other hand, for stage 2 with $l_1' = l_1 / 3 = 0.01$, we obtain $f_0' = 778$ Hz from

$$\omega_0' = (l_1')^{-1} (6L_1'C_1' + 8L_1'C_2' + 2L_1'C_5' + 8L_2'C_1' + 6L_2'C_2' + 2L_2'C_5' + 2L_5'C_1' + 2L_5'C_2')^{-1/2},$$

where L_i' and C_k' denote the quantities (3) for $N=9$. In this derivation, we neglect all higher terms with $\mathrm{O}(l_1'^3)$. As known from the analytic expressions of the resonance frequencies, f_0' simply becomes $3f_0$ due to $l_1' = l_1 / 3$, but is considerably reduced by the sum of products $L_i'C_k'$. This sum will reflect physically the interaction among acoustic waves in many smaller cavities that appeared at stage 2. Thus, we see that the ω_0' value reflects the fractal structure at higher stage. Alternatively, as the shortest parts of stage 2 are much smaller than those of stage 1, the fact that $\omega_0' > \omega_0$ might imply that the fractal at higher stage inclines to resonate with the acoustic wave with shorter wavelength. This result will support the claim that the localization of modes contributes to increase the acoustical losses in the fractal cavity.[3]

Third, we would like to comment on the experimental result that the second generation exhibits a broader resonance but a smaller absorption. This broadening is characterized by the quality factor $Q \equiv \omega_0 / 2\Gamma$. From $Q = 1/(\Im T_{11}|_{\omega=\omega_0} + (\rho c / l^2)\Im T_{21}|_{\omega=\omega_0})$, we obtain $Q = 3.8$ for stage 1. On the other hand, for stage 2, we obtain $Q' = 2.1$ from

$$\begin{aligned}\Im T_{11}|_{\omega=\omega_0'} = {} & (6L_1'G_1' + 4L_1'G_2' + 4L_1'G_4' + 2L_1'G_5' + 8L_2'G_1' \\ & + 3L_2'G_2' + 3L_2'G_4' + 2L_2'G_5' + 2L_5'G_1' + L_5'G_2' \\ & + L_5'G_4' + 6C_1'R_1' + 4C_1'R_2' + 4C_1'R_4' + 2C_1'R_5' \\ & + 8C_2'R_1' + 3C_2'R_2' + 3C_2'R_4' + 2C_2'R_5' + 2C_5'R_1' \\ & + C_5'R_2' + C_5'R_4')\omega_0' l_1'^2\end{aligned}$$

and $\Im T_{21}|_{\omega=\omega_0'} = (4C_1' + 4C_2' + C_5')\omega_0' l_1'$. An interesting point here is that the broadening that $Q' < Q$ stems from a rapid increase in the coupling terms as $L_i'G_k'$ and $C_i'R_k'$ describing the effect on the loss at stage 2. From the same reason, we can explain the result that $\alpha_M' < \alpha_M$. Alternatively, this result might be simply explained from the inner structure of fractals as follows: As a cavity in the fractal is physically equivalent to a hollow, the number of hollows in the fractal increases with degree of the fractal stage. Thus, it will be reasonable that the acoustic wave easily passes through the Menger sponge of stage 2 so that $\alpha_M' < \alpha_M$.

Finally, we would like to comment on the origin of peaks in the absorption curves. Intuitively, from Fig. 3, these peaks are expected to relate to the fourth wavelength in the lowest mode of the acoustic wave resonated in the Menger sponge cavity because these peaks are the first ones that appear as the frequencies increase from below. Figure 6 shows the profiles of the sound pressure P calculated from Eq. (2) at $f_0 = 703$ Hz (stage 1) and at $f_0' = 827$ Hz (stage 2). Since $P_1 = P(0)$ is zero and $P_{N+1} = P(0.09)$ becomes maximum, we see that the peaks of α originate from the lowest mode of the acoustic wave.

Our experimental study of the Menger sponge is done only up to stages 1 and 2 so that the results will be insufficient to fully clarify the role of the fractal object for acoustic waves. Since it is hard to study the acoustic property of the Menger sponge with higher stages from the model experiment, it will be necessary to improve the analytical calculation by using a more elaborated model such as the multiple-tube model based on the Webster equation.[6] From the present experimental results, although they are still preliminary, the fractal structure is expected to serve as a sound absorbing material workable for broad band of frequency because of its self-similar scaling property. Here it will be worth pointing out that a multiple noise filtering is imagined as one of possible applications of localized modes found from the investigation of Cantor-like waveguides.[7] Thus, the exploration of the interaction between sound waves and fractal objects must be important for designing a room acoustics, acoustic waveguides, and anechoic chambers.

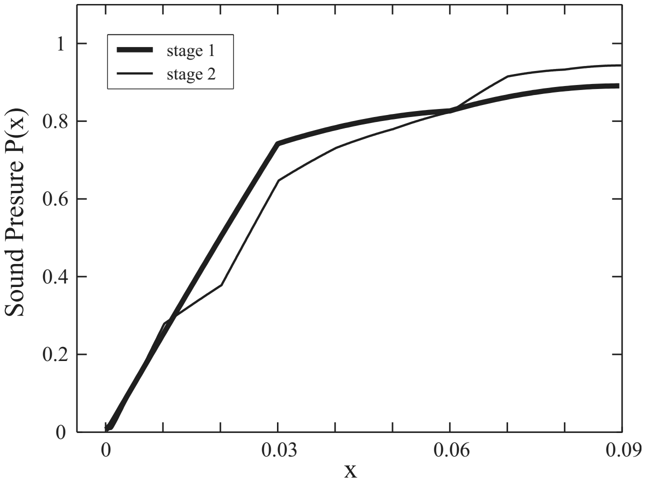


FIG. 6. Sound pressure $P(x)$ at $f_0 = 703$ Hz for stage 1 and at $f_0' = 827$ Hz for stage 2. The distance x is measured from the aperture of the cavity.

We are grateful to Hideo Torihara for his support with the measurement of the sound absorption coefficients by the two-microphone method, to Tokihiko Kaburagi for his advice on the electric equivalent circuit model, and to the members of Department of Acoustic Design of Kyushu University for useful discussions.

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).

[2] V. Gibiat, A. Barjau, K. Castor, and E. B. Chazaud. "Acoustical propagation in a prefractal waveguide," Phys. Rev. E **67**, 066609 (2003).

[3] B. Sapoval, O. Haeberlé, and S. Russ, "Acoustical properties of irregular and fractal cavities," J. Acoust. Soc. Am. **102**, 2014–2019 (1997).

[4] M. W. Takeda, S. Kirihara, Y. Miyamoto, K. Sakoda, and K. Honda, "Localization of electromagnetic waves in three-dimensional fractal cavities," Phys. Rev. Lett. **92**, 093902 (2004).

[5] J. Y. Chung and D. A. Blaser, "Transfer function method of measuring acoustic intensity in a duct system with flow," J. Acoust. Soc. Am. **68**, 1570–1577 (1980).

[6] J. L. Flanagan, *Speech Analysis Synthesis and Perception* (Springer-Verlag, New York, 1965).

[7] E. B. du Chazaud and V. Gibiat, "A numerical study of 1D self-similar waveguides: Relationship between localization, integrated density of state and the distribution of scatterers," J. Sound Vib. **313**, 631–642 (2008); S. Felix, M. Asch, M. Filoche, and B. Sapoval, "Localization and increased damping in irregular acoustic cavities," *ibid.* **299**, 965–976 (2007).

[8] I. G. Main, *Vibrations and Waves in Physics*, 3rd ed. (Cambridge University Press, New York, 1992).

Quantitative evaluation of fracture healing process of long bones using guided ultrasound waves: A computational feasibility study (L)

Xiasheng Guo, Di Yang, and Dong Zhang[a)]
Institute of Acoustics, Key Laboratory of Modern Acoustics, Ministry of Education, Nanjing University, Nanjing 210093, China

Weiguo Li and Yong Qiu
Drum Tower Hospital, Medical School of Nanjing University, Nanjing 210008, China

Junru Wu
Department of Physics, The University of Vermont, Burlington, Vermont 05405

(Received 10 October 2008; revised 25 February 2009; accepted 27 February 2009)

The feasibility of monitoring changes in guided waves' characteristics in a fractured long bone as modeled by a hollow cylinder and a callus at different healing stages is studied. Various guided wave modes are detected and extracted from a broadband signal at several discrete locations. The energy-spectrum and v_{FEP} (effective velocity of the first energy peak in callus region) of guided modes are found sensitive to the healing process in different aspects and stages. The healing process may be divided into several sub-courses, each of which can be evaluated by different combinations of guided wave modes. The energy-spectrum indicates that the longitudinal tube modes L(0,1) and L(0,2) are suitable for early healing; L(0,1), L(0,2), L(0,3), and L(0,5) for midway-course; and L(0,1) and L(0,3) for late consolidation, while L(0,2), L(0,5), and L(0,8) are suitable for detecting the change in callus geometrics. The v_{FEP} results suggest that L(0,5) for monitoring early-course; L(0,3) and L(0,7) for midway process; L(0,2) for later consolidation, and L(0,7) for monitoring geometrical variation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106526]

PACS number(s): 43.80.Ev, 43.80.Qf, 43.80.Jz [CCC] Pages: 2834–2837

I. INTRODUCTION

Bone fracture as a skeletal trauma becomes common in an aged society. The technique of quantitative ultrasound is playing an increasing role in monitoring the status of bone when it heals. This is because (1) ultrasound (US) is a nonionic radiation and relatively safe and (2) the US propagation property can reflect bone mechanical and structural properties.[1,2] As was pointed out by Moilanen,[3] the quantitative guided wave ultrasonography will become a serious competitor for existing methods of bone evaluation. Recently, interests have been focused on using a guided US wave propagating through a long bone to evaluate the mechanical properties during its healing process.[4–6] Compared with other types of mechanical waves such as the longitudinal and surface waves, guided waves can propagate both along the axis and across the thickness of bone, thus may provide a better characterization of bones' material properties and architecture. Typically, a broadband signal is used to excite wave propagation in long bones, and the detected multi-mode signals in time domain are processed with several signal processing techniques, such as the two-dimensional fast Fourier transform (2D-FFT) method,[4] the time-frequency (t-f) method,[5] and spectrum estimation methods.[6]

In previous studies the wave propagation in long bones was usually simplified to two-dimensional (2D) case interpreted with the Lamb wave theory.[7,8] The velocity of the antisymmetrical A0 Lamb mode differs significantly between healthy and osteoporotic bones.[8] However, analysis of *ex vivo* measurements from an intact bone indicated that the Lamb wave theory could not sufficiently describe the dispersion of propagating guided waves in long bones.[9] Moilanen *et al.*[10] demonstrated that the cortical thickness of long bones could be accurately predicted from the velocity of guided waves, and the velocity of the flexural tube mode F(1,1) is highly correlated with the bone cortical thickness. More recently, Protopappas *et al.*[11] presented a preliminary 2D (Ref. 9) and a more reasonable three-dimensional (3D) simulation study on monitoring the fracture healing of long bones using guided waves. Several fracture healing stages of long bones were established, in which the density and elastic properties of the formed callus gradually restore toward those of a healthy bone; energy distributions within different guided wave modes were calculated with the t-f method, and the characteristics of L(0,5) and L(0,8) modes were believed to be able to enhance the quality of monitoring the fracture healing process.[11] However, as pointed out by Protopappas *et al.*, no quantitative results could be provided using the t-f method for monitoring the healing course.

In this letter, we quantitatively study the changes in characteristics of different longitudinal guided wave modes in a bone pipe, L(0,1)– L(0,9), after they propagate through

[a)] Author to whom correspondence should be addressed. Electronic mail: dzhang@nju.edu.cn

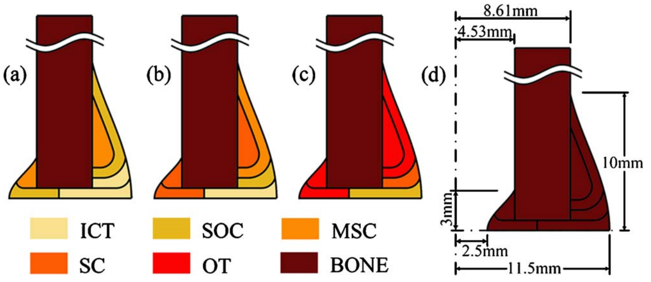


FIG. 1. (Color online) Geometrical and material features of one-quarter of the callus model (Refs. 11 and 12) in (a) stage 1, (b) stage 2, (c) stage 3, and (d) the hypothetical stage 0. (ICT—initial connective tissue, SOC—soft callus, MSC—intermediate stiff callus, SC—stiff callus, and OT—ossified tissue).

the fractured site during several simulated healing stages. Simulations are performed on a finite element (FE) model of healing bone similar to that in the initial studies.[10,11] A mode extraction technique is applied to study the energy variations and the effective velocity in callus region of longitudinal tube modes. The guided waves are excited with a broadband impulse, and vibrations in time domain are detected at a series of discrete positions; the detected signals are then analyzed with the mode extraction technique that gives waveforms of individual modes; energies and effective velocities of the first energy peak (FEP) in the callus region of extracted modes are calculated and discussed for different simulated cases. The objective of the present work is to present an effective method to extract guided wave mode features quantitatively from a broadband guided wave signal propagating in a long bone.

II. METHODS

A FE model similar to the work of Protopappas *et al.*[11] is adopted in this study. The bone is modeled as a 20 cm long hollow circular pipe with inner and outer radii equal to 4.53 and 8.61 mm, respectively. A model of the callus with circular endosteal and periosteal cross sections is located at the center of the pipe. The radius of the callus gradually changes from minimum inner radius (2.5 mm) to maximum outer radius (11.5 mm) along the axial direction according to the Hanning function, as shown in Fig. 1. The modeled callus contains six different regions; in each region the material is regarded as homogeneous isotropic and linearly elastic. The fracture healing is modeled as a three-stage process similar to previous studies.[11,12] Stages 1, 2, and 3 correspond to 1, 4, and 8 weeks after fracture. A hypothetical stage 0 in which the callus consists only of cortical bone, together with an intact bone, is also modeled and analyzed for comparison. Geometrical features of the model and the types of tissues involved in each healing stage are illustrated in Fig. 1.

In each case, since the geometry of the model is axisymmetrical, it is convenient to apply an axisymmetrical modeling method. Thus, we use the axisymmetrical elements of four-node (element type CAX4R, reduced integration with hourglass control) as dominant, and three-node (element type CAX3, linear) in the corner of some segments in ABAQUS, Version 6.6. The size of elements is approximately 0.08 mm in both radial and axial directions, leading to a number of 128 740 elements. At the two ends of the bone pipe, absorption boundaries (element type CINAX4) are applied to avoid reflections. A transmitter that covers a 5 mm annular area on the bone's outer surface is located 25 mm (center-to-center) from the callus, applying an impulse surface traction in radial direction. No receiver is modeled and the detected nodal vibrations are exported from the FE calculations. The excitation signal is a two-cycle 1 MHz sinusoidal tone-burst with its amplitude modulated by a time-dependent Hanning function, leading to a −6 dB bandwidth of 1 MHz. On the other side of the callus, nodal displacements in radial direction are detected on the outer surface of the bone pipe at 176 discrete positions, i.e., at the distance of 3–10 cm from the center of excitation in the axial direction, with a spatial interval of 0.4 mm. For each measured signal, a time step of 0.2 μs, corresponding to a sampling frequency of 5 MHz, is adopted for a 100 μs time period.

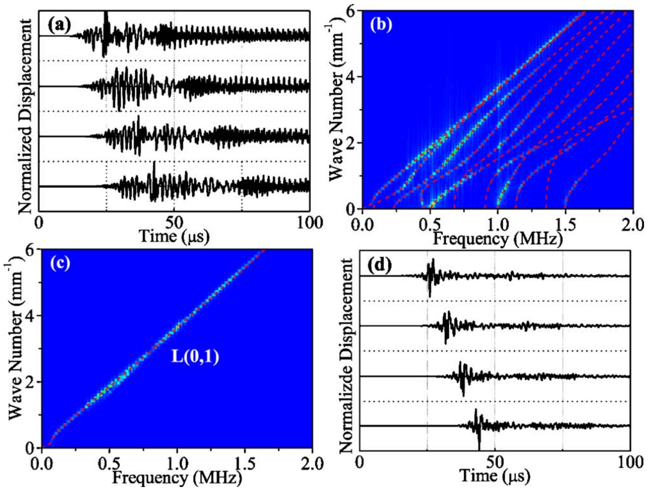


FIG. 2. (Color online) (a) Measured signals at four locations, (b) *k-f* map from 2D-FFT, (c) filtered *k-f* map combined with theoretical *k-f* curve of L(0,1), and (d) extracted L(0,1) signals.

The mode extraction technique[13] is then carried out according to the following procedures. First a 2D-FFT is applied to the recorded signals in both time and space domain. Then a window function is used to filter the obtained 2D *k-f* map to retain the mode of interest and eliminate the undesirable modes. The next step is to apply an inverse 2D-FFT on the filtered *k-f* map to derive the waveforms of the required wave mode at each measured position. The window function is shown in Eq. (1),[13]

$$F(k,f)=\begin{cases}0.5\{1+\cos[2\pi(f-f_m)/f_b]\}, & |f-f_m|<f_b/2\\ 0, & |f-f_m|\geq f_b/2,\end{cases} \qquad (1)$$

where k is the wave number and f is the frequency. f_m represents the theoretical *k-f* dispersion function of the desired mode. f_b is the length of the window in frequency domain. In this study we choose 200 kHz for f_b to ensure that signals for each guided mode, as weak as −30 dB (about 3%) compared to its peak energy, are included in calculations. An example of extracting L(0,1) mode from signals in an intact (healthy) bone model is given in Fig. 2.

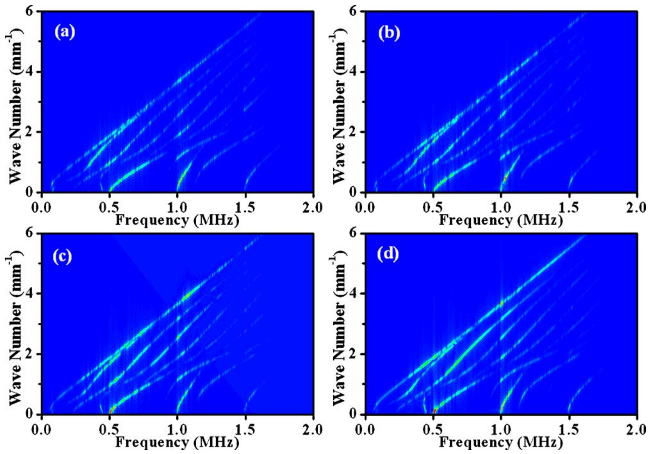


FIG. 3. (Color online) 2D-FFT presentations for (a) stage 1, (b) stage 2, (c) stage 3, and (d) the hypothetical stage 0.

III. RESULTS AND DISCUSSION

It should be noted that the models only describe the evolution of the material properties of the callus at stages from 1 to 3 and the hypothetical stage 0. The change in geometrical characteristics can be described by the comparison of stage 0 and the case of an intact bone. Thus the obtained signals can dynamically reveal the restoration progress of both mechanical and structural properties of the bone during the fracture healing process.

Figure 3 presents the 2D-FFT results for the simulated healing stages 1–3 and hypothetical stage 0. The energies of guided modes gradually recover toward the healthy bone case [Fig. 2(b)]. The longitudinal modes for each healing stage are extracted from Fig. 3, and the calculated normalized energies of nine longitudinal tube modes [L(0,1)–L(0,9), normalized by peak energy)] are compared at a certain position (e.g., the distance between the transmitter and receiver l_{T-R}=5 cm), as shown in Fig. 4. Generally speaking, the L(0,1), L(0,2), L(0,3), L(0,5), and L(0,8) modes possess the largest energy in most of the simulated cases.

To quantify the observed results, the healing process is divided into three sub-courses (i.e., the early-course from stage 1 to 2, the mid-course from stage 2 to 3, and the later-course from stage 3 to the hypothetical stage 0). The energy variations in L(0,1), L(0,2), L(0,3), L(0,5), and L(0,8) modes in each sub-course are presented in Table I corresponding to the data in Fig. 4 (l_{T-R}=5 cm). We find that different combinations of guided longitudinal tube modes could be selected for monitoring different healing sub-courses. The selection of mode combinations is based on the principle that either the modes' energy or their variations should be relatively large. For the early-course, L(0,1) and L(0,2) modes play a dominant role in monitoring, in which the energy increases by 35.0% and 27.2%, respectively. During the mid-course, L(0,1), L(0,2), L(0,3), and L(0,5) modes are significantly affected by the consolidation of the callus, with the energy increases of 101.8%, 59.7%, 102.0%, and 70.0%. For the later-course, the energies of L(0,1) and L(0,3) modes are increased by 16.9% and 104.3%. In terms of evaluating the geometrical change during healing which refers to the change in energy between stage 0 and intact bone, L(0,2), L(0,5), and L(0,8) modes might be significant, with their energy increases of 52.1%, 115.4%, and 528.0%.

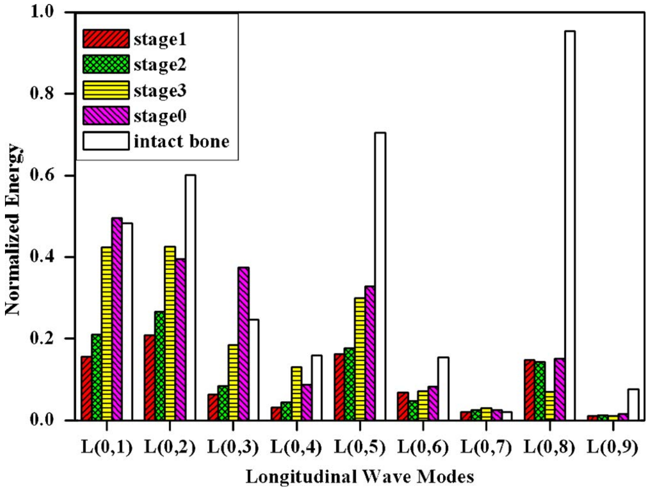


FIG. 4. (Color online) Normalized energy of L(0,1)–L(0,9) throughout healing stages (l_{T-R}=5 cm).

TABLE I. Energy variations in material and geometrics (l_{T-R}=5 cm).

	Energy variations in material (%)			
Modes	Early	Midway	Later	Energy variations in geometrics (%)
L(0,1)	35.0	101.8	16.9	−2.5
L(0,2)	27.2	59.7	−7.0	52.1
L(0,3)	31.9	102.0	104.3	−33.62
L(0,5)	8.6	70.0	9.2	115.4
L(0,8)	−3.5	−51.0	116.7	528.0

Calculating a mode's energy in the t-f domain with a similar methodology of filtering using t-f dispersion curves is widely accepted in analysis of guided wave signals. However, the application of t-f analysis is limited when dealing with long bones since the t-f results may suffer from a readability problem. Since the limited length of the bone (usually less than 50 cm) restricts the l_{T-R}, the dispersive guided signals could not spread out in time domain, which may cause the t-f analysis unable to resolve a single mode from the multi-mode signals.[11,14] By contrast, the mode extraction technique provides the possibility of obtaining signals in time domain for each single mode. Therefore, one can compute the features of each guided mode quantitatively.

Since time of flight (TOF) measurements allow calculation of the velocity of wave packages, we calculate the arrival time of the FEP of each mode in order to study the TOF features from the extracted mode signals. Figure 5 presents the effective velocity of FEP propagating (v_{FEP}) in the callus region for each simulated healing stage with the intact bone. The obtained FEP velocities have shown their sensitivity to the healing progress in a different manner from that of mode energies, i.e., L(0,5) for the early healing (increased by 255 m/s); L(0,3) and L(0,7) for the mid-way course (increased by 468 and 875 m/s, respectively); and L(0,2) for the later con-

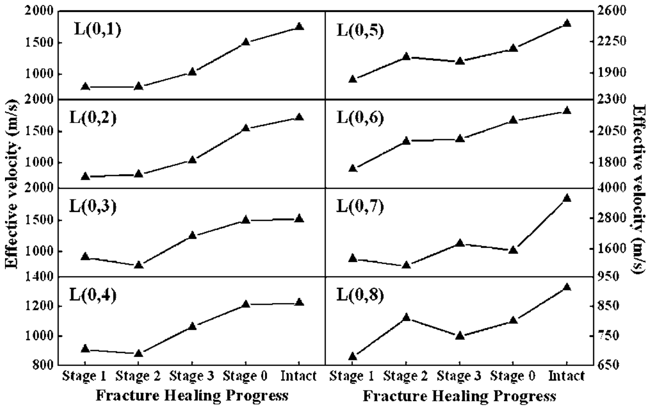


FIG. 5. Variations in v_{FEP} for L(0,1)–L(0,8) during the healing process.

solidation period (increased by 504 m/s), while L(0,7) is the best choice for accounting the geometrical variation (significantly increased by 2044 m/s).

Note that the applied mode extraction technique suffers from leakage noise induced by the FFT algorithm.[13] Our results have shown that the energy calculations are affected by less than 4%; furthermore, as the noise level is relatively low, it almost has no effect on the results of the FEP arrival time. Therefore, the influence of FFT leakage is limited.

The model used here is limited by its axisymmetrical geometry, the annular excitation, and simplified process of bone calcification. The axisymmetrical modeling method results in axisymmetrical excitations, which eliminates the torsional and flexural wave modes. In fact, long bones are irregular pipes where an infinite number of flexural modes may also propagate. Fortunately, dispersion characteristics of irregular geometries could be obtained with numerical simulation using FE or boundary element methods. Furthermore, some flexural wave modes are very close to longitudinal ones, e.g., F(1,8) vs L(0,5) and F(1,13) vs L(0,8).[13] Thus data measured from the real experiments could also be quantified using this mode extraction technique.

The simulated annular transducer used for excitation is far from being realistic; however, it could be carried out by several general contact transducers placed in a circumferential manner, which could restrain the propagation of most of the torsional and flexural wave modes. Also, it would be practically useful to simulate an annular transducer by multiple individuals in FE simulations when applying a full 3D analysis algorithm (not axisymmetrical). Additionally, fracture callus formation is a complex process; calcification in the callus represents a random and anisotropic pattern. The cortical surface at the fracture sites often shows complex topology which may affect wave propagation. These factors are needed to be addressed in future studies.

IV. CONCLUSION

A numerical feasibility study for quantitatively evaluating the healing process of long bones by the use of guided US waves is performed. Longitudinal tube modes are extracted from guided signals simulated at several different healing stages. The energies and v_{FEP} of extracted modes are calculated to test their capability of assessing the healing status of long bones. The energy results indicate that the L(0,1) and L(0,2) modes are appropriate to evaluate the early healing and the L(0,1), L(0,2), L(0,3), and L(0,5) modes are suitable to monitor the midway-course of the healing process, while the energy of the L(0,1) and L(0,3) modes changed significantly in the late consolidation course. It is also found that the L(0,2), L(0,5), and L(0,8) modes are sensitive to the change in the callus geometrics. Additionally, the v_{FEP} results recommend L(0,5) for monitoring the early-course; L(0,3) and L(0,7) for the mid-way process; L(0,2) for the later consolidation period; and L(0,7) accounts for the geometrical variation. It is suggested that the healing process may be divided into several sub-processes, each of which can be evaluated by using different combinations of guided wave modes.

ACKNOWLEDGMENTS

This research was supported by the Program for New Century Excellent Talents in University (Contract No. 06-0450) and the Scientific Research Foundation of Graduate School of Nanjing University (Contract No. 2006CL07).

[1] P. Laugier, "Quantitative ultrasound of bone: Looking ahead," Jt., Bone Spine **73**, 125–128 (2006).

[2] C. F. Njeh, J. R. Kearton, D. Hans, and C. M. Boivin, "The use of quantitative ultrasound to monitor fracture healing: A feasibility study using phantoms," Med. Eng. Phys. **20**, 781–786 (1999).

[3] P. Moilanen, "Ultrasonic guided waves in bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1277–1286 (2008).

[4] D. Alleyne and P. Cawley, "A two-dimensional Fourier transform method for the assessment of propagating multimode signals," J. Acoust. Soc. Am. **89**, 1159–1168 (1991).

[5] W. H. Prosser and M. D. Seale, "Time-frequency analysis of the dispersion of Lamb waves," J. Acoust. Soc. Am. **105**, 2669–2676 (1999).

[6] J. Vollmann and J. Dual, "High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium," J. Acoust. Soc. Am. **102**, 896–920 (1997).

[7] P. Moilanen, P. H. F. Nicholson, V. Kilappa, S. Cheng, and J. Timonen, "Measuring guided waves in long bones: Modeling and experiments in free and immersed plates," Ultrasound Med. Biol. **32**, 709–719 (2006).

[8] P. H. F. Nicholson, P. Moilanen, T. Kärkkäinen, J. Timonen, and S. Cheng, "Guided ultrasonic waves in long bones: Modeling, experiment and *in vivo* application," Physiol. Meas. **23**, 755–768 (2002).

[9] V. C. Protopappas, D. I. Fotiadis, and K. N. Malizos, "Guided ultrasound wave propagation in intact and healing long bones," Ultrasound Med. Biol. **32**, 693–708 (2006).

[10] P. Moilanen, P. H. F. Nicholson, V. Kilappa, S. Cheng, and J. Timonen, "Assessment of the cortical bone thickness using ultrasonic guided waves: Modeling and *in vitro* study," Ultrasound Med. Biol. **33**, 254–262 (2007).

[11] V. C. Protopappas, I. C. Kourtis, K. N. Malizos, C. V. Massalas, and D. I. Fotiadis, "Three-dimensional finite element modeling of guided ultrasound wave propagation in intact and healing long bones," J. Acoust. Soc. Am. **121**, 3907–3921 (2007).

[12] L. E. Claes and C. A. Heigele, "Magnitudes of local stress and strain along bony surfaces predict the course and type of fracture healing," J. Biomech. **32**, 255–266 (1999).

[13] T. Hayashi and K. Kawashima, "Single mode extraction from multiple modes of Lamb wave and its application to defect detecting," JSME Int. J., Ser. A **46**(4), 620–626 (2003).

[14] X. S. Guo, D. Zhang, D. Yang, X. F. Gong, and J. R. Wu, "Comment on "Three-dimensional finite element modeling of guided ultrasound wave propagation in intact and healing long bones,"," J. Acoust. Soc. Am. **123**, 4047–4050 (2008).

A simple method avoiding non-uniqueness in the boundary element method for acoustic scattering problem

Kunikazu Hirosawa[a)]
Nittobo Acoustic Engineering Co., Ltd., 1–21–10, Midori, Sumida-ku, Tokyo 130–0021, Japan

Takashi Ishizuka
Institute of Technology, Shimizu Corporation, 4–17, Etchujima 3–chome, Koto-ku, Tokyo 135–8530, Japan

Kyoji Fujiwara
Faculty of Design, Kyushu University, Shiobaru 4–9–1, Minami, Fukuoka 815–8540, Japan

(Received 19 February 2008; revised 12 March 2009; accepted 12 March 2009)

The boundary element method (BEM) is widely used for sound field analysis problems; however, it has a non-uniqueness problem in the exterior domain. Various methods to avoid this problem have been developed; however, these are not easily applied to the BEM. In this paper, a simple method called the “ICA-Ring (inner cavity ringing) method” is proposed for avoiding the non-uniqueness problem, and this method is applied to the BEM in both single and plural domains. The concept of the ICA-Ring method is that a scatterer in free space is hollowed as a shell and the volume is smaller; the eigenfrequencies are shifted to a higher range. Next, the mechanism of the non-uniqueness problem in plural domains and a reason of the application of the ICA-Ring method to the case of plural domains are explained. Finally, some results calculated by the BEM using the ICA-Ring method are shown. The calculational condition is that a cylinder with radius 0.125 m floats in two-dimensional free space. In this case, no calculational errors exist in 1–6000 Hz in both single and plural domains, when the thickness of the shell is 20 mm. The ICA-Ring method does not need to modify an existing computer program of conventional BEM.
© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3111856]

PACS number(s): 43.20.Rz, 43.20.Fn [KA] Pages: 2838–2846

I. INTRODUCTION

The boundary element method (BEM) is widely used in sound field analysis problems. One of the most significant advantages of the BEM compared to the finite element method is that the problem’s dimensionality is reduced by 1, so that only the boundary of each domain has to be discretized. The BEM is also suitable for infinite domain problems, because the far-field radiation condition is always satisfied. In an exterior problem, however, it is known that BEM has non-unique solutions at characteristic frequencies which are the eigenfrequencies of the interior Dirichlet problems. This is a purely mathematical issue; it has no physical meaning for the exterior problems under investigation.

To avoid the non-uniqueness problem, two conventional techniques have been applied to acoustic scattering problems. One is the combined Helmholtz integral equation formulation (CHIEF),[1] in which the Helmholtz integral equation with source points in the interior domain of the scattering body is applied in the exterior domain. By using the Helmholtz integral equation of the interior domain, CHIEF creates an overdetermined system of simultaneous equations, and these equations can be solved using a least-squares technique. CHIEF has been widely used for acoustic scattering and radiation problems, and various improvements to the original method have been proposed.[2–7] In CHIEF, however, the researcher has to search for interior points (called the CHIEF points). Generally, only trial and error determines the number and positions of the CHIEF points. The other conventional technique applied to acoustic scattering problems to avoid the non-uniqueness problem in the BEM method was proposed by Burton and Miller.[8] This method uses a complex linear combination of the Helmholtz integral equation and its normal derivative equation for only the exterior domain. It has been proven that the linear combination of these two equations will yield a unique solution when the coefficient is a complex number. As with CHIEF, the Burton–Miller method has been used and improved.[9–16] However, these two methods do not work in every situation. Marburg and Amini discussed these methods in detail.[17]

As an alternative to using conventional methods, Ishizuka and Fujiwara proposed a simple technique to avoid the non-uniqueness problem of the BEM.[18] In this method, the scatterer is modeled as a shell to reduce the bounded area while maintaining the geometric configuration. Sakuma *et al.* also proposed a technique similar to the one proposed by Ishizuka and Fujiwara.[19]

The above discussion limits single domain; the non-uniqueness problem also arises in plural domains. Cremers *et al.* investigated non-uniqueness problem in multi-domain for the direct collocation BEM,[20] and they showed that the problem in multi-domain can be avoided by using the infinite

[a)] Author to whom correspondence should be addressed. Electronic mail: hirosawa@noe.co.jp

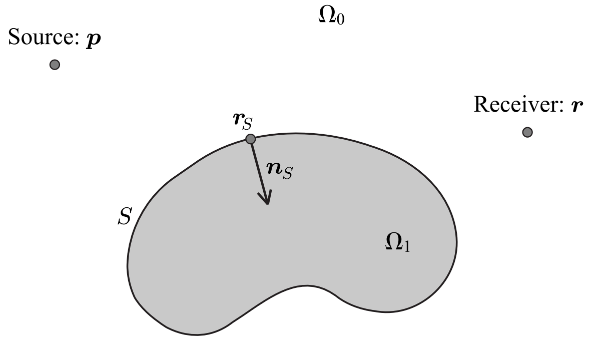


FIG. 1. Geometry of the acoustic scattering problem.

element. In their paper, however, the only case was that all subdomains are in external domain and the internal domain of the scatterer is not considered.

For both single domain and plural domains, the special techniques must be applied to the BEM for avoiding the non-uniqueness problem. This seems to be difficult to non-BEM specialists such as general engineers, and means that those involved in the BEM have to modify the computer program of conventional BEM.

In this paper, we propose a simple method of improving the technique proposed by Ishizuka and Fujiwara to avoid the non-uniqueness problem in the BEM. Our proposed method is called the "ICA-Ring (inner cavity ringing) method" because, in this method, a scattering body is regarded as a shell. First, we explain the application of the ICA-Ring method to the most basic single domain problem. Then, we explain the application of the ICA-Ring method for the avoidance of the non-uniqueness problem in plural domains, for example, in the case in which a scatterer consists of a porous material with air space (this is treated as a coupled problem for exterior and interior domains). Finally, we include several calculation results of two-dimensional acoustic scattering problems around a cylinder. The merits of the ICA-Ring method are that it is a simpler logic than those of CHIEF and Burton and Miller method, and it does not need to modify the existing computational program of conventional BEM.

II. FORMULATION OF THE ICA-RING METHOD

A. The basic integral equation and the non-uniqueness problem

Consider the situation shown in Fig. 1, which shows the sound source $\boldsymbol{p}$, the receiver $\boldsymbol{r}$, and the finite body with smooth surface S in the exterior domain Ω_0. $\boldsymbol{r}_S$ is the source point anywhere on S, and the interior domain of the finite body is Ω_1. In the exterior domain Ω_0, the well-known Helmholtz–Huygens integral equation is given by

$$c(\boldsymbol{r})p(\boldsymbol{r}) = G(\boldsymbol{p},\boldsymbol{r}) - \int_S \left\{ p(\boldsymbol{r}_S)\frac{\partial G(\boldsymbol{r}_S,\boldsymbol{r})}{\partial n_S} - G(\boldsymbol{r}_S,\boldsymbol{r})\frac{\partial p(\boldsymbol{r}_S)}{\partial n_S} \right\} dS(\boldsymbol{r}_S), \qquad (1)$$

where $p(\boldsymbol{r})$ is the sound pressure at the arbitrary point $\boldsymbol{r}$ in Ω_0; on the smooth surface S, $\boldsymbol{n}_S$ at $\boldsymbol{r}_S$ is the unit vector normal to the surface S into the interior domain Ω_1. $c(\boldsymbol{r})$ is equal to 1 when $\boldsymbol{r}$ exists in Ω_0, $1/2$ when $\mathbf{r}$ exists on S, and 0 when $\boldsymbol{r}$ exists in Ω_1. G is the free space Green's function, which is given by the following in two and three dimensions, respectively:

$$G(\boldsymbol{r}_0,\boldsymbol{r}) = \frac{1}{4j}H_0^{(2)}(kR), \quad \text{(for 2D)},$$

$$G(\boldsymbol{r}_0,\boldsymbol{r}) = \frac{e^{-jkR}}{4\pi R}, \quad \text{(for 3D)}, \qquad (2)$$

where $\boldsymbol{r}_0$ is the source point, R is the distance between $\boldsymbol{r}_0$ and $\boldsymbol{r}$, ω is the angular frequency, c_0 is the sound speed in Ω_0, $k=\omega/c_0$ is the wave number, and j is the imaginary unit. Also, in this paper, $\exp(j\omega t)$ is used as the time factor, where t is time. Since the sound pressure on $\boldsymbol{r}_S$ is $p(\boldsymbol{r}_S)$, the particle velocity $\boldsymbol{u}(\boldsymbol{r}_S)$ is described as follows:

$$\boldsymbol{u}(\boldsymbol{r}_S) = -\frac{1}{j\omega\rho_0}\frac{\partial p(\boldsymbol{r}_S)}{\partial n_S}, \qquad (3)$$

where ρ_0 is the density of the medium in Ω_0. Assuming that S is "locally reactive,"

$$\frac{\partial p(\boldsymbol{r}_S)}{\partial n_S} = -jk\beta(\boldsymbol{r}_S)p(\boldsymbol{r}_S), \qquad (4)$$

where $\beta(\boldsymbol{r}_S)$ is the normal acoustic admittance on the point $\boldsymbol{r}_S$, which is normalized by the characteristic impedance $\rho_0 c_0$ of the medium in Ω_0. Hence, Eq. (1) is rewritten as follows:

$$c(\boldsymbol{r})p(\boldsymbol{r}) = G(\boldsymbol{p},\boldsymbol{r}) - \int_S p(\boldsymbol{r}_S)\left\{ \frac{\partial G(\boldsymbol{r}_S,\boldsymbol{r})}{\partial n_S} + jk\beta(\boldsymbol{r}_S)G(\boldsymbol{r}_S,\boldsymbol{r}) \right\} dS(\boldsymbol{r}_S). \qquad (5)$$

The Helmholtz–Huygens integral equation in Eq. (1) is discretized into N_0 boundary elements for estimation of values on S. This discretization yields a simultaneous equation with N_0 unknowns, which is expressed in matrix form as

$$[\boldsymbol{A}^{(0)}]\{\boldsymbol{p}^{(0)}\} = \{\boldsymbol{p}_d\}, \qquad (6)$$

where $[\boldsymbol{A}^{(0)}]$ is the coefficient matrix $(N_0 \times N_0)$ in Ω_0 and is given by

$$[\boldsymbol{A}^{(0)}] = \begin{bmatrix} a_1^{(0)}(\boldsymbol{r}_1) & \cdots & a_{N_0}^{(0)}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ a_1^{(0)}(\boldsymbol{r}_{N_0}) & \cdots & a_{N_0}^{(0)}(\boldsymbol{r}_{N_0}) \end{bmatrix}. \qquad (7)$$

If a surface is discretized into constant boundary elements, $a_j^{(0)}(\boldsymbol{r}_i)$ is given by

$$a_j^{(0)}(\boldsymbol{r}_i) = \begin{cases} c(\boldsymbol{r}_i) + h_{S_j}(\boldsymbol{r}_i) + g_{S_j}(\boldsymbol{r}_i), & i=j \\ h_{S_j}(\boldsymbol{r}_i) + g_{S_j}(\boldsymbol{r}_i), & i \neq j \end{cases}, \qquad (8)$$

$$h_{S_j}(\boldsymbol{r}_i) = \int_{S_j} \frac{\partial G(\boldsymbol{r}_{S_j},\boldsymbol{r}_i)}{\partial n_{S_j}} dS(\boldsymbol{r}_{S_j}),$$

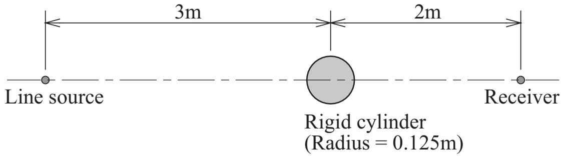


FIG. 2. Arrangement of an acoustic scattering problem including a floating rigid cylinder in two-dimensional free space.

$$g_{S_j}(\boldsymbol{r}_i) = jk\beta(\boldsymbol{r}_{S_j})\int_{S_j} G(\boldsymbol{r}_{S_j},\boldsymbol{r}_i)dS(\boldsymbol{r}_{S_j}), \tag{9}$$

and $c(\boldsymbol{r}_i)=1/2$. Note that in Eqs. (8) and (9), the index $j(1 \leq j \leq N_0)$ represents the element numbers on a surface and the index $i(1 \leq i \leq N_0)$ represents the source points at node points on each element, respectively. $\{\boldsymbol{p}^{(0)}\}$ is the vector which consists of N_0 unknown sound pressures at the node points on the boundary elements for S, and $\{\boldsymbol{p}_d\}$ is the direct sound vector that indicates the contribution of radiation from a sound source to the N_0 boundary elements. Solving Eq. (6), the sound pressure vector $\{\boldsymbol{p}^{(0)}\}$ on S is evaluated, and then values at arbitrary receivers can be estimated by substitution of $\{\boldsymbol{p}^{(0)}\}$ into the Helmholtz–Huygens integral equation, Eq. (5).

In order to observe the non-uniqueness problem, we compute the acoustic scattering around a cylinder in two dimensions, and compare the boundary element solution with the exact solution.[21] The arrangement for this numerical analysis is shown in Fig. 2, which shows a line source, a receiver, and a floating rigid cylinder in two-dimensional free space. The radius of the cylinder is 0.125 m. In the application of BEM, we presume the constant boundary element for discretization of the surface of the cylinder. In order to observe the behavior of the results closely, the calculations are conducted at 1 Hz steps over the indicated frequency range. The results are shown in Fig. 3, which show that the conventional BEM solution seriously decreases accuracy around the eigenfrequencies. Several eigenfrequencies are described in Table I. This calculation error is caused by the reason that the rank of the coefficient matrix in Eq. (7) is insufficient to solve the simultaneous equation (6) in the neighborhood of the eigenfrequencies of Ω_1.

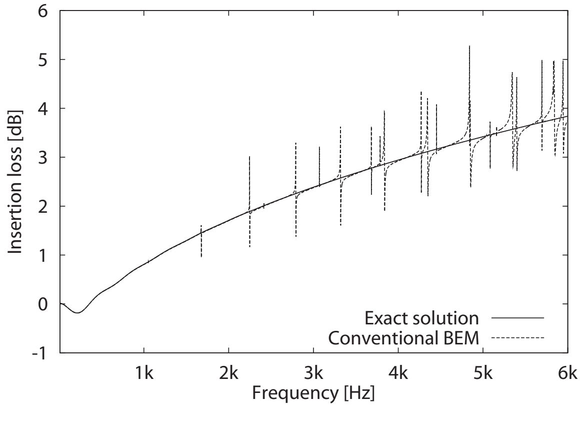


FIG. 3. Comparison of the IL computed according to the exact solution and by conventional BEM for a floating rigid cylinder in two-dimensional free space.

TABLE I. Several eigenfrequencies for a rigid cylinder whose radius is 0.125 m.

(i,j)	Freq. (Hz)
(0,1)	1042.5
(0,2)	2389.7
(0,3)	3746.2
(1,1)	1658.8
(1,2)	3037.1
(1,3)	4404.1
(2,1)	2223.2
(2,2)	3643.8
(2,3)	2762.0

B. The ICA-Ring method for single domains

To avoid the non-uniqueness problem described in Sec. II A, Ishizuka and Fujiwara proposed the following improvement technique:[18] a tiny hole, which is small enough so that it will not affect the sound field around the scatterer, is made in the original surface, and the original scatterer is hollowed out, as shown in Fig. 4. The interior surface is assumed to be absorptive, to suppress re-radiation from the opening. This improvement method is based on the idea that the eigenfrequencies of the non-uniqueness problem depend on the volume or area of a scatterer, and are shifted to higher frequencies when the volume or area of the scatterer becomes smaller. Using this technique, Ishizuka and Fujiwara succeeded almost completely in avoiding the inaccuracies caused by the non-uniqueness problem in the frequency range of which they considered.

In this paper, we aim to perfectly remove the influence of the tiny opening in the scatterer. As shown in Fig. 4, we suppose that the exterior surface containing the tiny opening is S', the interior surface is S'_A, and the tiny opening's surface is S'_R and rigid. Δ and δ are the width of the tiny opening and the thickness of the modified scatterer, respectively. Then, Eq. (5) is rewritten as follows:

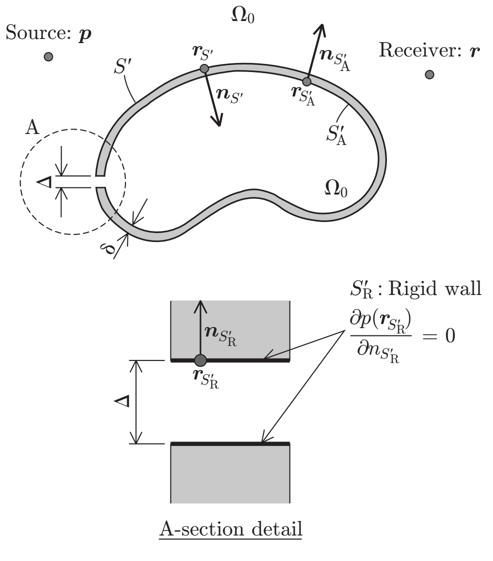


FIG. 4. Illustration of the method proposed by Ishizuka and Fujiwara to avoid the non-uniqueness problem in BEM.

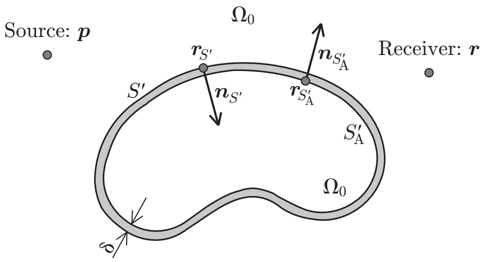


FIG. 5. Illustration of the ICA-Ring method given by Eq. (12) for a single domain.

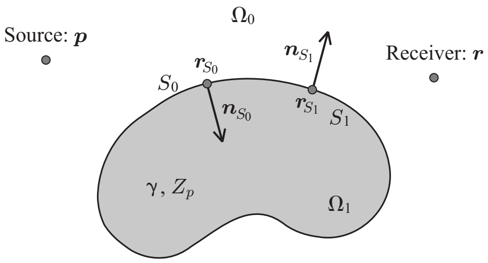


FIG. 6. Geometry of the acoustic scattering problem for plural domains.

$$c(\boldsymbol{r})p(\boldsymbol{r}) = G(\boldsymbol{p},\boldsymbol{r}) - \int_{S'+S'_A+S'_R} p(\boldsymbol{r}_S)\left\{\frac{\partial G(\boldsymbol{r}_S,\boldsymbol{r})}{\partial n_S} + jk\beta(\boldsymbol{r}_S)G(\boldsymbol{r}_S,\boldsymbol{r})\right\}dS(\boldsymbol{r}_S) = G(\boldsymbol{p},\boldsymbol{r}) - \int_{S'} p(\boldsymbol{r}_{S'}) \times \left\{\frac{\partial G(\boldsymbol{r}_{S'},\boldsymbol{r})}{\partial n_{S'}} + jk\beta(\boldsymbol{r}_{S'})G(\boldsymbol{r}_{S'},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S'}) - \int_{S'_A} p(\boldsymbol{r}_{S'_A})\left\{\frac{\partial G(\boldsymbol{r}_{S'_A},\boldsymbol{r})}{\partial n_{S'_A}} + jkG(\boldsymbol{r}_{S'_A},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S'_A}) - \int_{S'_R} p(\boldsymbol{r}_{S'_R})\frac{\partial G(\boldsymbol{r}_{S'_R},\boldsymbol{r})}{\partial n_{S'_R}}dS(\boldsymbol{r}_{S'_R}). \tag{10}$$

We note that the interior is absorptive and that the tiny opening's surface is a rigid wall. In this paper, to minimize the reflection from S'_A, $\beta(\boldsymbol{r}_{S'_A})$ is 1, and $\beta(\boldsymbol{r}_{S'_R})$ is 0. Moreover, when we assume that the tiny opening's surfaces are parallel with each other, the unit normal vectors on the surface S'_R point are in opposite directions. If the tiny opening is made small without limit, $p(\boldsymbol{r}_{S'_R})$ on each of the surfaces must approach equal values. Then, the third integral term in Eq. (10) becomes 0:

$$\lim_{\Delta\to 0}\int_{S'_R} p(\boldsymbol{r}_{S'_R})\frac{\partial G(\boldsymbol{r}_{S'_R},\boldsymbol{r})}{\partial n_{S'_R}}dS(\boldsymbol{r}_{S'_R}) = 0. \tag{11}$$

Now, Eq. (10) can be rewritten as follows:

$$c(\boldsymbol{r})p(\boldsymbol{r}) = G(\boldsymbol{p},\boldsymbol{r}) - \int_{S'} p(\boldsymbol{r}_{S'})\left\{\frac{\partial G(\boldsymbol{r}_{S'},\boldsymbol{r})}{\partial n_{S'}} + jk\beta(\boldsymbol{r}_{S'})G(\boldsymbol{r}_{S'},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S'}) - \int_{S'_A} p(\boldsymbol{r}_{S'_A}) \times \left\{\frac{\partial G(\boldsymbol{r}_{S'_A},\boldsymbol{r})}{\partial n_{S'_A}} + jkG(\boldsymbol{r}_{S'_A},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S'_A}). \tag{12}$$

Equation (12) means that a scatterer S as a single domain in Ω_0, as shown by Fig. 1, can be dealt with as a shell with thickness δ, as shown in Fig. 5. Although both the exterior and the interior of a shell are the same field Ω_0, S is separated into two boundaries, S' and S'_A. By the discretization of S' and S'_A into N_0 and M_0 boundary elements, respectively, the integral equation (12) is expressed in matrix form as

$$[\widetilde{\boldsymbol{A}}^{(0)}]\{\widetilde{\boldsymbol{p}}^{(0)}\} = \{\widetilde{\boldsymbol{p}}_d\}, \tag{13}$$

where $[\widetilde{\boldsymbol{A}}^{(0)}]$ is the coefficient matrix $((N_0+M_0)\times(N_0+M_0))$, $\{\widetilde{\boldsymbol{p}}^{(0)}\}$ is the vector which consists of N_0+M_0 unknown sound pressures on S' and S'_A, and $\{\widetilde{\boldsymbol{p}}_d\}$ is the direct sound vector that indicates the contribution of radiation from a sound source to the N_0+M_0 boundary elements. We have used the tilde symbol $(\widetilde{X})$ to indicate the application of the ICA-Ring method.

By applying BEM to Eq. (12) and solving the simultaneous equations (13), we can avoid the non-uniqueness problem in our required frequency range. The ICA-Ring method is very simple compared with the CHIEF or the Burton-Miller method, because the ICA-Ring method requires only the consideration of a new boundary S'_A. The interior boundary S'_A depends on the thickness of the shell or required frequency range, the decision does not need trial and error many times.

C. The ICA-Ring method for plural domains

In Fig. 1, when the interior domain Ω_1 is a material like glass wool, which is able to transmit sound waves, calculation of the acoustic field is generally treated as the coupled problem in which Ω_1 is combined with Ω_0 by continuity conditions of sound pressure and particle velocity on the boundary surfaces S_0 and S_1.

In this case, we presume that Ω_1 is an isotropic medium with the propagation constant γ and characteristic impedance Z_p, as shown in Fig. 6. Since the boundary conditions on S_0 in Ω_0 and S_1 in Ω_1 cannot be determined only by normal acoustic admittances $\beta(\boldsymbol{r}_{S_0})$ and $\beta(\boldsymbol{r}_{S_1})$, we also need to introduce the particle velocities $u(\boldsymbol{r}_{S_0})$ on S_0 and $u(\boldsymbol{r}_{S_1})$ on S_1 to give the boundary conditions. The Helmholtz–Huygens integral equations in Ω_0 and Ω_1 are written as follows, from Eq. (1):

$$c(\boldsymbol{r})p(\boldsymbol{r}) = G(\boldsymbol{p},\boldsymbol{r}) - \int_{S_0}\left\{p(\boldsymbol{r}_{S_0})\frac{\partial G(\boldsymbol{r}_{S_0},\boldsymbol{r})}{\partial n_{S_0}} + jk\rho_0 c_0 u(\boldsymbol{r}_{S_0})G(\boldsymbol{r}_{S_0},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S_0}) \quad \text{in } \Omega_0, \tag{14}$$

$$c(\boldsymbol{r})p(\boldsymbol{r}) = -\int_{S_1}\left\{p(\boldsymbol{r}_{S_1})\frac{\partial G(\boldsymbol{r}_{S_1},\boldsymbol{r})}{\partial n_{S_1}} + jk_pZ_pu(\boldsymbol{r}_{S_1})G(\boldsymbol{r}_{S_1},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S_1}) \quad \text{in } \Omega_1, \tag{15}$$

where the following relations are used in Eqs. (14) and (15):

$$\frac{\partial p(\boldsymbol{r}_{S_0})}{\partial n_{S_0}} = -\rho_0\frac{\partial u(\boldsymbol{r}_{S_0})}{\partial t} = -jk\rho_0c_0u(\boldsymbol{r}_{S_0}),$$

$$\frac{\partial p(\boldsymbol{r}_{S_1})}{\partial n_{S_1}} = -\rho_p\frac{\partial u(\boldsymbol{r}_{S_1})}{\partial t} = -jk_pZ_pu(\boldsymbol{r}_{S_1}), \tag{16}$$

where ρ_p is the density of the material in the interior domain Ω_1; k_p is the complex wave number which described by the propagation constant γ of Ω_1 as

$$k_p = -j\gamma \Rightarrow \gamma = \zeta + j\eta \quad (\zeta > 0, \eta > 0). \tag{17}$$

ζ is the attenuation constant, and η is the acoustic phase constant of the material in Ω_1, respectively.

We can solve Eqs. (14) and (15) for the sound pressure and the particle velocity on the boundary surfaces S_0 and S_1 with their continuity conditions, which are that at points $\boldsymbol{r}_{S_0}$ on S_0 and $\boldsymbol{r}_{S_1}$ on S_1, the sound pressures are equal, and the particle velocities are equal but directly opposite to each other, that are

$$p(\boldsymbol{r}_{S_0}) = p(\boldsymbol{r}_{S_1}),$$

$$u(\boldsymbol{r}_{S_0}) = -u(\boldsymbol{r}_{S_1}). \tag{18}$$

However, Eqs. (14) and (15) are combined only by the continuity conditions of sound pressure and particle velocity; they are mutually independent mathematically. If we leave Eq. (14) in Ω_0 alone, then the non-uniqueness problem is caused in Ω_0. It is mentioned in Sec. III that the non-uniqueness problem occurs in spite of considering both domains Ω_0 and Ω_1 and applying a coupled method. Therefore, we have to avoid non-uniqueness as described in Sec. II B for Eq. (14) in Ω_0.

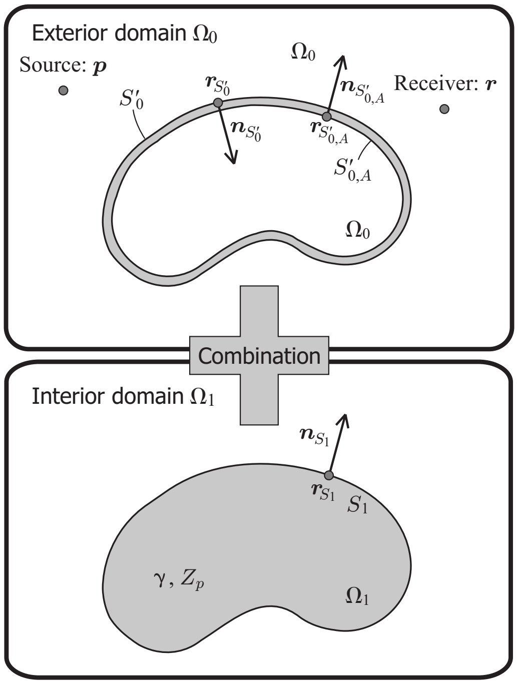


FIG. 7. The analysis model for BEM as a combination of the exterior domain Ω_0 and the interior domain Ω_1.

By following Sec. II B, the body in Ω_0 is treated as the shell shown on the upper side of Fig. 7. When the boundary S_0 in Ω_0 is separated into $S'_0+S'_{0,A}$, we assume that the outside boundary condition of S'_0 is given by the particle velocity $u(\boldsymbol{r}_{S'_0})$, and that the inside surface, $S'_{0,A}$, is absorptive [in this paper, the normal acoustic admittance $\beta(\boldsymbol{r}_{S'_{0,A}}) = 1$]. Then, Eq. (14) is rewritten as

$$c(\boldsymbol{r})p(\boldsymbol{r}) = G(\boldsymbol{p},\boldsymbol{r}) - \int_{S'_0}\left\{p(\boldsymbol{r}_{S'_0})\frac{\partial G(\boldsymbol{r}_{S'_0},\boldsymbol{r})}{\partial n_{S'_0}} + jk\rho_0c_0u(\boldsymbol{r}_{S'_0})G(\boldsymbol{r}_{S'_0},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S'_0}) - \int_{S'_{0,A}}\left\{p(\boldsymbol{r}_{S'_{0,A}})\frac{\partial G(\boldsymbol{r}_{S'_{0,A}},\boldsymbol{r})}{\partial n_{S'_{0,A}}} + jkp(\boldsymbol{r}_{S'_{0,A}})G(\boldsymbol{r}_{S'_{0,A}},\boldsymbol{r})\right\}dS(\boldsymbol{r}_{S'_{0,A}}). \tag{19}$$

S'_0, $S'_{0,A}$, and S_1 are discretized into N_0, M_0, and N_1 boundary elements, respectively. Provided that N_0=N_1, the integral equations (19) and (15) can be combined and expressed in matrix forms as follows:

$$[\widetilde{\boldsymbol{H}}^{(0)}_{S'}]\{\widetilde{\boldsymbol{p}}^{(0)}_{S'}\} + [\widetilde{\boldsymbol{G}}^{(0)}_{S'}]\{\widetilde{\boldsymbol{u}}^{(0)}_{S'}\} + [\widetilde{\boldsymbol{A}}^{(0)}_{S'_A}]\{\widetilde{\boldsymbol{p}}^{(0)}_{S'_A}\} = \{\widetilde{\boldsymbol{p}}_d\}$$

$$[\boldsymbol{H}^{(1)}]\{\boldsymbol{p}^{(1)}\} + [\boldsymbol{G}^{(1)}]\{\boldsymbol{u}^{(1)}\} = \{\boldsymbol{0}\} \tag{20}$$

Applying the continuity conditions (18) on $\boldsymbol{r}_{S_0}$=$\boldsymbol{r}_{S_1}$,

$$[\widetilde{\boldsymbol{H}}^{(0)}_{S'}]\{\widetilde{\boldsymbol{p}}^{(0)}_{S'}\} + [\widetilde{\boldsymbol{G}}^{(0)}_{S'}]\{\widetilde{\boldsymbol{u}}^{(0)}_{S'}\} + [\widetilde{\boldsymbol{A}}^{(0)}_{S'_A}]\{\widetilde{\boldsymbol{p}}^{(0)}_{S'_A}\} = \{\widetilde{\boldsymbol{p}}_d\}$$

$$[\boldsymbol{H}^{(1)}]\{\widetilde{\boldsymbol{p}}^{(0)}_{S'}\} - [\boldsymbol{G}^{(1)}]\{\widetilde{\boldsymbol{u}}^{(0)}_{S'}\} = \{\boldsymbol{0}\}$$

$$\Leftrightarrow \begin{bmatrix} [\widetilde{\boldsymbol{H}}^{(0)}_{S'}] & [\widetilde{\boldsymbol{A}}^{(0)}_{S'_A}] & [\widetilde{\boldsymbol{G}}^{(0)}_{S'}] \\ [\boldsymbol{H}^{(1)}] & [\boldsymbol{0}] & -[\boldsymbol{G}^{(1)}] \end{bmatrix} \begin{bmatrix} \{\widetilde{\boldsymbol{p}}^{(0)}_{S'}\} \\ \{\widetilde{\boldsymbol{p}}^{(0)}_{S'_A}\} \\ \{\widetilde{\boldsymbol{u}}^{(0)}_{S'}\} \end{bmatrix} = \begin{bmatrix} \{\widetilde{\boldsymbol{p}}_d\} \\ \{\boldsymbol{0}\} \end{bmatrix}. \tag{21}$$

The matrix on the left hand side of Eq. (21) is the global coefficient matrix $((N_0+M_0+N_1)\times(N_0+M_0+N_1) = (2N_0+M_0)\times(2N_0+M_0))$ in Ω_0 and Ω_1. The component matrices $[\widetilde{\boldsymbol{H}}^{(0)}_{S'}]$, $[\boldsymbol{H}^{(1)}]$, $[\widetilde{\boldsymbol{G}}^{(0)}_{S'}]$, $[\boldsymbol{G}^{(1)}]$, $[\widetilde{\boldsymbol{A}}^{(0)}_{S'_A}]$ in the global coefficient matrix will be described in Appendix A. $[\{\widetilde{\boldsymbol{p}}^{(0)}_{S'}\},\{\widetilde{\boldsymbol{p}}^{(0)}_{S'_A}\},\{\widetilde{\boldsymbol{u}}^{(0)}_{S'}\}]^T$ is the total unknown vector: $\{\widetilde{\boldsymbol{p}}^{(0)}_{S'}\}$ consists of N_0 unknown sound pressures on S'_0, $\{\widetilde{\boldsymbol{p}}^{(0)}_{S'_A}\}$ consists of M_0 unknown sound pressures on $S'_{0,A}$, and $\{\widetilde{\boldsymbol{u}}^{(0)}_{S'}\}$, consists of N_0 unknown particle velocities on S'_0. $\{\widetilde{\boldsymbol{p}}_d\}$ is the direct sound vector that indicates the contribution of radiation from a sound source to the N_0+M_0 boundary elements on S'_0 and $S'_{0,A}$.

The sound pressures $p(\boldsymbol{r}_{S'_0})$, $p(\boldsymbol{r}_{S'_{0,A}})$, and $p(\boldsymbol{r}_{S_1})$ and the particle velocities $u(\boldsymbol{r}_{S'_0})$ and $u(\boldsymbol{r}_{S_1})$ are estimated by solving the simultaneous equations in Eq. (21). When we apply this method, we get valid values that avoid the non-uniqueness problem in our required frequency range.

In this paper, we describe a plural domain problem in which the domain Ω_1 is a porous material. However, this formulation can be applied to various cases, for example, the case in which Ω_1 and Ω_0 are different materials, and only a partial surface connects Ω_0 with Ω_1.

III. NUMERICAL CALCULATION EXAMPLES AND DISCUSSION

In this section, we show some results calculated by the BEM in which we have applied the ICA-Ring method. These calculations are performed under conditions identical to those described in Sect. II A. A cylinder with radius 0.125 m floats in two-dimensional free space. There is a line source 3 m away from the center of the cylinder and a receiver 2 m away from the center of the cylinder on the opposite side from the line source. We calculate results over the frequency range 10–6000 Hz in 1 Hz steps. We calculate the insertion loss (IL) of the cylinder to evaluate the effectiveness of the ICA-Ring method.

A. Single domain

In the case of a rigid cylinder, an exact solution exists, so we can compare our BEM solutions with this exact solution.[21] These results are shown in Fig. 8.

The top figure in Fig. 8 shows the conventional BEM solution in which ICA-Ring has not been applied. This figure shows the inaccuracies of the conventional BEM solutions that appear around the eigenfrequencies of the interior Dirichlet problems. In a cylinder, these eigenfrequencies correspond with the wave numbers $k_{i,j}$ which are the solutions of the equation $J_i(k_{i,j}r)=0$, where r is a radius of a cylinder and J_i is Bessel function of the i-th order. Table I shows several eigenfrequencies of the (i,j) orders that are (0,1)–(2,3), when the radius of the cylinder is 0.125 m.

The second to fourth figures in Fig. 8 show results calculated by BEM using the ICA-Ring method. Three types of the shell thickness used in these calculations are 40, 30, and 20 mm, respectively. These figures show that the eigenfrequencies are shifted to a higher range by a smaller cross-sectional area according to the thickness of the shell, and therefore the total cross-sectional area of the shell decreases. When the thickness of shell is 20 mm, the BEM solution calculated using ICA-Ring method is almost equal to the exact solution in our required frequency range.

B. Double domains

This section considers a cylinder that consists of an isotropic and permeable material, for example, a porous material such as glass wool. Since sound waves are also propagated inside such a cylinder, we will apply the coupled method to BEM, as described in Sec. II C. We suppose that the flow resistivity of the material inside the cylinder is 9000 N s/m^4, and that the propagation constant and the characteristic impedance of the material are estimated by the Miki model.[22] Figure 9 shows the results of the ICA-Ring method for this case and those of a rigid cylinder.

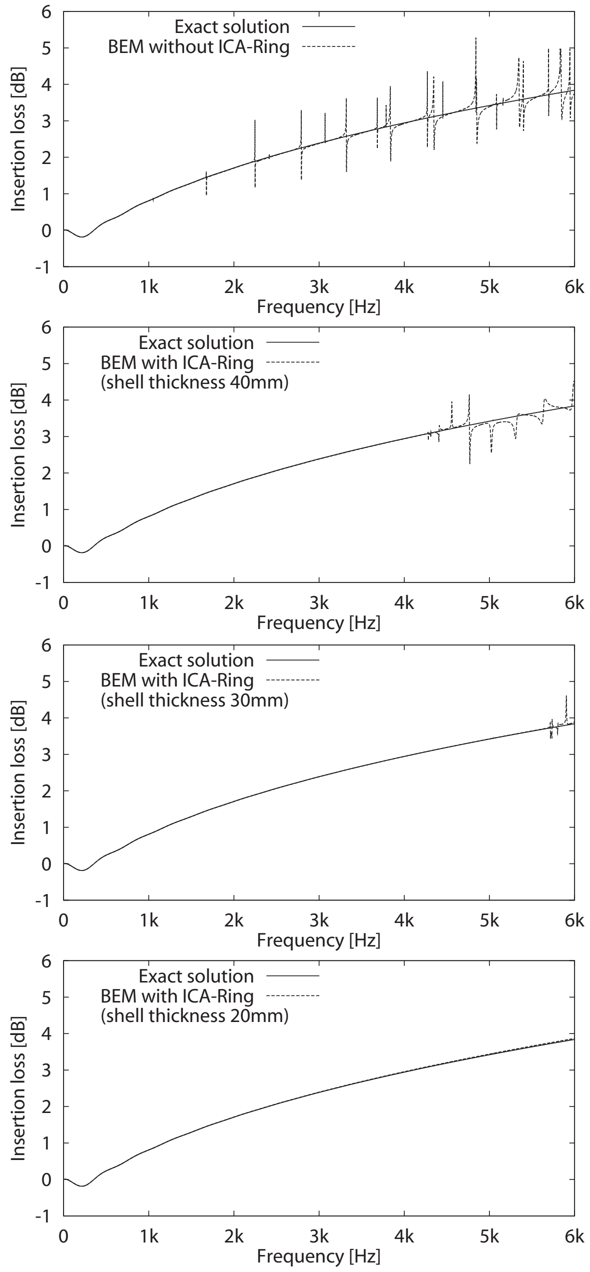


FIG. 8. The IL of a floating rigid cylinder in two-dimensional free space.

The top figure of Fig. 9 shows conventional BEM solutions without the application of the ICA-Ring method. The second through fourth figures show the BEM solutions using the ICA-Ring method. It can be noted, as in the case of single domains, the accuracy of the conventional BEM solution is seriously decreased by the non-uniqueness of the interior Dirichlet problem in the neighborhood of the eigenfrequencies of a permeable cylinder, and by applying ICA-Ring

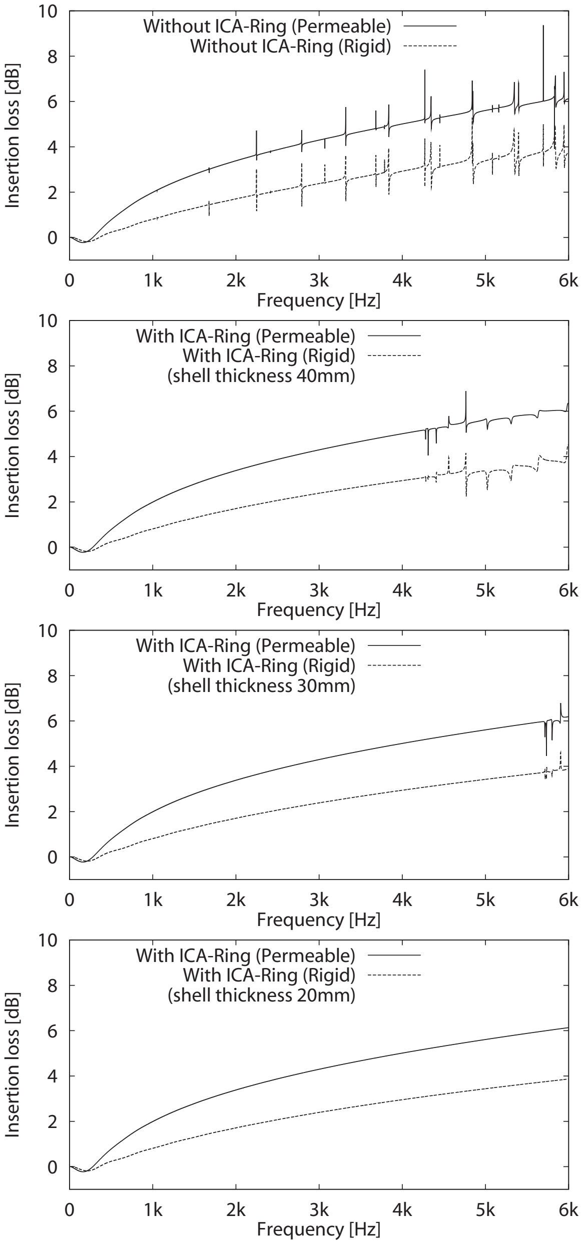


FIG. 9. The IL of floating permeable and rigid cylinders in two-dimensional free space. The permeable cylinder consists of an isotropic material that is able to transmit sound waves, with a flow resistivity of 9000 N s/m^4.

method, the eigenfrequencies effectively shift to a higher range by making the cross-sectional area smaller.

It is significant that the eigenfrequencies of a single domain and of plural domains are equal. In the case considered here, the eigenfrequencies of the permeable cylinder include those in Table I, because our BEM formulation of plural domains treats each domain as mathematically independent.

Now, consider the case in which the ICA-Ring method is not applied to the BEM formulation. In Fig. 6, we assume that surfaces S_0 and S_1 are discretized into N_0 and N_1 boundary elements, respectively, and that $N_0 = N_1$. Then, matrix expressions of the discretized Helmholtz–Huygens integral equations in Ω_0 and Ω_1 with continuity conditions (18) on $\boldsymbol{r}_{S_0} = \boldsymbol{r}_{S_1}$ are described as

$$\begin{bmatrix} [\boldsymbol{H}^{(0)}] & [\boldsymbol{G}^{(0)}] \\ [\boldsymbol{H}^{(1)}] & -[\boldsymbol{G}^{(1)}] \end{bmatrix} \begin{bmatrix} \{\boldsymbol{p}^{(0)}\} \\ \{\boldsymbol{u}^{(0)}\} \end{bmatrix} = \begin{bmatrix} \{\boldsymbol{p}_d\} \\ \{\boldsymbol{0}\} \end{bmatrix}. \qquad (22)$$

The matrix on the left hand side of this equation is the global coefficient matrix $((N_0+N_1)\times(N_0+N_1)=2N_0\times 2N_0)$ in Ω_0 and Ω_1. The component matrices $[\boldsymbol{H}^{(0)}]$, $[\boldsymbol{H}^{(1)}]$, $[\boldsymbol{G}^{(0)}]$, and $[\boldsymbol{G}^{(1)}]$ in the global coefficient matrix are described in Appendix B. $[\{\boldsymbol{p}^{(0)}\},\{\boldsymbol{u}^{(0)}\}]$ is the total unknown vector: $\{\boldsymbol{p}^{(0)}\}$ consists of N_0 unknown sound pressures on S_0, and $\{\boldsymbol{u}^{(0)}\}$ consists of N_1 unknown particle velocities on S_1. $\{\boldsymbol{p}_d\}$ is the direct sound vector that indicates the contribution of radiation from a sound source to the N_0 boundary elements on S_0. The component matrices $[\boldsymbol{H}^{(0)}]$ and $[\boldsymbol{G}^{(0)}]$ for Ω_0 in the global coefficient matrix in Eq. (22) are virtually equal to $[\boldsymbol{A}^{(0)}]$ in Eq. (6). This is due to the fact that taking into account the relationship between Eqs. (3) and (4), Eqs. (B2) and (B6) are equal to Eq. (9), and the simultaneous equation regarding Ω_0 in Eq. (22) means the same as in Eq. (6). Since the rank of the component matrices $[\boldsymbol{H}^{(0)}]$ and $[\boldsymbol{G}^{(0)}]$ are insufficient to solve the simultaneous equation in Eq. (6) around the eigenfrequencies, the rank of the global coefficient matrix is also insufficient to solve the simultaneous equation in Eq. (22). As noted above, the eigenfrequencies of the plural domains are equal to those of the single domain. Therefore, in the exterior domain Ω_0, the formulations of single and plural domains are treated equivalently, and the non-uniqueness problem must be avoided for both cases.

IV. CONCLUSIONS

In this paper, we have discussed the inaccuracies of BEM calculations in the neighborhood of the eigenfrequencies of the interior Dirichlet problems for both single and plural domains, and proposed a new simple method, called ICA-Ring, which corrects these inaccuracies by avoiding the non-uniqueness problem in the BEM formulation.

We have applied the ICA-Ring method to BEM for both single and plural domain problems. To compare our method to conventional BEM, we computed the IL for rigid and permeable cylinders in two-dimensional free space. In the conventional BEM solution, the equation cannot be solved properly around the eigenfrequencies of the interior Dirichlet problems. This non-uniqueness problem is present in both the single and plural domain cases: because the exterior domain in the plural domains is treated as single domain, consequently the eigenfrequencies are equal to those of the single domain.

The ICA-Ring method effectively shifts the problematic eigenfrequencies to outside the frequency range of researchers' interest. This is accomplished by reducing the area or the volume of a scatterer by modeling it as a shell with a modified boundary. The thickness of this shell determines the upper frequency of the analysis range: a thinner shell increases the frequency range that can be accurately calculated without interference from the non-uniqueness problem. The ICA-Ring method is much simpler than conventional techniques, such as CHIEF or the Burton–Miller method, which are usedto avoid the non-uniqueness problem, because the conventional BEM computer program is not needed to modify and the addition of a boundary inside the modeled scatterer is only required. However, adding a boundary increases the computation cost for ICA-Ring method and it is larger than conventional techniques, CHIEF or the Burton–Miller method. In the future, to further improve the ICA-Ring method in view of the reduction in the computational cost, we will investigate the optimum size and condition of boundary elements and determine whether computation time could be reduced without loss of accuracy.

APPENDIX A: COMPONENT MATRICES OF THE GLOBAL COEFFICIENT MATRIX IN EQUATION (21)

Assuming that the surfaces S_0', $S_{0,A}'$, and S_1 are discretized into constant boundary elements, the component matrices of the global coefficient matrix given in Eq. (21) are described as follows:

$$[\tilde{\boldsymbol{H}}^{(0)}_{S'}] = \begin{bmatrix} h_{S'_{0,1}}(\boldsymbol{r}_1) & \cdots & h_{S'_{0,N_0}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ h_{S'_{0,1}}(\boldsymbol{r}_{N_0+M_0}) & \cdots & h_{S'_{0,N_0}}(\boldsymbol{r}_{N_0+M_0}) \end{bmatrix}, \tag{A1}$$

$$h_{S'_{0,j}}(\boldsymbol{r}_i) = \begin{cases} \dfrac{1}{2} + \displaystyle\int_{S'_{0,j}} \dfrac{\partial G(\boldsymbol{r}_{S'_{0,j}},\boldsymbol{r}_i)}{\partial n_{S'_{0,j}}} dS(\boldsymbol{r}_{S'_{0,j}}) & (i=j) \\ \displaystyle\int_{S'_{0,j}} \dfrac{\partial G(\boldsymbol{r}_{S'_{0,j}},\boldsymbol{r}_i)}{\partial n_{S'_{0,j}}} dS(\boldsymbol{r}_{S'_{0,j}}) & (i \neq j) \end{cases}, \tag{A2}$$

$$[\boldsymbol{H}^{(1)}] = \begin{bmatrix} h_{S_{1,1}}(\boldsymbol{r}_1) & \cdots & h_{S_{1,N_1}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ h_{S_{1,1}}(\boldsymbol{r}_{N_1}) & \cdots & h_{S_{1,N_1}}(\boldsymbol{r}_{N_1}) \end{bmatrix}, \tag{A3}$$

$$h_{S_{1,j}}(\boldsymbol{r}_i) = \begin{cases} \dfrac{1}{2} + \displaystyle\int_{S_{1,j}} \dfrac{\partial G(\boldsymbol{r}_{S_{1,j}},\boldsymbol{r}_i)}{\partial n_{S_{1,j}}} dS(\boldsymbol{r}_{S_{1,j}}) & (i=j) \\ \displaystyle\int_{S_{1,j}} \dfrac{\partial G(\boldsymbol{r}_{S_{1,j}},\boldsymbol{r}_i)}{\partial n_{S_{1,j}}} dS(\boldsymbol{r}_{S_{1,j}}) & (i \neq j) \end{cases}, \tag{A4}$$

$$[\tilde{\boldsymbol{G}}^{(0)}_{S'}] = \begin{bmatrix} g_{S'_{0,1}}(\boldsymbol{r}_1) & \cdots & g_{S'_{0,N_0}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ g_{S'_{0,1}}(\boldsymbol{r}_{N_0+M_0}) & \cdots & g_{S'_{0,N_0}}(\boldsymbol{r}_{N_0+M_0}) \end{bmatrix}, \tag{A5}$$

$$g_{S'_{0,1}}(\boldsymbol{r}_i) = jk\rho_0 c_0 \int_{S'_{0,1}} G(\boldsymbol{r}_{S'_{0,1}},\boldsymbol{r}_i) dS(\boldsymbol{r}_{S'_{0,1}}), \tag{A6}$$

$$[\boldsymbol{G}^{(1)}] = \begin{bmatrix} g_{S_{1,1}}(\boldsymbol{r}_1) & \cdots & g_{S_{1,N_1}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ g_{S_{1,1}}(\boldsymbol{r}_{N_1}) & \cdots & g_{S_{1,N_1}}(\boldsymbol{r}_{N_1}) \end{bmatrix}, \tag{A7}$$

$$g_{S_{1,j}}(\boldsymbol{r}_i) = jk_p Z_p \int_{S_{1,j}} G(\boldsymbol{r}_{S_{1,j}},\boldsymbol{r}_i) dS(\boldsymbol{r}_{S_{1,j}}) \tag{A8}$$

$$[\tilde{\boldsymbol{A}}^{(0)}_{S'_A}] = \begin{bmatrix} a_{S'_{0,A,N_0}}(\boldsymbol{r}_1) & \cdots & a_{S'_{0,A,N_0+M_0}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ a_{S'_{0,A,N_0+1}}(\boldsymbol{r}_{N_0+M_0}) & \cdots & a_{S'_{0,A,N_0+M_0}}(\boldsymbol{r}_{N_0+M_0}) \end{bmatrix}, \tag{A9}$$

$$a_{S'_{0,A,j}}(\boldsymbol{r}_i) = \begin{cases} \dfrac{1}{2} + \displaystyle\int_{S'_{0,A,j}} \left\{ \dfrac{\partial G(\boldsymbol{r}_{S'_{0,A,j}},\boldsymbol{r}_i)}{\partial n_{S'_{0,A,j}}} + jkG(\boldsymbol{r}_{S'_{0,A,j}},\boldsymbol{r}_i) \right\} dS(\boldsymbol{r}_{S'_{0,A,j}}) & (i=j) \\ \displaystyle\int_{S'_{0,A,j}} \left\{ \dfrac{\partial G(\boldsymbol{r}_{S'_{0,A,j}},\boldsymbol{r}_i)}{\partial n_{S'_{0,A,j}}} + jkG(\boldsymbol{r}_{S'_{0,A,j}},\boldsymbol{r}_i) \right\} dS(\boldsymbol{r}_{S'_{0,A,j}}) & (i \neq j) \end{cases}, \tag{A10}$$

APPENDIX B: COMPONENT MATRICES OF THE GLOBAL COEFFICIENT MATRIX IN EQUATION (22)

Assuming that the surfaces S_0 and S_1 are discretized into constant boundary elements, the component matrices of the global coefficient matrix given in Eq. (22) are described as follows:

$$[\boldsymbol{H}^{(0)}] = \begin{bmatrix} h_{S_{0,1}}(\boldsymbol{r}_1) & \cdots & h_{S_{0,N_0}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ h_{S_{0,1}}(\boldsymbol{r}_{N_0}) & \cdots & h_{S_{0,N_0}}(\boldsymbol{r}_{N_0}) \end{bmatrix}, \tag{B1}$$

$$h_{S_{0,j}}(\boldsymbol{r}_i) = \begin{cases} \dfrac{1}{2} + \displaystyle\int_{S_{0,j}} \dfrac{\partial G(\boldsymbol{r}_{S_{0,j}}, \boldsymbol{r}_i)}{\partial n_{S_{0,j}}} dS(\boldsymbol{r}_{S_{0,j}}) & (i = j) \\ \displaystyle\int_{S_{0,j}} \dfrac{\partial G(\boldsymbol{r}_{S_{0,j}}, \boldsymbol{r}_i)}{\partial n_{S_{0,j}}} dS(\boldsymbol{r}_{S_{0,j}}) & (i \neq j) \end{cases}, \tag{B2}$$

$$[\boldsymbol{H}^{(1)}] = \begin{bmatrix} h_{S_{1,1}}(\boldsymbol{r}_1) & \cdots & h_{S_{1,N_1}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ h_{S_{1,1}}(\boldsymbol{r}_{N_1}) & \cdots & h_{S_{1,N_1}}(\boldsymbol{r}_{N_1}) \end{bmatrix}, \tag{B3}$$

$$h_{S_{1,j}}(\boldsymbol{r}_i) = \begin{cases} \dfrac{1}{2} + \displaystyle\int_{S_{1,j}} \dfrac{\partial G(\boldsymbol{r}_{S_{1,j}}, \boldsymbol{r}_i)}{\partial n_{S_{1,j}}} dS(\boldsymbol{r}_{S_{1,j}}) & (i = j) \\ \displaystyle\int_{S_{1,j}} \dfrac{\partial G(\boldsymbol{r}_{S_{1,j}}, \boldsymbol{r}_i)}{\partial n_{S_{1,j}}} dS(\boldsymbol{r}_{S_{1,j}}) & (i \neq j) \end{cases}, \tag{B4}$$

$$[\boldsymbol{G}^{(0)}] = \begin{bmatrix} g_{S_{0,1}}(\boldsymbol{r}_1) & \cdots & g_{S_{0,N_0}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ g_{S_{0,1}}(\boldsymbol{r}_{N_0}) & \cdots & g_{S_{0,N_0}}(\boldsymbol{r}_{N_0}) \end{bmatrix}, \tag{B5}$$

$$g_{S_{0,j}}(\boldsymbol{r}_i) = jk\rho_0 c_0 \int_{S_{0,j}} G(\boldsymbol{r}_{S_{0,j}}, \boldsymbol{r}_i) dS(\boldsymbol{r}_{S_{0,j}}), \tag{B6}$$

$$[\boldsymbol{G}^{(1)}] = \begin{bmatrix} g_{S_{1,1}}(\boldsymbol{r}_1) & \cdots & g_{S_{1,N_1}}(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ g_{S_{1,1}}(\boldsymbol{r}_{N_1}) & \cdots & g_{S_{1,N_1}}(\boldsymbol{r}_{N_1}) \end{bmatrix}, \tag{B7}$$

$$g_{S_{1,j}}(\boldsymbol{r}_i) = jk_p Z_p \int_{S_{1,j}} G(\boldsymbol{r}_{S_{1,j}}, \boldsymbol{r}_i) dS(\boldsymbol{r}_{S_{1,j}}). \tag{B8}$$

[1] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).

[2] A. F. Seybert and T. K. Rengarajan, "The use of chief to obtain unique solutions for acoustic radiation using boundary integral equations," J. Acoust. Soc. Am. **81**, 1299–1306 (1987).

[3] T. W. Wu and A. F. Seybert, "A weighted residual formulation for the chief method in acoustics," J. Acoust. Soc. Am. **90**, 1608–1614 (1991).

[4] D. J. Segalman and D. W. Lobitz, "A method to overcome computational difficulties in the exterior acoustics problem," J. Acoust. Soc. Am. **91**, 1855–1861 (1992).

[5] R. A. Marschall, "Boundary element solution of a body's exterior acoustic field near its internal eigenvalues," J. Comput. Acoust. **1**, 335–353 (1993).

[6] P. Juel, "A numerical study of the coefficient matrix of the boundary element method near characteristic frequencies," J. Sound Vib. **175**, 39–50 (1994).

[7] Z. S. Chen, G. Hofstetter, and H. A. Mang, "A symmetric galerkin formulation of the boundary element method for acoustic radiation and scattering," J. Comput. Acoust. **5**, 219–241 (1997).

[8] A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).

[9] W. L. Meyer, W. A. Bell, B. T. Zinn, and M. P. Stallybrass, "Boundary integral solutions of three dimensional acoustic radiation problems," J. Sound Vib. **59**, 245–262 (1978).

[10] T. Terai, "On calculation of sound fields around three dimensional objects by integral equation methods," J. Sound Vib. **69**, 71–100 (1980).

[11] Z. Reut, "On the boundary integral method for the exterior acoustic problem," J. Sound Vib. **103**, 297–298 (1985).

[12] K. A. Cunefare and G. Koopmann, "A boundary element method for acoustic radiation valid for all wavenumbers," J. Acoust. Soc. Am. **85**, 39–48 (1988).

[13] C. C. Chien, H. Rajiyah, and S. N. Atluri, "An effective method for solving the hypersingular integral equations in 3-d acoustics," J. Acoust. Soc. Am. **88**, 918–937 (1990).

[14] S. A. Yang, "A boundary integral equation method for two-dimensional acoustic scattering problems," J. Acoust. Soc. Am. **105**, 93–105 (1999).

[15] Z. Y. Yan, K. C. Hung, and H. Zheng, "Solving the hypersingular boundary integral equation in three-dimensional acoustics using a regularization relationship," J. Acoust. Soc. Am. **113**, 2674–2683 (2003).

[16] S. A. Yang, "An integral equation approach to three-dimensional acoustic radiation and scattering problems," J. Acoust. Soc. Am. **116**, 1372–1380 (2004).

[17] S. Marburg and S. Amini, "Cat's eye radiation with boundary elements: Comparative study on treatment of irregular frequencies," J. Comput. Acoust. **13**, 21–45 (2005).

[18] T. Ishizuka and K. Fujiwara, "Performance of noise barriers with various edge shapes and acoustical conditions," Appl. Acoust. **65**, 125–141 (2004).

[19] T. Sakuma, Y. Kosaka, Y. Yasuda, and T. Oshima, "Application of thin layer boundary modelling in boundary element sound field analysis," in Technical Committee Meeting on Architectural Acoustics (sponsored by Acoustical Society of Japan) (2006), in Japanese.

[20] L. Cremers, K. R. Fyfe, and P. Sas, "A variable order infinite element for multi-domain boundary element modelling of acoustic rafiation and scattering," Appl. Acoust. **59**, 185–220 (2000).

[21] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes*, revised printing (Hemisphere, Philadelphia, 1987).

[22] Y. Miki, "Acoustical properties of porous materials—modifications of Delany–Bazley models," J. Acoust. Soc. Jpn. **11**, 19–24 (1990).

Low-frequency geoacoustic model for the effective properties of sandy seabottoms

Ji-Xun Zhou and Xue-Zhen Zhang
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405 and National Laboratory of Acoustics, Institute of Acoustics, The Chinese Academy of Sciences, Beijing 100190, China

D. P. Knobles
The Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713

(Received 7 May 2008; revised 2 February 2009; accepted 4 February 2009)

The debate on the sound speed dispersion and the frequency dependence of sound attenuation in seabottoms has persisted for decades, mainly due to the lack of sufficient experimental data in the low-frequency (LF) to high-frequency speed/attenuation transition band. This paper analyzes and summarizes a set of LF measurements in shallow water that have resulted in the identification of nonlinear frequency dependence of sound attenuation in the effective media of sandy seabottoms. The long-range acoustic measurements were conducted at 20 locations in different coastal zones around the world. The seabed attenuations, inverted from different acoustic field measurements and characteristics, exhibit similar magnitude and nonlinear frequency dependence below 1000 Hz. The resulting effective sound attenuation can be expressed by $\alpha(\mathrm{dB/m})=(0.37 \pm 0.01)(f/1000)^{(1.80\pm0.02)}$ for 50–1000 Hz. The corresponding average sound speed ratio at the bottom-water interface in the 50–600 Hz range is 1.061 ± 0.009. Both the LF-field-derived sound speed and attenuation can be well described by the Biot–Stoll model with parameters that are consistent with either theoretical considerations or experimental measurements. A combination of the LF-field-inverted data with the SAX99, SAX04, and other high-frequency measurements offers a reference broadband data set in the 50–400 000 Hz range for sonar prediction and sediment acoustics modeling. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3089218]

PACS number(s): 43.30.Ma, 43.30.Pc, 43.30.Zk, 43.20.Jr [RCG] Pages: 2847–2866

I. INTRODUCTION

The emphasis in ocean acoustic research has markedly shifted from deep sea to shallow water (SW) in recent years. Strong seabed and sea surface interactions, multipath (multimode), and water column variability characterize sound propagation and signal fluctuations in SW. An accurate solution to the Helmholtz equation in SW waveguide requires accurate seabed acoustic parameters to define a bottom boundary condition. Reliable predictions of transmission loss (TL), echo to reverberation ratio, spatial/temporal decorrelation, time/angle spreading of signals, etc., all need a seabed geoacoustic model that incorporates the physics of acoustic interaction with the bottom.

In the past three decades a significant effort has been devoted to development of seabed geoacoustic models and to *in situ* measurements of sediment properties. There are several well known seabed geoacoustic models that have resulted from this effort: the Hamilton visco-elastic model,[1–4] the Biot–Stoll poro-elastic model,[5–11] the Buckingham VGS (viscous-grain-shearing) model,[12–14] and the Chotiros–Isakson BICSQS (Biot–Stoll with squirt flow and shear) model.[15]

Hamilton and co-worker[1–4] published a series of review papers on his work over many years, classified the sediments in continental zones into nine types, and developed successful geoacoustic models for practical purposes. They derived empirical relationships to connect sound speed and attenuation to sediment type (mean grain size or porosity) and frequency. Their empirical models suggest that sound speed is approximately independent of frequency and the attenuation of sound in marine sediments increases linearly with frequency over the full frequency range (a few hertz to megahertz) of interest in ocean acoustics. Attenuation values computed with the model of Hamilton and co-worker[1–4] are in reasonable agreement with experimental values at high frequencies (HFs) (>10 kHz) or for finer-grained (soft) sediments with a high porosity. Their JASA papers,[3,4] are still highly regarded today, indicating the importance of their contribution.

On the basis of the porous nature of marine sediments, Biot[5–7] and Stoll and co-worker[8–11] theorized that the relative motion of the pore fluid and the sediment frame should lead to viscous damping of the sound wave. The Biot–Stoll model predicts that the acoustic attenuation in sediments should exhibit a nonlinear frequency dependence, particularly in sandy and sand-silt mixture bottoms (we refer to both of these as sandy seabottoms in the interest of brevity). It also predicts that the sound speed in these sediments should exhibit strong nonlinear dispersion. Because of a lack of enough convincing experimental evidence to support the Biot–Stoll model, the debate on the frequency dependence of sound attenuation in the seabottom, especially in sandy bottoms at low frequencies (LFs), has persisted for decades. The small sound attenuation in sediments at LFs prevents meaningful laboratory measurements, because the required dimensions associated with an experiment are at least hundreds or thousands of meters to achieve a detectable amount of sound attenuation variation. Thus, few quality LF seabed attenuation data are available for testing the validity of the visco-elastic model[1–4] or the poro-elastic model.[5–11]

Considering grain-to-grain contacts in marine sediments, Buckingham[12,13] developed a grain-shearing (GS) model to correlate the extensive HF data sets, published by Hamilton[1–4] and others, and the physical properties of the sediments. The GS theory has recently been extended to a new version (designated as the VGS theory) to include the effects of the viscosity of the molecularly thin layer of pore fluid separating contiguous grains.[14] At lower frequency, the sound speed dispersion curve and the frequency dependence of sound attenuation in marine sediments predicted by the VGS theory are similar to those of the Biot model. At HF, the VGS speed dispersion curve approaches those of the GS theory asymptotically; the VGS attenuation frequency dependence at high frequency is close to f^1 but the Biot dependence is close to $f^{1/2}$.

Apparent in the extant published experimental data in water-saturated sands, the sound speed dispersion is significantly greater than that predicted by the Biot–Stoll model with constant coefficients. In order to explain this, Chotiros and Isakson[15] proposed the BICSQS model, an extension of the Biot–Stoll model. The extension uses a physical model of the grain-to-grain contact, which includes squirt flow and shear drag, to compute the frame moduli.

In recent years, a number of careful studies have offered some data to support the Biot–Stoll model and demonstrated the inadequacy of the visco-elastic model.[16–19] Reference 18 cites a complete lack of experimental data for p-wave attenuation at frequencies below 1 kHz, and Ref. 19 notes that the largest difference between model and data occurs at the lowest frequencies. Isakson and Neilsen[20] indicated that inferences on sediment properties from reflection measurements at HFs are by themselves insufficient to predict broadband behavior. However, seabed inversions that employ reflection data that include a 1-decade band about the sound speed transition region (from LF to HF) successfully predict the dispersion, the attenuation, the normal reflection coefficient, and the shear speed over a 4-decade band (100 Hz–1 MHz). In order to reveal the physics of sediment acoustics based on the recent published data, broadband data-model comparisons have been made in many papers.[12–21] In the comparisons, however, there were no LF sound attenuation data as a constraint.

A LF database would be helpful for analyzing the physics of sound propagation in the marine sediments, and for broadband data-model comparisons. Kibblewhite[22] published a comprehensive review paper in 1989 with a special emphasis on the LF attenuation of sound in marine sediments. At that time, the LF-field data for sandy bottoms were very limited. Since then, the accumulation of LF attenuation data has progressed. The need for supplementary attenuation data below 2 kHz and the availability of new measurements prompted the current study to analyze and summarize relevant LF-field measurements by the SW acoustics community. The relevant SW measurements, conducted in 20 locations in different coastal zones around the world, include bottom reflection loss, TL, and broadband pulse propagation. The various analyses involve normal-mode spatial filtering, inferences on dispersion, transition range of sound propagation decay laws, vertical coherence of reverberation/propagation, signal time series spreading, matched-field processing (MFP), and Hankel transform methods. The results were summarized and presented at the 149th and 151st meetings of the Acoustical Society of America.[23,24] This paper attempts to explore whether the various analyses of these LF data can provide additional insight on this subject. Holmes *et al.*[25] recently published a short review letter on nonlinear frequency dependence attenuation in sandy sediments. In contrast to this previous work, the current study discusses the details of how we obtain the LF-field-inverted attenuation values from data analysis or communications with original authors. Additionally, we provide most of the numerical values of bottom attenuation obtained from our data review/analysis rather than only the slope of the bottom attenuation with frequency.

Most of the inverted LF sound speed and attenuation in this paper were obtained from long-range acoustic field data for which the surficial sediment layer with a thickness on the order of a few wavelengths plays the dominant role. Many inversion methods treated the seabed as a fluid half-space without sediment layer-structure information. Thus, it would be better to interpret the LF-field-inverted bottom sound speed and attenuation values, as well as the Biot parameter values presented in this paper, as the average acoustic properties of an effective medium equivalent to sandy bottoms. While the proposed half-space model does not consider the effects of layering and sediment gradients, the effective properties derived from long-range (sub-critical angles) field measurements both exhibit a high degree of consistency and, moreover, are to a fidelity needed in many applied problems.

According to Hamilton,[3] "a *geoacoustic model* is defined as a model of the real seafloor with emphasis on measured, extrapolated, and predicted values of those properties important in underwater acoustics and those aspects of geophysics involving sound transmission." Following this definition, the paper is organized as follows: Section I deals with the motivation/background of this paper. The LF-field measurements that resulted in the inference of a nonlinear frequency dependence of bottom attenuation are described in Sec. II. Some comments on the geoacoustic inversion methods used to make the inferences on the attenuation values are also briefly discussed. All the sound attenuations and sound speeds in the sandy bottoms, inferred from LF measured data and discussed in Sec. II, are summarized as a function of frequency in Sec. III. The inferred sound speeds and attenuations derived from 20 sites are compared with the Biot–Stoll model in Sec. IV with emphasis on the extrapolated and predicted values of the effective Biot parameters that are important for LF sound transmission and sediment acoustic modeling in SW. A combination of the LF data with the SAX99 data[19] and other mid-frequency to HF measurements is given in Sec. V, offering a reference data set on sound speed and attenuation in the 50–400 000 Hz band. The broadband data-model comparisons and discussions are also given in Sec. V. This section shows that there is a pressing need to obtain high quality data of sound speed and attenuation in seabottoms that cover the LF- to HF transition band. Finally, Sec. VI summarizes and discusses the results of this paper.

II. LF SHALLOW-WATER MEASUREMENTS

This section summarizes LF measurements, the geoacoustic inversion methods, as well as the inversion results from 20 locations with sandy and sandy mixture bottoms that have a sound speed higher than the speed in the adjoining water, along with what supporting data are available. Among the first 12 locations, 10 of them are in the Yellow Sea and the East China Sea. (The results from the other two locations are used to compare with the results obtained from the China seas using similar inversion methods.) The other eight locations are in the eastern coastal zones of the United States. Numerical values of the sound speed and attenuation for each site are presented. The averaged values over different locations will be provided in Sec. III.

In general, the sound attenuation in the sediment may exhibit complex frequency dependence over a broadband. For our effective medium model, it is assumed that in a given frequency band, the attenuation can approximately be expressed by an empirical form of a power law

$$\alpha = k_b f^n \text{ dB/m}, \tag{1}$$

where f is frequency in units of kHz. The logarithm of both sides of Eq. (1) is taken, and then the least-squares method is used to curve-fit the LF-field-inverted bottom attenuation data to obtain the values of the constants k_b and n, and their standard deviations (that represent only the curve-fit uncertainty, not any of the other uncertainties in the measurements or methods).

A. Frequency dependence of bottom reflection loss at small grazing angles from three sites in the Yellow Sea

For a homogeneous half-space, the bottom reflection coefficient $V(\vartheta,f)$ is approximately[26–29,36]

$$-\ln|V(\vartheta,f)| = \begin{cases} Q(f)\vartheta, & 0 \le \vartheta \le \vartheta_c \\ -\ln|V_0| = \text{const}, & \vartheta_c \le \vartheta \le \dfrac{\pi}{2}. \end{cases} \tag{2}$$

Here ϑ is the grazing angle and ϑ_c is the critical angle. $Q(f)$ is the reflection loss gradient (Np/rad) at grazing incidence. The bottom reflection loss (dB) at small grazing angles is

$$-20 \log|V(f,\vartheta)| = 8.686 Q(f)\vartheta. \tag{3}$$

If the bottom attenuation is expressed by Eq. (1), and the shear wave in the sediment is neglected, then one can obtain[26–29,75]

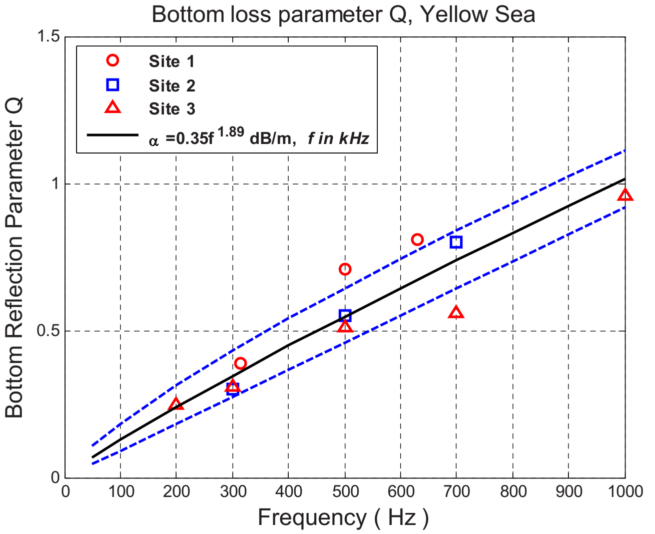


FIG. 1. (Color online) Reflection loss gradient Q vs frequency at three sites in the Yellow Sea.

$$\begin{aligned} Q(f) &= 0.0366 \frac{(c_w^2/c_b)(\rho_b/\rho_w)}{[1-(c_w/c_b)^2]^{3/2}} \times \frac{\alpha}{f} \\ &= 0.0366 \frac{(c_w^2/c_b)(\rho_b/\rho_w)}{[1-(c_w/c_b)^2]^{3/2}} k_b f^{(n-1)}, \end{aligned} \tag{4}$$

where α is the sound attenuation (dB/m) in the bottom and f is frequency in kHz. c_w and c_b are the sound speeds (km/s) in the water and the bottom, respectively. ρ_w and ρ_b are the densities (kg/m^3) in the water and the bottom, respectively.

If the bottom attenuation depends linearly on frequency, as the Hamilton model predicts, the frequency exponent in Eq. (1) is unity ($n=1$). In this case Q in Eqs. (3) and (4) should be independent of frequency. However, experimental measurements often show Q increasing with frequency. Figure 1 shows the values for Q obtained from three different sites (designated in this paper as sites 1–3) in the Yellow Sea.[30–32] From modal measurements and dispersion analysis, the sound speed in similar sediments in the vicinity of these three sites has been derived. The ratio of the sound speed in the bottom to the water sound speed at the water/sediment interface is about 1.056.[33–35] With this sound speed ratio, the best match between the measured LF Q values and the predictions of Eq. (4) requires that the bottom attenuation at sites 1–3 is

$$\alpha = (0.35 \pm 0.03) f^{(1.89 \pm 0.11)} \text{ dB/m}, \tag{5}$$

where f is frequency in units of kHz. Theoretical Q values obtained from the average bottom attenuation given by Eq. (5) are plotted in Fig. 1 by a solid line. Two dashed lines in the figure express the regression uncertainty.

B. The second transition range of decay laws in Pekeris model (north of Elba Island in the Mediterranean Sea)

The Pekeris waveguide consists of an isospeed water layer over an isospeed half-space bottom. Average sound intensity as a function of range in the Pekeris isovelocity SW can be divided into four regions with different "decay

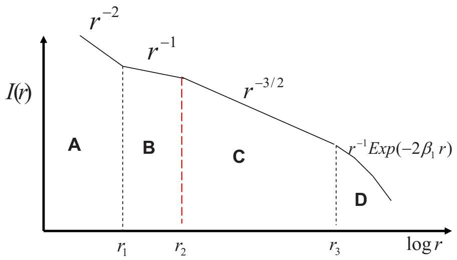


FIG. 2. (Color online) Transition from one decay law to another in the Pekeris SW waveguide.

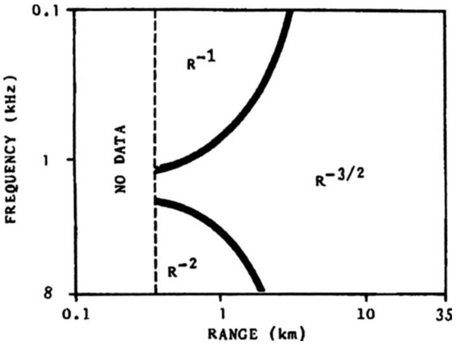


FIG. 3. Transitions between decay laws for the winter data at a site north of Elba Island (from Ref. 39).

laws"[26,36,37] as shown in Fig. 2. The average sound intensity in four propagation regions can be expressed (the sound absorption in the water column is omitted) by the following.[26,36,37]

(A) For $r_0 < r < r_1$, spherical spreading region $\sim r^{-2}$,

$$I_A(r) \approx \frac{4}{-\ln|V_0|}\frac{1}{r^2}. \tag{6}$$

(B) For $r_1 < r < r_2$, cylindrical spreading region $\sim r^{-1}$,

$$I_B(r) = \frac{2\vartheta_c}{H}\frac{1}{r}. \tag{7}$$

(C) For $r_2 < r < r_3$, three-halves law region $\sim r^{-3/2}$,

$$I_C(r) = \sqrt{\frac{\pi}{QH}}\frac{1}{r^{3/2}}. \tag{8}$$

(D) For $r > r_3$, single mode (first mode) decay region,

$$I_D(r) = \frac{2\pi}{kH^2}\frac{1}{r}\exp\left[-\frac{Q\pi^2 r}{k^2H^3}\right]. \tag{9}$$

Three transition ranges in the Pekeris model are defined by[26,36,37]

$$r_1 = \frac{H}{(-\ln|V_0|)\vartheta_c}, \tag{10}$$

$$r_2 = \frac{H}{Q\vartheta_c^2}, \tag{11}$$

$$r_3 = \frac{k^2H^3}{Q\pi^2}, \tag{12}$$

where H is the water depth and k is wave number. In general, the three-halves law region is of common interest in SW and was first derived by Brekhovskikh.[38]

Using the experimental data collected from a SACLANTCEN trial zone (designated site 4 in this paper), north of Italian Elba Island (H=110 m) during winter with a quasi-isovelocity profile, Murphy and Olesen[39] offered convincing evidence for the three-halves law. Their experimental results showed that from about 2 to 35 km the standard deviation from the three-halves law for 20 one-third octave bands is less than 1 dB. Figure 3 is copied from Fig. 6 in Ref. 39, summarizing the transitions from one decay law to another for the winter data collected from site 4. In this area, at 400 Hz, the normal-mode attenuation increases abruptly around the 17th mode (see Fig. 9 in Ref. 40). This means that the equivalent critical angle of seabottom reflection in this trial zone is about 17°, i.e., $\vartheta_c \approx 16.7°$, $c_b/c_w \approx 1.044$. Using this value of the critical angle and the second transition range (from the cylindrical law to three-halves law shown in Fig. 3) in Eq. (11), one can derive the reflection loss gradient Q as a function of frequency shown in Fig. 4. Again, Q at small grazing angles increases with increasing frequency. The best match between the second transition range-derived Q and the frequency dependence given by Eq. (4) implies that the bottom attenuation for the site north of Elba Island is

$$\alpha = (0.45 \pm 0.02)f^{(1.82\pm0.06)}\ \text{dB/m}, \tag{13}$$

where f is frequency in units of kHz. The solid line in Fig. 4 is obtained from Eq. (4) with the averaged bottom attenuation given by Eq. (13). Two dashed lines represent the curve-fit uncertainty.

C. Bottom sound attenuation vs frequency from normal-mode measurements

1. Gulf of Mexico

Ingenito[41] first used measured individual mode attenuation coefficients in SW to deduce the bottom attenuation for

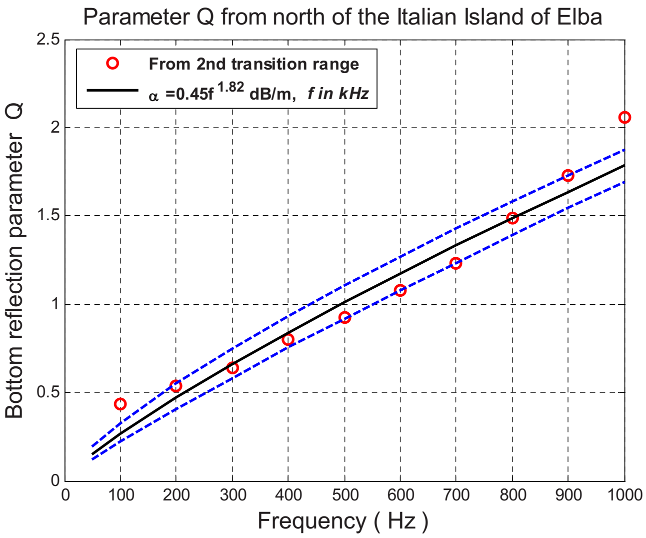


FIG. 4. (Color online) Reflection loss gradient Q vs frequency at a north site of Elba Island.

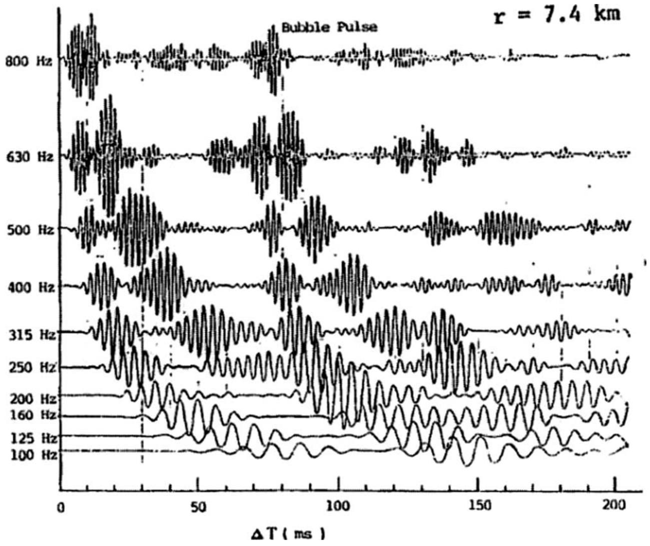


FIG. 5. Dispersion waveform of an explosive signal at 7.4 km from the source location. Abscissa is the arrival time relative to the first 5-kHz arrival.

two frequencies of 400 and 750 Hz. The measurements were conducted in an area of the Gulf of Mexico (designated site 5 in this paper) about 20 km off Panama City, FL. There the water depth is about 30 m, and the seabed consists of sand. The acoustic normal modes were separated by two methods: a temporal method based on the differing group velocities of the modes and a spatial filtering method based on their vertical pressure distribution. Measured normal-mode attenuations required an $f^{1.75}$ frequency dependence of the bottom attenuation to fit the measured data, i.e.,

$$\alpha = 0.50 f^{1.75} \text{ dB/m}, \tag{14}$$

where f is frequency in units of kHz. In Ref. 41 the density and sound speed in the bottom were determined from a bottom reflection experiment at 3.5 kHz by Ferris and Kuperman.[42] The sediment density was determined to be $\rho_b = 1.8 \text{ g/cm}^3$. The sound speed ratio at the bottom-water interface was determined to be about 1.034. These parameters were not inverted from an independent LF measurement.

2. In the Yellow Sea

Both bottom sound speed and attenuation were *simultaneously* deduced from broadband measurements for ten one-third octave bands (80–800 Hz) at two sites (designated sites 6 and 7 in this paper) in the Yellow Sea by Zhou[33] and Zhou *et al.*[35] The bottom at these two sites is very flat. The mean grain sizes of the surficial sediments at these sites are 0.0492 mm (4.35ϕ) and 0.0892 mm (3.49ϕ), respectively. Figure 5, reproduced from Fig. 1 of Ref. 35, shows the received waveform dispersion at a distance of 7.4 km from the explosive source. The water depth is 28.5 m. The time series implies that for a given frequency the mode group speed depends on mode order. The arrival time difference between modes 1 and 2 at a given distance r is given by

$$\Delta T_{12}(f) = \left[\frac{1}{V_g^2(f)} - \frac{1}{V_g^1(f)} \right] r = K_{12}(f) r. \tag{15}$$

For a given mode, the mode group speed depends on frequency. For example, for the first mode, the arrival time difference between a higher frequency (Hi) and a lower frequency (Lo) is proportional to distance,

$$\Delta T_{HL}(f) = \left[\frac{1}{V_{g1}(f_{\text{Lo}})} - \frac{1}{V_{g1}(f_{\text{Hi}})} \right] r = K_{HL}(f) r. \tag{16}$$

The two slopes (gradients), K_{12} and K_{HL}, in Eqs. (15) and (16) are sensitive to the sound speed in the bottom. If the sound speed profiles in water column c_w and bottom density ρ_b are known, the values of K_{12} and K_{HL} can be used to invert for the bottom sound speed c_b,[33,35] from the expression for the modal group speed[43]

$$\begin{aligned} V_g^{(n)} = k_n \left(\rho_w \int_0^H |U_n^w(z)|^2 dz + \rho_b \int_H^\infty |U_n^b(z)|^2 dz \right) \\ \times \left\{ \omega \left[\rho_w \int_0^H \left(\frac{|U_n^w(z)|^2}{c_w^2(z)} \right) dz \right. \right. \\ \left. \left. + \frac{\rho_b}{c_b^2} \int_{hH}^\infty |U_n^b(z)|^2 dz \right] \right\}^{-1}, \end{aligned} \tag{17}$$

where k_n and $U_n(z)$ are the eigenvalue and eigenfunction, respectively, of the nth mode. Figure 6, reproduced from Fig. 6 of Ref. 33, shows the arrival time delay between LF and HF signals for the first mode [K_{HL} in Eq. (16) is expressed by K_t' in Fig. 6]. At a given range, the experimental data (dots) are averaged values obtained from the waveforms produced by five explosive charges. The purpose of re-showing Fig. 6 here is to address an important issue that is often encountered in dispersion-based geoacoustic inversion. If an analog filter or MATLAB's "filter" command is used in the dispersion analysis for broadband signal (such as explosive) data, a least-squares fit to the data would yield a time intercept of $\Delta T_0(f)$, as shown in Fig. 6. Compared with the theoretical prediction in Eq. (16), the experimental dispersion data (such as Fig. 6) often result in a false relation

$$\begin{aligned} \Delta T_{HL}(f) &= \left[\frac{1}{V_{g1}(f_{\text{Lo}})} - \frac{1}{V_{g1}(f_{\text{Hi}})} \right] r + \Delta T_0(f) \\ &= K_{HL}(f) r + \Delta T_0(f), \end{aligned} \tag{18}$$

which involves an instrumentation error, $\Delta T_0(f)$. It is caused by a fact that a filter often has a different response time (phase shift) for different central frequencies. In such a case, if only one time-frequency diagram at a given range (such as Fig. 5) is used to invert sound speed in the bottom, the filter phase shift, corresponding to $\Delta T_0(f)$, may cause an unacceptable error.

From the slopes shown in Fig. 6, Zhou[33] inverted for sound speed in the bottom at two sites in a frequency range of 50–800 Hz. The sound speed ratio at the water/sediment interface is about 1.056.

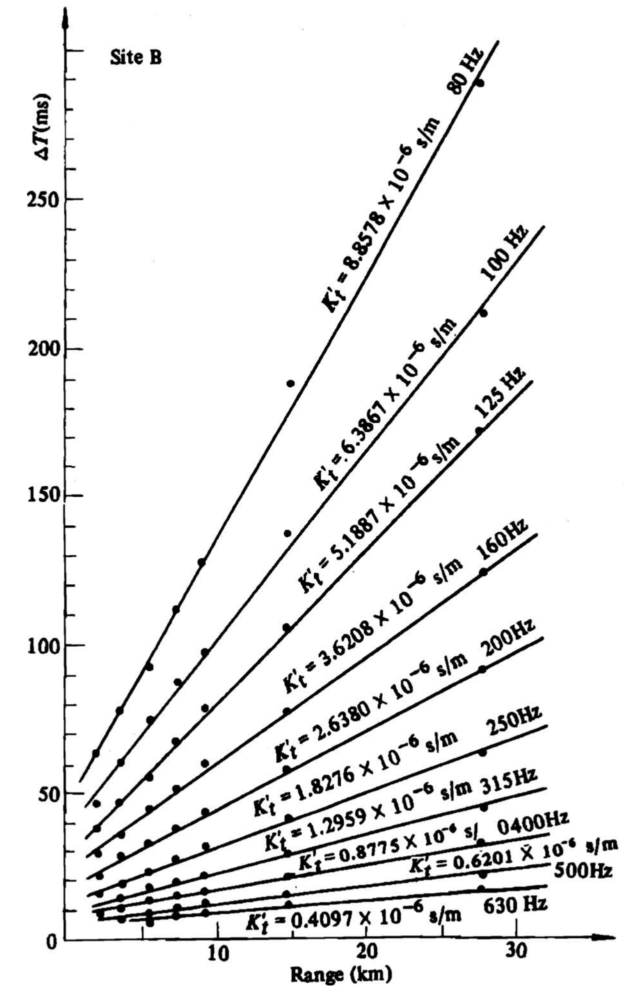


FIG. 6. The mode 1 arrival time delay, $\Delta T(f)$, between LF and 6.3 kHz as a function of range at site 7 with a time intercept of $\Delta T_0(f)$ (site B from Ref. 33, ©1985 Acoustical Society of America).

After the sediment sound speed is extracted from mode dispersion measurements, the only unknown quantity in the expression for mode attenuation is the sediment attenuation. From the data/model comparison for the attenuation of an individual normal mode, Zhou[33] inverted bottom attenuation as a function of frequency for sites 6 and 7 in the 80–800 Hz band. As discussed in Ref. 33, the higher the signal frequency, the larger the measurement error for the dispersion-inverted bottom acoustic parameters. Thus, only data in the 80–500 Hz band are used for describing the bottom attenuation at these two sites. The averaged attenuation from sites 6 and 7 is listed in Table I. Using a power law fitting, it can approximately be expressed by

$$\alpha = (0.38 \pm 0.08) f^{(1.87\pm0.13)} \text{ dB/m}, \tag{19}$$

where f is frequency in units of kHz.

3. "Speed-attenuation coupling" in TL-only based inversions

Here we discuss an important issue on the seabed geoacoustic inversion in SW: speed-attenuation coupling. This concept will be used in Secs. II G and II M, and in Sec. IV.

The attenuation factor β_n of the nth normal mode in SW can be expressed[43]

$$\beta_n(\omega) = \frac{\omega}{k_n N_n} \int_0^\infty \frac{\rho_b(z)\alpha(z)}{c_b(z)} |U_n(z)|^2 dz. \tag{20}$$

According to the WKB approximation, it can alternatively and approximately be expressed in terms of a bottom reflection coefficient $V(\vartheta)$ as

$$\beta_n(\vartheta) \approx \frac{\ln|V(\vartheta)|}{S_n}.$$

For waveguide modes with small grazing angles, using Eq. (4) one has

$$\beta_n(\vartheta_n) \approx -\frac{Q\vartheta_n}{S_n} = -0.0366 \frac{(c_w^2/c_b)(\rho_b/\rho_w)}{[1-(c_w/c_b)^2]^{3/2}} \times \frac{\alpha}{f} \times \frac{\vartheta_n}{S_n}, \tag{21}$$

where ϑ_n is a grazing angle of the nth mode ray; S_n is the cycle distance of nth mode and is mainly controlled by the sound speed profile in the water column. The LF sound TL in SW is controlled by modal attenuation factors (i.e., energy loss in the seabottom). Equations (20) and (21) show that, for a set of given single frequency modal attenuation factors, an increase (decrease) in bottom sound speed can be compensated for by an increase (decrease) in bottom attenuation. Thus, if only sound energy loss such as TL or mode attenuation is used to invert seabottom attenuation, one must have ground truth measurements on sound speed in the bottom as an inversion constraint. Otherwise, the inverted attenuation might involve unacceptable errors. We call this speed-attenuation coupling in TL-only bottom geoacoustic inversion. The sound speeds in the bottom at sites 6 and 7 were independently derived from LF mode measurements, and then used as a constraint for the inversion of sound attenuation in bottom. Thus, it was the first time that both broadband bottom sound speed and attenuation were deduced from the same LF-field measurements at two SW sites.

D. Geoacoustic inversion from the Yellow Sea 1996 experiment

1. Sound speed from the maximum correlation between the measured and predicted waveforms; attenuation from transmission loss

The first joint China–United States underwater acoustic experiment was conducted in 1996 in the Yellow Sea where the water depth is 75 ± 1 m with a strong thermocline. A core

TABLE I. Bottom attenuation inverted from modal measurements at sites 6 and 7.

f (Hz)	80	100	125	160	200	250	315	400	500
α (dB/m)	0.0022	0.0059	0.0108	0.0137	0.0186	0.0274	0.0388	0.0571	0.1136

TABLE II. Bottom attenuation at Yellow Sea 1996 site (site 8a).

f (Hz)	200	600	1000	1200	1500
α (dB/m)	0.0139	0.1059	0.2710	0.4234	0.6144

analysis of the seabed constituents implies that the surface sediment in this area (designated site 8a in this paper) can be categorized as fine sand-silty sand. The mean grain diameter of the sediment M_d=0.0643 mm (3.96ϕ). During the TL measurements, the sea surface was smooth.

The geoacoustic inversions of the Yellow Sea 1996 data by Li and Zhang[44] and Rogers *et al.*[45] exploited the availability of data taken on a vertical array. The approach was to compare the measured and modeled waveforms produced at each hydrophone in the array from a source at the relatively short range of 2.6 km. The theoretical time series waveforms were obtained from the normal-mode model by Fourier synthesis. The average sediment sound speed at the Yellow Sea 1996 site, inverted from the time waveform correlations, is 1587.5 m/s in the 200–1000 Hz range. The sound speed ratio at the water-sediment interface is 1.073. With this bottom speed as a constraint, the best match between measured and numerical TL, requires the values listed in Table II for the attenuation of sound in the bottom at five frequencies. Using power law fitting, the bottom attenuation in a frequency range of 200–1500 Hz can be approximately expressed by

$$\alpha = (0.29 \pm 0.01) f^{(1.89 \pm 0.03)} \text{ dB/m}, \tag{22}$$

where f is frequency in units of kHz.

2. Sound speed from broadband cross-spectral density matrices and modal dispersion; attenuation from transmission losses

Wan *et al.*[46] used a cross-spectral density matrix (CSDM) method, mode dispersion characteristics, and TL for the seabed geoacoustic inversion from the Yellow Sea 1996 data (designated site 8b in this paper). Broadband signals measured on a vertical line array from an explosion at long range were used to form the CSDM for extracting the depth function of normal modes.[47] Then the bottom sound speed and mode eigenvalue were inverted from the best match between measured and theoretical depth functions of the normal mode. The CSDM-inverted sound speed in the bottom at 100 Hz is 1588.1 m/s. The arrival time differences between modes 1–3 were also used to derive the bottom speed at 150–400 Hz. This resulted in an average bottom sound speed of 1588.3 m/s. The average speed ratio at the bottom-water interface, obtained from two methods, is 1.074 for the 100–400 Hz range.

The broadband TL data were carefully analyzed as a function of range, frequency, receiver depth, and azimuth in two radial directions up to 55 km and along a quarter of a circle with a radius of 34 km. With the inverted bottom speed as a constraint, the best match between the measured and the modeled TL, obtained from hydrophones located both below and above the thermocline, resulted in sound attenuations in the bottom for the Yellow Sea 1996 site (site 8b) that are listed in Table III. Using the power law fitting, the bottom attenuation at site 8b in a 80–1000 Hz range can be expressed by

$$\alpha = (0.36 \pm 0.03) f^{(1.95 \pm 0.06)} \text{ dB/m}, \tag{23}$$

where f is frequency in units of kHz.

E. Geoacoustic inversion from the Yellow Sea 2002 experiment

The Yellow Sea 2002 experiment was conducted at three sites in the Yellow Sea.[48] Two sites (designated sites 9 and 10 in this paper) are in the vicinity of site 6 (cf. Sec. II C 2) The bottom at site 9 is silty sand; site 10 is a sandy-clay mixture. To account for the different sensitivities of acoustic field parameters to seabed acoustic parameters, Li and Zhang[48] combined several inversion methods to invert for acoustic parameters in the bottom: (1) The characteristic impedance ($\rho_b c_b$) of the bottom was inverted from bottom vertical reflection coefficients; (2) the reflection-inverted bottom characteristic impedance was used as a constraint in the MFP inversion method to obtain sound speed in the bottom at short source ranges where the MFP was sensitive to sediment sound speed but less sensitive to attenuation in the bottom; (3) then, with inverted density and sound speed in the bottom as a constraint, measured TL data from all hydrophones (located above or below the thermocline) were used to infer the sediment sound attenuation.

In a frequency range of 70–200 Hz, the inverted sound speed ratios at the bottom/water interface at sites 9 and 10 are 1.069 and 1.045, respectively. The inverted bottom attenuations from sites 9 and 10 are listed in Table IV. By using power-law fitting, we have the following.

For site 9,

$$\alpha = (0.32 \pm 0.02) f^{(1.88 \pm 0.07)} \text{ dB/m} \quad \text{for } 100-1000 \text{ Hz}$$

and

$$\alpha = (0.28 \pm 0.01) f^{(1.76 \pm 0.04)} \text{ dB/m} \quad \text{for } 100-2000 \text{ Hz}. \tag{24}$$

for site 10,

$$\alpha = (0.28 \pm 0.02) f^{(2.23 \pm 0.08)} \text{ dB/m} \quad \text{for } 100-1000 \text{ Hz}$$

and

$$\alpha = (0.25 \pm 0.01) f^{(2.16 \pm 0.04)} \text{ dB/m} \quad \text{for } 100-2000 \text{ Hz}, \tag{25}$$

where f is frequency in units of kHz.

TABLE III. Bottom attenuation at Yellow Sea 1996 site (site 8b).

f (Hz)	80	100	125	160	200	250	315	400	500	630	800	1000
α (dB/m)	0.002 29	0.004 02	0.006 39	0.009 12	0.022 58	0.021 89	0.041 38	0.065 03	0.080 55	0.138 92	0.186 36	0.442 71

TABLE IV. Bottom attenuation inverted from Yellow Sea 2002.

					Site 9					
f (Hz)	100	200	300	400	500	600	700	800	900	1000
α (dB/m)	0.003 75	0.0150	0.045	0.050	0.100	0.128	0.166	0.190	0.236	0.313
f (Hz)	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
α (dB/m)	0.344	0.405	0.488	0.508	0.525	0.580	0.701	0.743	0.736	0.825
					Site 10					
f (Hz)	100	200	300	400	500	600	700	800	900	1000
α (dB/m)	0.001 28	0.007 69	0.0269	0.0410	0.0577	0.0846	0.126	0.164	0.196	0.244
f (Hz)	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
α (dB/m)	0.310	0.369	0.433	0.485	0.539	0.656	0.806	0.900	0.999	1.128

F. Geoacoustic inversion from the ASIAEX site in the East China Sea

The Asian Sea International Acoustic Experiment (ASIAEX) was conducted in 2001 in the East China Sea where the water depth is about 104 m.[49] The core analysis shows that the sediment at this area (designated site 11 in this paper) is very fine sand-silty sand. The mean grain diameter over the entire ASIAEX area is 0.043 mm (4.54 ϕ); close to the receiving array, the grain diameter is 0.090 mm (3.47 ϕ).

1. Geoacoustic inversion based on modal filtering, dispersion analysis, and transmission loss

Using normal-mode spatial filtering and dispersion analysis, Peng *et al.*[50] derived a bottom density of 1.82 g/cm^3 and a bottom speed of 1607 m/s at the ASIAEX site[50] (sound speed ratio=1.058 for 50 and 200 Hz). With the inverted LF bottom sound speed and density as constraints, the bottom attenuation was inverted from measured TL in a frequency range of 100–500 Hz. The results are listed in Table V, and are approximately expressed as

$$\alpha = (0.19 \pm 0.02) f^{(1.55 \pm 0.08)} \text{ dB/m}, \tag{26}$$

where f is frequency in units of kHz.

2. Geoacoustic inversion from a combination of MFP, bottom reflection, and vertical coherence of sound propagation

Using the MFP inversion, combined with bottom vertical reflection coefficients and vertical correlation of sound propagation, Li *et al.*[51] derived the following bottom acoustic parameters: density=1.86 g/cm^3; sound speed=1610 m/s in 100–200 Hz range (the sound speed ratio=1.060); the sound attenuations in the seabed for frequencies in the range 100–600 Hz are also listed in Table V. With the power-law fitting, it can be expressed by

$$\alpha = (0.34 \pm 0.04) f^{(1.80 \pm 0.09)} \text{ dB/m}, \tag{27}$$

where f is frequency in units of kHz.

3. Geoacoustic inversion from reverberation vertical coherence

From reverberation vertical coherence (RVC) measurements, Zhou *et al.*[52] derived the average bottom speed and attenuation at the center of the ASIAEX area within a radius of 9 km. The reverberation measurements were conducted at the same location on June 3 and June 5, 2001. The pair of bottom sound speed and attenuation values for which the numerical RVC curves best match the experimental RVC data in a least-error-squared sense is determined using a numerical search. The average value of sound speed in the bottom, inverted from RVC measurements on June 3 and June 5, is 1612 m/s (sound speed ratio of 1.061). The sound attenuation in the bottom, inverted from two RVC measurements, exhibits a nonlinear frequency dependence in the frequency range of 100–1200 Hz. From the RVC measurement made on June 3, the inverted sound attenuations are listed in Table V. It can be approximated as

$$\alpha = (0.31 \pm 0.11) f^{(1.84 \pm 0.27)} \text{ dB/m} \quad \text{for } 100-600 \text{ Hz}$$

and

TABLE V. Bottom attenuation at the ASIAEX site in the East China Sea. (Site 11).

					Peng *et al.*[a]					
f (Hz)	100	200	300	400	500					
α (dB/m)	0.005	0.015	0.032	0.047	0.056					
					Li *et al.*[b]					
f (Hz)	100	200	300	400	500	600				
α (dB/m)	0.0050	0.0199	0.0466	0.0745	0.0901	0.1230				
					Zhou *et al.*[c]					
f (Hz)	100	200	300	400	500	600	700	800	1000	1200
α (dB/m)	0.0055	0.0134	0.0312	0.0320	0.1360	0.1386	0.1106	0.0864	0.2220	0.2316

[a] Reference 50.
[b] Reference 51.
[c] Reference 52.

TABLE VI. Bottom attenuation at site 12 in the East China Sea.

f (Hz)	30	50	80	100	160	200	250	300	350	400	450	500	600	700	800
α (dB/m)	0.000 38	0.001 06	0.002 71	0.004 24	0.010 84	0.016 93	0.026 45	0.0381	0.0518	0.0676	0.0855	0.1054	0.1516	0.2056	0.2678

$$\alpha = (0.20 \pm 0.04) f^{(1.57\pm0.16)}\ \text{dB/m} \quad \text{for } 100-1200\ \text{Hz}, \tag{28}$$

where f is frequency in units of kHz.

4. Geoacoustic inversion using decoupling inversion method

A decoupling inversion method was employed by Knobles *et al.*[53,54] to estimate the surface sediment sound speed based on the ASIAEX time series data for 50–350 and 350–500 Hz bands. The inversion resulted in a sound speed of 1605–1621 m/s for a top sediment layer thickness of 6.7 m at the ASIAEX site. The averaged sound speed is 1613 m/s (sound speed ratio of 1.062). The best fit between predicted and measured time series in the 50–500 Hz band requires that the bottom attenuation at the ASIAEX site has the following value (see Table II of Ref. 53):

$$\alpha \approx 0.56 f^{2.0}\ \text{dB/m}, \tag{29}$$

where f is frequency in units of kHz. The different frequency exponents of sound attenuation from one site may reflect uncertainties and errors, cased by different measurements and methodologies in different frequency bands.

G. Geoacoustic inversion from the wideband transmission loss at a site in the East China Sea

Using a simulated annealing algorithm, Knobles *et al.*[53] inverted bottom sound speed and attenuation from the broadband TL in the 25–800 Hz band, acquired at a location in the East China Sea (designated site 12 in this paper). The experimental site was centered at 29°05'N, 126°43'E. It was about 65 km south of the ASIAEX site (site 11). Knobles *et al.*[53] reviewed and combined the geophysical survey data by different groups, and concluded that the first layer of sediments at sites 12 and 11 had similar properties. They constructed an approximate ground truth description of the seabed as a fine to medium sand sediment with a sand content ranging from 65% to 80%. The porosity was about 45%, average grain size was on the order of 3.3ϕ, densities ranged from 1.7 to 1.95 g/cm^3, and surface sound speed ranged from 1590 to 1655 m/s.

Five representations of seabed cases (models) with different combinations of hypothetical frequency dependence of the sound speed and attenuation were tested in the analysis of the multi-frequency octave average TL data taken at site 12.[53] Two cases with linear frequency dependence of seabottom attenuation were rejected by the existing geophysical data. Only those cases (cases 2, 4, and 5) that had a nonlinear frequency dependence of attenuation led to a reasonable data/model comparison. Resultant bottom attenuations as a function of frequency, derived from five cases, were given in Fig. 7b in Ref. 53. Although cases 2, 4, and 5 all gave a reasonable match with the data, in this paper we adopt the result of case 5, because it is a modified Biot model with a more solid physical base. The broadband TL data/model comparison in case 5 required the bottom attenuation to vary as $0.99f^{2.00}$ in a frequency range of 30–800 Hz. The sound attenuation in the bottom at site 12 exhibits a similar frequency dependence as the attenuation data obtained from sites 1–11, but the attenuation magnitude is larger at site 12. One possible reason is the higher bottom speed ratio of 1.11, i.e., due to the speed-attenuation coupling in the TL-only based inversion, discussed in Sec. II C 3. The site is about 65 km south of the ASIAEX experimental location. These two sites have similar bottom properties. Thus, the speed ratio at site 12 should be close to 1.06, obtained from site 11 by four different groups using different LF-field characteristics. Figures 8 and 12 will also indicate that the average LF sound speed ratio in sandy bottoms is around 1.061 ± 0.009 (cf. Sec. III B) with no apparent speed dispersion at LFs. Based on these considerations, changing the speed ratio from 1.11 to 1.06 and keeping mode attenuation rates and TL data unchanged, the broadband TL data/model comparison in case 5 will result in a "normalized" bottom attenuation shown in Table VI for site 12. It can approximately be expressed by

$$\alpha = 0.42 f^{2.00}\ \text{dB/m}, \tag{30}$$

where f is frequency in units of kHz.

H. Geoacoustic inversion in an area, south of Long Island, NY

A number of measurements on TL were conducted in an area 40 nm (75 km) south of Montauk, Long Island, NY (designated site 13 in this paper) with a water depth of 55–60 m.[55–58] The bottom is known to consist of medium grain size sand. All of the TL measurements were made under downward refracting conditions in the water column. Broadband sources were used in measurements in 1965 and 1967, and narrowband sources at 1000 and 500 Hz were used in 1991 and 1992 measurements. An earlier analysis of the acoustics measurements found a good agreement in data-model TL comparisons at higher frequencies for a fluid half-space model of the seabed with the attenuation linear with frequency. However, modeled TLs at 100 and 500 Hz, obtained from a bottom model that matched the higher frequency measurements on the basis of a linear extrapolation, were much larger than the measured TL data.

Two sediment sound speed measurements (a LF modal dispersion measurement and a HF core measurement) taken in the vicinity of the test area showed evidence of frequency dispersion.[55,56] The LF (100 Hz) measurement showed a sound speed ratio of 1.050,[55] while the HF (400 kHz) core measurement showed an averaged speed ratio of 1.13.[56] Thus, instead of the traditional homogeneous fluid or solid representation of the seabed with a linear frequency dependence of bottom attenuation, Tattersall, Chizhik, Cole, and

TABLE VII. Bottom attenuation at the south of Montauk, Long Island, NY (site 13).

f (Hz)	50	80	100	125	160	200	250	300	315	400	500	600
α (dB/m)	0.0016	0.0039	0.0059	0.0090	0.0143	0.0220	0.0337	0.0477	0.0524	0.0822	0.1244	0.1729
f (Hz)	700	800	900	1000	1200	1500	1600	2000	3000	4000	5000	
α (dB/m)	0.2265	0.2840	0.3443	0.4063	0.5325	0.7194	0.7799	1.0105	1.5214	1.9809	2.4133	

DiNapoli (TCCD) (Ref. 56) used a Biot geoacoustic model to redo the analysis of experimental data. The Biot geophysical parameters they chose are listed in Table X labeled by TCCD. A porosity of 0.43 was obtained from saturated density measurements of core samples. Both pore size (3.0×10^{-5} m) and permeability (1.25×10^{-11}) for the Biot model were adjusted to match measured TL for 1.0–3.5 kHz. The bottom attenuation obtained from the Biot model with these Biot parameters is listed in Table VII. Numerical TL based on the Biot model with TCCD parameters is in good agreement with the 1965 and 1967 broadband measurements from 100 to 8000 Hz as well as with later 500 and 1000 Hz narrowband measurements in 1991 and 1992. With the power-law fitting, the bottom attenuation (listed in Table VII) in the New York Bight in a frequency range of 50–1000 Hz can be expressed by

$$\alpha = (0.44 \pm 0.01) f^{(1.86 \pm 0.01)} \text{ dB/m}, \qquad (31)$$

where f is frequency in units of kHz.

I. Geoacoustic inversion on the continental shelf off New Jersey

Evans and Carey[59] also argued that the extrapolation of bottom attenuation with an assumption of linear frequency dependence would result in a disagreement between predictions and measurements on sound TL. The measurements were conducted on the continental shelf off New Jersey, near AMCOR borehole 6010 (designated site 14a in this paper). The bottom was characterized as sandy. The water depth was about 73 m. TL measurements, taken in 1988 and 1993 along the same cross slope track, were examined in a frequency range of 50–1000 Hz. The bottom attenuation at 50 Hz in this area was determined by Badiey *et al.*[60] (see Table I of Ref. 60). With the value of the bottom attenuation at 50 Hz, the calculated TL was in good agreement with measured TL at 50 and 75 Hz. However, computed TL at higher frequencies, based on a geoacoustic model with a linearly extrapolated sediment volume attenuation, was much lower than measured. Evans and Carey[59] found that a geoacoustic model with a nonlinear frequency dependence of bottom attenuation could provide a good fit to data for frequencies between 50 and 1000 Hz. The best fit between predictions and TL data from separate years' measurements required that the sound attenuation in the top 5 m layer of bottom for the 50–1000 Hz had the following value:

$$\alpha = 0.29 f^{1.57} \text{ dB/m}, \qquad (32)$$

where f is frequency in units of kHz.

Dediu *et al.*[61] recently reexamined the TL data from the same measurements in this site (designated site 14b). Extensive comparisons between measured and calculated TLs for frequencies between 400 and 1000 Hz showed that the sound attenuation in the upper sediments can be expressed by[25,61]

$$\alpha = (0.34 \pm 0.10) f^{(1.85 \pm 0.15)} \text{ dB/m}, \qquad (33)$$

where f is in kHz.

J. Bottom attenuation inverted from sound transmission loss in the Strait of Korea

The experiment called the Acoustic Characterization Test III was performed in the oceanographically complex Strait of Korea in August and September of 1995 (designated site 15 in this paper).[62] Bottom sampling and sub-bottom surveys coupled with archival geophysical information provided the basis for the geoacoustic depth profiles of sound speed, density, and attenuation. The bottom was a sandy sediment. Using the sediment compressional wave sound speed profiles, the measured bathymetry, and the water sound speed as input parameters, Rozenfeld *et al.*[62] found that good agreement was obtained between measured and calculated narrowband TLs with an attenuation profile with a $f^{1.8}$ frequency dependence in the near sediment-water interface layer. This power law was determined by using an effective attenuation coefficient and a least-squares comparison of calculated and measured sound transmissions of five narrowband tones between 47 and 604 Hz. The resulting geoacoustic model was used to compare measured and calculated broadband sound transmissions and signal spreads, and excellent agreement was found. In the study, grab sample measurements yielded a sound speed ratio of 1.08 ± 0.01 and a density of 1.89 ± 0.03 kg/m^3. (The sound speed ratio was not obtained by an independent LF inversion method. Only a HF could be used for sound measurement for a grab sample.) The sound attenuation in the top layer of sediments in the Strait of Korea was expressed by[62]

$$\alpha = 0.28 f^{1.84} \text{ dB/m}, \qquad (34)$$

where f is frequency in kHz.

K. Bottom attenuation from the complex pressure field vs range in Nantucket Sound, MA

Using data from an autonomous underwater vehicle-towed hydrophone array and a coherent synthetic aperture Hankel transform method, Holmes *et al.*[63] inverted sound attenuation in the bottom at a site in the Nantucket Sound, MA (designated site 16 in this paper). The seabed at site 16 is described as a sandy bottom. The porosity of sediments is 0.49, the mean grain size is 0.115 mm (3.12 ϕ), and the water depth is 13 m. The Hankel transform-inverted sound attenuations in the seabed at site 16 are listed in Table VIII and can be expressed by a power law

TABLE VIII. Bottom attenuation in Nantucket Sound, MA (site 16).

f (Hz)	220.5	415	635	823	1031	1228
α (dB/m)	0.0177	0.0521	0.1680	0.1209	0.2627	0.3859

$$\alpha = (0.25 \pm 0.04) f^{(1.73\pm0.20)} \text{ dB/m}, \tag{35}$$

where f is frequency in units of kHz.

L. Sound attenuation in the sediments at the SW06 site on the New Jersey continental shelf

A comprehensive SW 2006 experiment (SW06) was conducted on the continental shelf off New Jersey. Using two *L*-arrays and the combustive sound source (CSS) as well as continuous wave (CW) signals in 50–3000 Hz, Knobles *et al.*[64] inverted sound speed and attenuation in the bottom at this site (designated site 17 in this paper). On the propagation track along a sand ridge, sediment grain sizes lie between 0.8 and 1.5ϕ, placing them in the medium-to-coarse sand category. The LF signal arrival structure is sensitive to sound speed and layer structure. The LF interferences between normal modes as a function of range are also sensitive to sound speed in the bottom. These facts were used to derive the sound speed in the bottom from the best match between measured and modeled CSS time series and mode interference patterns. The average sound speed in the top sediments layer (about 3 m in thick) is about 1650 m/s; in the second layer of 20 m the speed is about 1590 m/s. Using the LF inverted bottom sound speed as a constraint the sound attenuation in the bottom was derived from TL data. The inverted attenuation data in a frequency range of 53–2003 Hz are listed in Table IX. The attenuation can be expressed by

$$\begin{aligned}\alpha &= (0.42 \pm 0.05) f^{(1.74\pm0.07)} \text{ dB/m} \quad \text{for } 53-503 \text{ Hz}\\ &= (0.61 \pm 0.06) f^{(1.92\pm0.07)} \text{ dB/m} \quad \text{for } 53-2003 \text{ Hz},\end{aligned} \tag{36}$$

where f is frequency in units of kHz. Here, we might again find the speed-attenuation coupling at higher frequencies ($f > 700$ Hz): A top 3 m layer sediment with a high sound speed ratio results in higher bottom attenuation.

M. Bottom attenuation inverted from transmission loss off Daytona Beach, FL

Experimental and predicted TLs from several SW sites were compared by Beebe *et al.*[65] The predicted loss was obtained using a normal-mode model with the bottom attenuation calculated using the Biot sediment model. The bottom attenuation was predicted to vary as $f^{1.76}$ for a site off Daytona Beach, FL (designated site 18 in this paper) having a medium-to-coarse sand bottom (the mean grain size was 0.85 ϕ and the porosity was 0.38). The agreement between measured and predicted TLs was acceptable for frequencies from 100 to 600 Hz. From a TL data/model comparison, the bottom attenuation for the site off Daytona Beach, FL was expressed by[65]

$$\alpha = 2.096 f^{1.76} \text{ dB/m}, \tag{37}$$

where f is frequency in units of kHz.

The authors obtained a nonlinear frequency dependence for the bottom attenuation that is similar to sites 1–17. However, the magnitude of sediment sound attenuation they obtained was much higher than obtained for sites 1–17. It might be also due to the speed-attenuation coupling problem in the TL-only-based inversion, discussed in Sec. II C 3. The high sound speed of 1723 m/s (sound speed ratio of 1.126), obtained from seismic-refraction analysis (which might be from a very LF and for much deeper bottom), was used for the TL data/model comparison that resulted in much higher bottom attenuation amplitudes.

Based on the following physical considerations, we suggest using a lower speed ratio to get normalized bottom attenuation from same set of TL data at site 18. (i) Our experimental data shown in Fig. 8 indicate that the LF sound speed ratio at the bottom-water interface for sandy bottom, including the medium-to-coarse sand, is between 1.04 and 1.08 with an average value of 1.061 ± 0.009. (ii) Beebe *et al.*[65] used the Biot model to get the bottom attenuation for TL data/model comparison at site 18; the sound speed, however, was not calculated from the Biot model. The seabottom at this site is medium-to-coarse sand with a porosity of 0.38. (iii) Two measurements at the SAX99 site with a similar sediment (porosity=0.385) resulted in average speed ratios of 1.055 ± 0.010 at 125 Hz and 1.049 ± 0.033 at 400 Hz.[19] Direct measurements at the SAX04 site with a similar sandy bottom resulted in the average speed ratio of 1.046–1.068 for 600, 800, and 1000 Hz.[66] Changing the speed ratio from 1.126 into 1.06–1.08 (upper values shown in Fig. 8) and keeping TL and normal-mode attenuation rates unchanged [cf. Eqs. (21) and (22)], the broadband TL data/model comparison at site 18 may result in a normalized bottom attenuation $\alpha = (0.740-1.115) f^{1.76}$ dB/m in the 100–600 Hz range.

N. Bottom attenuation inverted at the New England Bight

Potty *et al.*[67] used long-range sediment tomography techniques to invert for sound speed and attenuation at a location in the shelf-break region in the Middle Atlantic Bight south of New England (designated site 19 in this paper). The data were obtained from the Shelf Break Primer Experiment, conducted in August 1996. The sound speed in the bottom was estimated using a hybrid inversion scheme based on the dispersion behavior of broadband acoustic propagation (with 0.82 kg explosive charges). This inversion scheme is a combination of a genetic algorithm and the Levenberg–Marquardt method. They reported that the sound speed in the bottom had a strong positive gradient (see Fig.

TABLE IX. Bottom attenuation at the SW'06 site on New Jersey shelf (site 17).

f (Hz)	53	103	203	253	303	403	503	703	953	1153	1503	2003
α (dB/m)	0.002 78	0.007 52	0.025 51	0.0379	0.0466	0.0779	0.1596	0.3977	0.8253	0.9559	1.3759	1.7470

19 in Ref. 67). In the top 5 m layer of the sediment the average sound speed was 1590.8 m/s (speed ratio=1.073). The sound attenuation values in the sediments were estimated based on modal amplitude ratios, received at 40 km away from the source. They were 0.000 98, 0.0012, 0.001 49, and 0.0019 dB/m for 30, 40, 50, and 60 Hz, respectively. The frequency band might be too narrow to offer a frequency dependence of the bottom attenuation. However, the bottom attenuation magnitudes at site 19 match the attenuation values obtained from sites 1–17, and much smaller than that extrapolated from Hamilton empirical model with an assumption of linear frequency dependence.

O. Sound attenuation in the sediments at SWARM 95 site off the New Jersey coast

Applying MFP-based optimal and Bayesian inversion techniques to air gun data collected from the Shallow Water Acoustics in the Random Media (SWARM'95) experiment, Jiang *et al.*[68] inverted the geoacoustics parameters at the SWARM'95 site on the New Jersey shelf (designated site 20 in this paper), including sound speed and attenuation in the sediments. The frequencies spanned two LF bands: 35–90 and 120–180 Hz, centered at 65 and 150 Hz, respectively. Sediment *p*-wave attenuation was given in the form of a maximum posterior estimate and its 95% credibility interval at each frequency band. The inverted sound attenuations in the sediments (excluding shear wave effects) are 0.005 36 dB/m for 65 Hz and 0.024 98 dB/m for 150 Hz (averaged from Tables V and VII in Ref. 68). A little higher value of the constant coefficient might partly be caused by a higher sound speed ratio of 1.14 at the water-sediment interface, i.e., by the speed-attenuation coupling discussed in Sec. II C 3.

III. SUMMARY OF LF ATTENUATIONS AND SOUND SPEEDS IN SANDY SEABOTTOMS

A. LF sound attenuation

The LF-field-derived sound attenuations (all are unnormalized original data from above-discussed papers) in sandy bottoms, obtained from 20 locations in differential coastal zones around the world, are presented in Fig. 7. Although all the inversion methods unavoidably involved their own uncertainties, the effective attenuation data derived from different LF characteristics at the 20 locations with sandy bottoms exhibit similar nonlinear frequency dependence. The majority of the attenuation amplitudes from different locations (except site 18) are so similar that many data symbols for some locations cannot be seen in the figure, because they cover each other. For comparison, the two f^1 frequency dependences shown in this figure (black dotted lines) are based on Hamilton's empirical formula:[3,4] $\alpha=0.679f^1$ dB/m and $\alpha=0.138f^1$ dB/m (f in kHz). Hamilton[3,4] classified the sediments in continental zones into nine types. According to Hamilton's empirical prediction,[3,4] the sound attenuation for most of fast sediments in continental zones (types 1–7, i.e., from coarse sand to sandy-clay bottom) should fall between these two f^1 lines. The LF derived attenuation data do match Hamilton's prediction[3,4] very well around 1 kHz. However, the LF bottom attenuation, derived from different data and analysis techniques, exhibits an apparent nonlinear frequency dependency in a frequency range of 50–2000 Hz. Also shown in Fig. 7 is frequency dependence that varies as $f^{1.80}$ (black solid line).

A currently available directly-measured bottom attenuation data set below 2 kHz, obtained by Turgut and Yamamoto[16] from the cross-hole tomography experiment in saturated beach sand, is also plotted in Fig. 7. It is readily

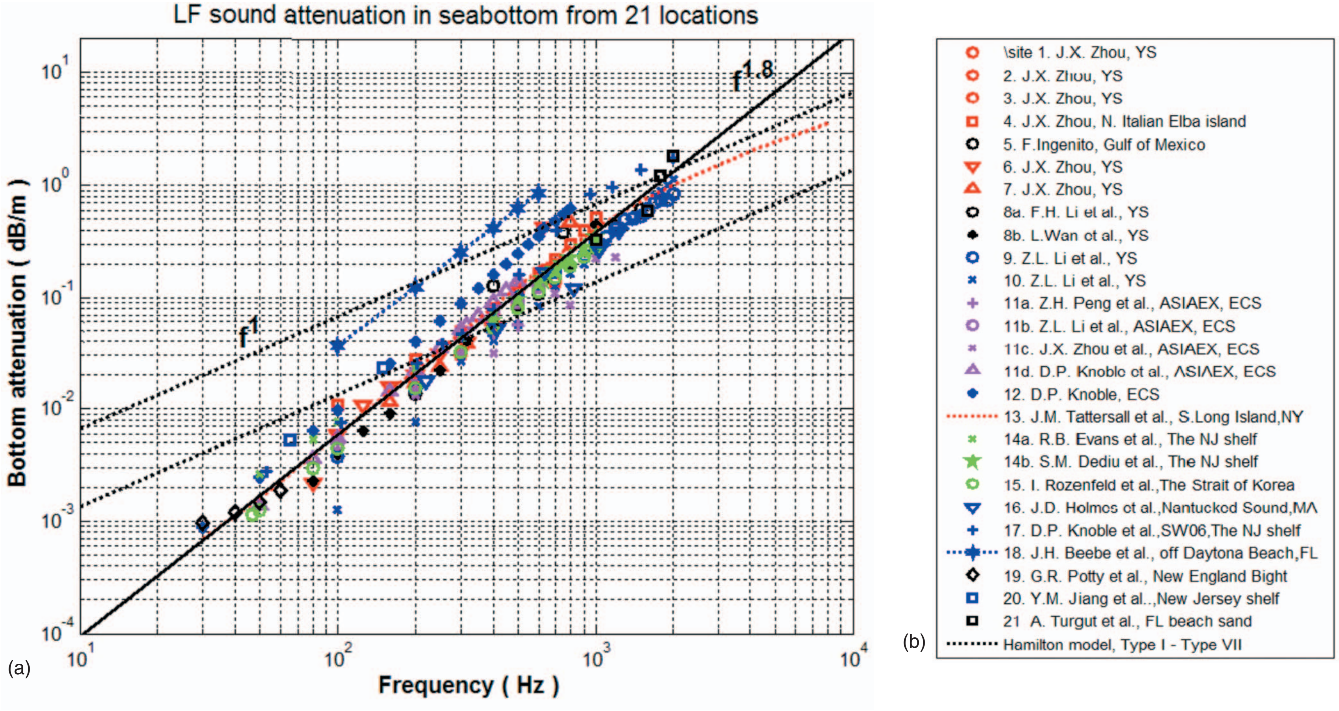


FIG. 7. (a) and (b) LF sound attenuation vs frequency in sandy bottoms from 21 sites.

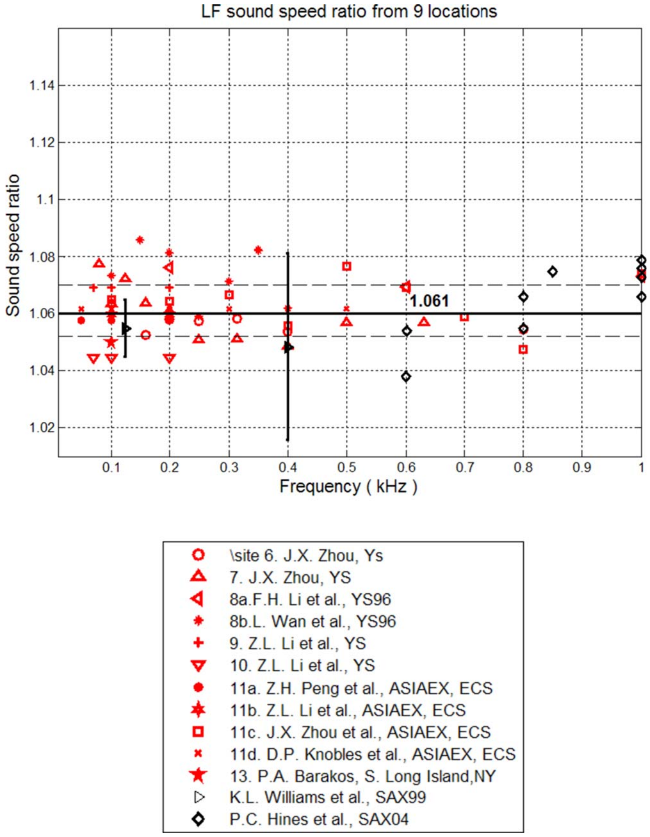


FIG. 8. (Color online) LF sound speed ratio vs frequency in sandy bottoms from nine sites.

apparent that all the LF bottom attenuation data shown in Fig. 7, obtained from 21 locations, exhibit strong nonlinear frequency dependence.

B. LF sound speed ratio

Among sites 1–20, there were 7 sites where the sound speeds in the seabed were obtained from independent LF inversion methods by 11 research groups (and using a half-space bottom model). All the data (53 data points) were obtained from sites 6–11 in the Yellow Sea and the East China Sea, except one at 100 Hz from New York Bight, south of Long Island (site 13). The sound speed ratios are plotted in Fig. 8 by red symbols. Because the sound speed in the top layer of sediment may vary with the seasonal variations of temperature in the water column near the bottom,[69,70] it is more appropriate to use sound speed ratio, instead of the bottom sound speed itself.

Figure 8 shows that the sound speed ratios at the water-sediment interface, inverted from the LF measurements, are 1.04–1.08. The mean value, averaged over 53 data points in a frequency range of 100−1000 Hz, is 1.061 with a standard deviation of ±0.009. (The mean value and STD averaged over 47 data points in a frequency range of 50−600 Hz are the same.) However, these LF measurements do not exhibit sound speed dispersion. Sound speed dispersion for $f < 1$ kHz may not have been observed because (1) most of the 50 data points are in 50−600 Hz range (47 points), (2) the Biot model shows that the speed dispersion in sandy bottoms in 50−600 Hz range is very weak or does not exist (see Figs. 10 and 12), and (3) speed dispersion in sandy bottoms below 1 kHz is too weak to be precisely inverted by those inversion methods; in general, the higher the frequency, the larger the error of geoacoustic inversion in SW will be, i.e., the sound speed inversion may be masked by uncertainties in geophysical parameters (especially water depth and SSP) for $f > 600$ Hz.

Williams *et al.*[19] reported the LF sound speed ratios at 125 and 400 Hz, obtained from the SAX99 site with a sediment porosity of 0.385. Hines *et al.*[66] reported the speed ratios in the bottoms at the SAX04 site (that was near the SAX99 site), obtained from direct time-of-flight measurements along all three Cartesian axes at 600, 800, and 1000 Hz. For comparison, the LF speed ratio data from the SAX99 and SAX04 are also plotted in Fig. 8 by black symbols. The speed ratio below 1000 Hz, averaged over the LF-field-inversion data (53 points), the SAX99 data (2 points), and the SAX06 data (9 points), is around 1.062±0.010.

IV. COMPARISON OF LF RESULTS WITH THE BIOT–STOLL MODEL

Following Hamilton's definition for the seafloor *geoacoustic model*,[1–3] and using the LF measurements shown in Sec. III, this section discusses extrapolated and predicted values of the Biot parameters in the effective medium of sandy bottoms that are important for sound transmission and sediment acoustic modeling.

Several representative data sets for the Biot geophysical parameters in sandy bottoms have been published and are listed in Table X. Each data set is referred by the authors as initials. The Biot parameters in the column labeled TCCD were used by Tattersall *et al.*[56] to match the many years of measurements of both broadband and narrowband TLs in the New York Bight, south of Long Island. (The bottom consists of medium grain sized sand.) The parameters in the column labeled TY were used by Turgut and Yamamoto[16] to match their *in situ* cross-hole measurements on sound speed and attenuation in homogeneous, unconsolidated sands in the 1−30 kHz band. In the column labeled WJTTS in Table X are the parameters reported by Williams *et al.*[19] from the Sediment Acoustics Experiment of 1999 (SAX99) in the Gulf of Mexico, where the Biot parameters of the sediments were extensively measured by using both traditional and newly developed methods. The column labeled "Historical" summarizes the range of Biot parameters by Stoll for sandy sediments.[71] The pore size is not listed in several columns in Table X. In such cases, it is calculated by the relationship

$$a = \sqrt{\frac{8\alpha_0\kappa_s}{\beta}}. \qquad (38)$$

In the basic Biot theory, it is largely three parameters that control the fluid motion that results in the nonlinear dispersion that is important in the coarser granular sediments. These are the permeability κ_s, the porosity β, and the tortuosity α_0, which define the necessary amount of "added mass" in the Biot–Stoll model. These three parameters define the pore-size parameter by Eq. (38). Data/model comparisons (trials) tell us that the LF sound speed ratio is most

TABLE X. Input parameters to Biot model.

Symbols	Units	TCCD	TY	Historical	WJTTS	LF fit	BB fit	
							(A)	(B)
				Bulk properties				
(1) Porosity, β		0.43	0.44	0.36–0.47	0.385(0.415)	0.42	0.39	0.45
(2) Grain density, ρ_s	$(\mathrm{kg/m^3})$	2650	2650	2650	2690	2690	2690	2690
(3) Fluid density, ρ_f	$(\mathrm{kg/m^3})$	1024	1000	1000	1023	1023	1023	1023
(4) Grain bulk modulus, K_s	(Pa)	3.6×10^{10}	3.6×10^{10}	$(3.6–4.0)\times10^{10}$	3.2×10^{10}	3.2×10^{10}	3.2×10^{10}	3.2×10^{10}
(5) Fluid bulk modulus, K_f	(Pa)	2.38×10^{10}	2.3×10^{9}	$(2.0–2.3)\times10^{9}$	2.395×10^{9}	2.395×10^{9}	2.395×10^{9}	2.395×10^{9}
				Fluid motion				
(6) Viscosity, η	kg/m s	0.001 01	0.001	0.001	0.001 05	0.001 05	0.001 05	0.001 05
(7) Permeability, κ_s	$(\mathrm{m^2})$	1.25×10^{-11}	1.75×10^{-11}	$(0.65–10.0)\times10^{-11}$	2.50×10^{-11}	1.0×10^{-11}	2.3×10^{-11}	0.5×10^{-11}
(8) Pore size, a	(m)	3.0×10^{-5}	5.2×10^{-5}	$(1.00–5.00)\times10^{-5}$				
(9) Tortuosity of sediment, α_0		1.75	1.25	1.00–1.25	1.35 (1.12)	1.35	1.25	1.45
				Frame response				
(10) Frame shear modulus, $\mu0$	(Pa)	5.0×10^{7}	2.4×10^{7}	$(2.61–11.9)\times10^{7}$	2.92×10^{7}	2.92×10^{7}	2.92×10^{7}	2.92×10^{7}
(11) Shear log decrement, δ_s		0.02	0.0477	0.00–0.15	0.0616	0.0616	0.0616	0.0616
(12) Frame bulk modulus, K_0	(Pa)	5.0×10^{7}	5.2×10^{7}	$(4.36–43.6)\times10^{7}$	4.36×10^{7}	4.36×10^{7}	4.23×10^{7}	4.36×10^{7}
(13) Bulk log decrement, δ_b		0.02	0.0477	0.00–0.15	0.0477	0.0477	0.0477	0.0477

sensitive to the porosity of sediment and is less sensitive to the permeability and tortuosity. The LF sound attenuation is most sensitive to the permeability of the sediment and is less sensitive to the porosity and tortuosity. Thus, the porosity of 0.42 and the permeability of 1.0×10^{-11} were determined as the values that provide the best match between the Biot model and the LF inverted attenuation and speed ratio shown in Figs. 7 and 8. The other 11 Biot parameters are same as those from the SAX99 measurement.[19]

The Biot parameters in the column labeled "LF fit" in Table X produce the Biot model results for the sound attenuation and speed ratio as a function of frequency shown in Figs. 9 and 10, respectively. The data/model comparisons show that the Biot model matches LF sound attenuation in sandy bottoms in a frequency range of 50–2000 Hz. Neither the sound speed data nor the predictions of the Biot model predict sound speed dispersion at lower frequencies. All the Biot parameters, labeled "LF fit" in Table X for our modeling, are derived from the LF-field inversion and the SAX99 measurements. These parameters are consistent with either theoretical considerations or historical experimental measurements. Therefore, the sound attenuation and speed in the effective medium equivalent to sandy bottoms, inverted from LF SW acoustics measurements, may imply a solid physically based model for the actual medium. The average LF sound speed and attenuation, obtained from 20 locations, are in good agreement with the physics-based Biot model. Thus, they naturally satisfy the Kramers–Kronig relationship between the frequency dependence of the sound speed and attenuation.

For more clarity, the averaged LF-field-inverted sound attenuation from sites 1–20 is listed in Table XI as a function of frequency. The numbers of available data sets for the av-

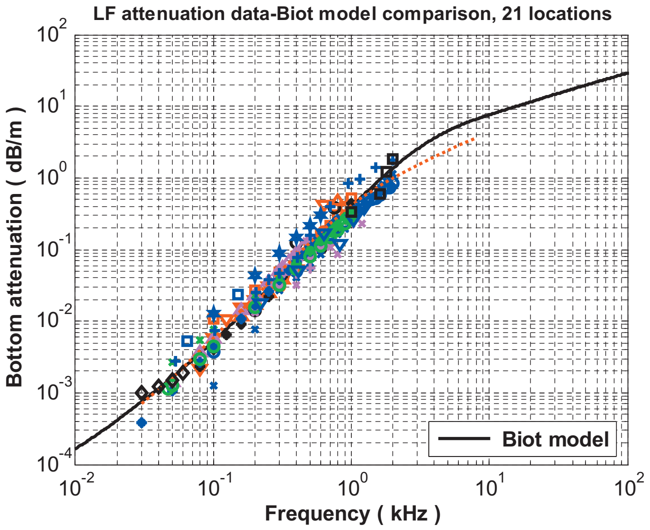


FIG. 9. Comparison between the LF bottom attenuation data (21 locations) and predictions based on the Biot theory.

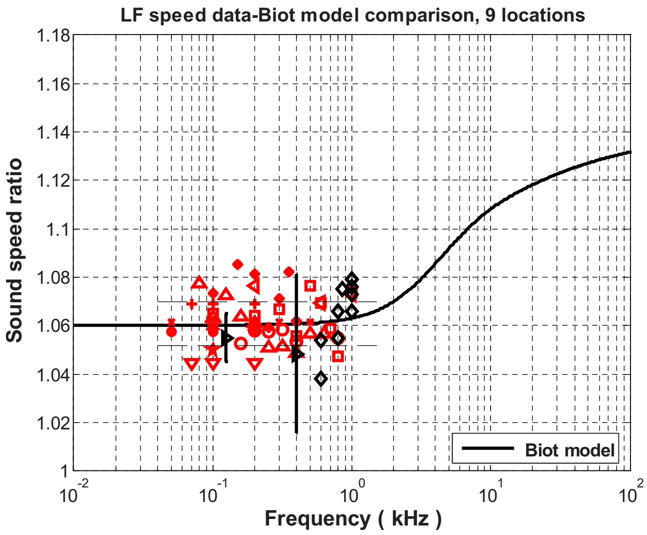


FIG. 10. (Color online) Comparison between the LF sound speed ratio data (nine locations) and predictions based on the Biot theory.

TABLE XI. The average sound attenuation in sandy seabottoms.

f (Hz)	50	80	100	125	160	200	250	300	315	400	500
α (dB/m)	0.001 75	0.003 30	0.005 98	0.008 78	0.015 10	0.019 78	0.029 50	0.042 71	0.043 68	0.070 12	0.109 26
Data sets	(7)	(7)	(15)	(5)	(10)	(19)	(9)	(15)	(8)	(19)	(19)
f (Hz)	600	630	700	750	800	900	1000	1200	1500	1600	2000
α (dB/m)	0.145 15	0.184 79	0.188 88	0.236 60	0.223 65	0.275 11	0.365 04	0.471 93	0.754 64	0.672 10	1.177 68
Data Sets	(12)	(9)	(12)	(6)	(12)	(7)	(13)	(8)	(5)	(3)	(4)

erage are also listed. Using the power-law fitting, the average attenuation in sandy seabottoms can approximately be expressed as the following nonlinear frequency dependence:

$$\alpha = (0.37 \pm 0.01) f^{(1.80\pm0.02)} \text{ dB/m} \quad \text{for } 50-1000 \text{ Hz}$$

or

$$\alpha = (0.35 \pm 0.01) f^{(1.78\pm0.02)} \text{ dB/m} \quad \text{for } 50-2000 \text{ Hz}, \tag{39}$$

where f is frequency in units of kHz.

The sound attenuation in the sandy bottoms, averaged over sites 1–20, as a function of frequency is shown in Fig. 11 by circles. Theoretical prediction by the Biot model with the parameters listed in a column labeled “LF fit” in Table X is also shown in Fig. 11 by a solid line. They are in very good agreement. It is interesting to note that these Biot parameters are very close to TCCD parameters listed in Table X that successfully predicted many years' measurements of TL for both broadband and narrowband in the New York Bight, south of Long Island[56] (also see Sec. II H).

For the data-model comparisons, in this section we have used the normalized attenuation for sites 12 and 18. If the bottom attenuations at sites 12 and 18 are not normalized with a speed ratio of 1.06, the resultant bottom attenuations, averaged over 20 locations, are plotted in Fig. 11 by x-markers. The power-law fitting to x-markers will result in $\alpha = (0.42 \pm 0.02) f^{(1.80\pm0.04)}$ dB/m for 50–1000 Hz and $\alpha = (0.38 \pm 0.02) f^{(1.74\pm0.03)}$ dB/m for 50–2000 Hz, where f is in units of kHz. To match the un-normalized data shown in Fig. 11, a small change in the permeability from 1.0×10^{-11} to 1.1×10^{-11} is sufficient. That is, with or without the normalization of seabottom attenuation for sites 12 and 18, the conclusions are nominally identical.

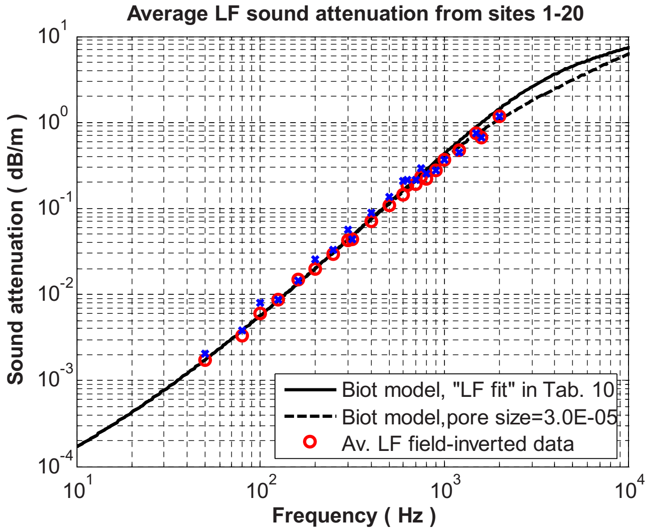


FIG. 11. (Color online) Comparison of the average LF inverted sound attenuation, obtained from sites 1–20, with the Biot model.

It is interesting to note that above 1000 Hz, the average LF inverted attenuations are slightly smaller than the Biot model prediction. This can be explained by the speed-attenuation coupling described in Sec. II C 3. In general, most of the LF inversion methods in this paper derived bottom sound speeds in the 50–500 Hz range. The Biot model predicts that the sound speed should increase with increasing frequency around 1 kHz. Applying a (lower) sound speed obtained from LFs to HF inversion may result in lower values of sound attenuation in the bottom. Thus, the field-inverted attenuation is less reliable for frequencies above 1000 Hz. One may increase the tortuosity or pore size listed in a column of “LF fit” to a higher value to obtain a perfect match with all the data shown Fig. 11, including data above 1000 Hz. As an example, using a pore size of 3.0×10^{-5} in the Biot model instead of Eq. (38) will yield a theoretical curve as shown in Fig. 11 by the dashed line. However, increasing the pore size or tortuosity needs to be justified (constrained) by quality sound speed dispersion data in the LF to HF speed transition band, because the Biot model with a larger pore size or tortuosity predicts a smaller sound speed dispersion.

V. A REFERENCE BROADBAND DATA SET FOR SANDY BOTTOMS IN A FREQUENCY RANGE OF 50–400 000 Hz

A. The LF data are consistent with the SAX99 and other mid-to high-frequency measurements

In the fall of 1999, a comprehensive field experiment named “SAX99” (Sediment Acoustics Experiments) was conducted in the Gulf of Mexico near Fort Walton Beach, FL (referred to as SAX99 site).[19,72,73] The top 1–2 m of the sediment at the SAX99 site was composed of a coarse-to-medium sand containing numerous small shell fragments. During the SAX99 experiment, the Biot parameters of the sediments were extensively measured by using both traditional and newly developed methods. The sound speed over a frequency range of 125 Hz–400 kHz and attenuation over a frequency range of 2.6–400 kHz were measured by scientists from different institutions. The data are unique both for the frequency range spanned at a common location and for the extensive environmental characterization. Williams *et al.*[19] summarized the results on sound speed and attenuation measurements, discussed possible effects of the uncertainties of measured environmental parameters on sound speed and

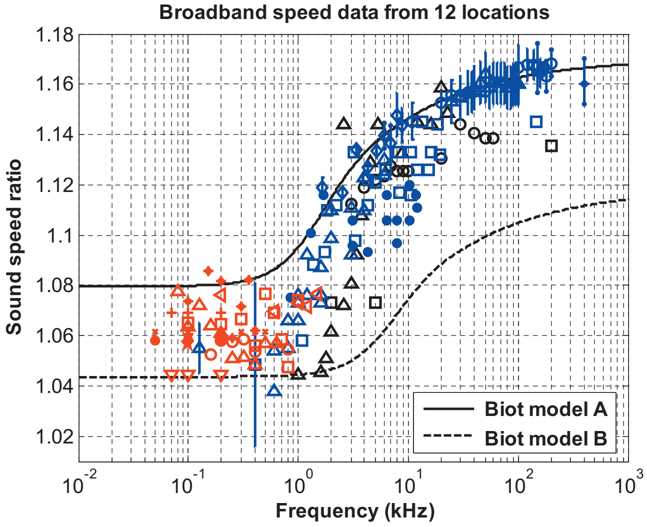


FIG. 12. Broadband sound speed ratio in sandy seabottoms, and data/model comparison. Blue with and without uncertainty bars: SAX99 data (Ref. 19) and SAX04 (Ref. 66); other colors: LF-field-derived data from seven locations of this paper; black: from mid-frequency to HF direct measurements (see text for details). The Biot parameters for models *A* and model *B* are listed in Table X.

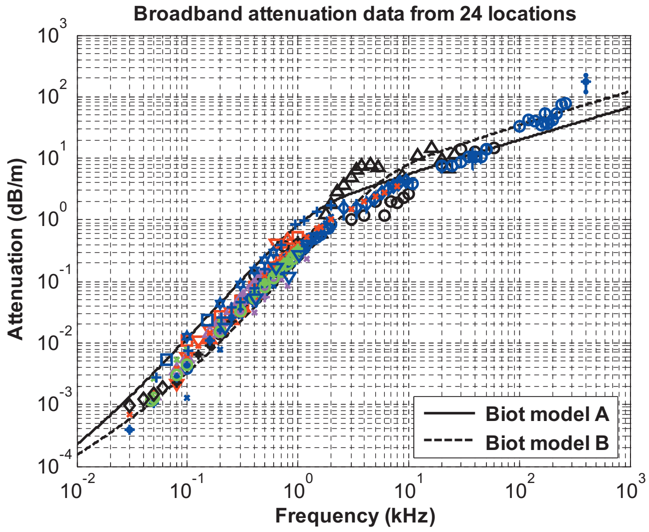


FIG. 13. Broadband sound attenuation in sandy seabottoms, and data/model comparison. Blue with uncertainty bars: SAX99 data (Ref. 19); other colors: LF-field-inverted data from 20 locations; black: from mid-frequency to HF direct measurements (see text for details). The Biot parameters for models *A* and *B* are listed in Table X.

attenuation, and made a comparison between the experimental data and predictions based on the Biot theory of porous media. The data in Figs. 3 and 4 in Ref. 19, obtained by different groups with different methods, will be used as a benchmark data set for sound attenuation and speed ratio in the mid-frequency to HF band. This paper will compare the LF data from sites 1–20 with the SAX99 results. The main purpose of doing this is to (1) validate the LF-field-derived seabed sound speed and attenuation; (2) see if we can extrapolate and predict the values of LF bottom sound speed and attenuation from mid-frequency to HF measurements by the Biot model; and (3) offer a reference broadband data set of sound speed and attenuation for sandy bottoms, either for practical field prediction in SW or for theoretical research on sediment acoustics.

The SAX99 data are shown in Figs. 12 and 13 by blue symbols with uncertainty bars. The LF-field-derived data are plotted by other color symbols. Figure 12 shows that the sound speed ratios at 125 and 400 Hz, obtained from the SAX99 measurement, are within the range of the LF data. Figure 13 shows that the LF sound attenuation in sandy bottoms, obtained from sites 1–20, and the SAX99 data are smoothly joined. For comparison, the speed ratio and attenuation measurements from the cross-hole tomography experiment of Turgut and Yamamoto[16] are shown in Figs. 12 and 13 by black up-triangles. The Biot parameters at the cross-hole experiment site are listed in Table X. The sound speed ratio at the water-sediment interface, obtained near Marciana Marina on the north side of Elba Island, Italy by Maguer *et al.*[17] using a parametric source and a buried hydrophone array, is shown in Fig. 12 by black squares. The sediments at this site were classified as medium sand with a permeability of 1.7×10^{-11} m^2 and a porosity of 45.8%. The measurements of Simpson *et al.*[74] using a buried vertical synthetic array in St. Andrews Bay, FL are shown in Figs. 12 and 13 by black circles. The sound speed ratio at the SAX04 site, obtained by Hines *et al.*[66] using direct time-of-flight measurements along all three Cartesian axes, are also indicated in Fig. 12 with blue symbols. These HF measurement sites have sandy bottom properties that are similar to those at the LF-field measurement sites.

B. Data-model comparison for the broadband data

The sound speed ratios and attenuations, calculated from two sets of the Biot parameters labeled "BB fit" in Table X, are shown in Figs. 12 and 13. All of the broadband data for sound speed ratio and attenuation are covered by (or close to) these two theoretical curves. However, using the broadband data, data-model comparisons show that, with one set of adjustable input parameters, the Biot model may perfectly match either the broadband speed dispersion or broadband attenuation, but not both. That is, there seems to be a problem if one tries to extrapolate mid-frequency to HF results for LF-field applications by using the Biot model (or the Hamilton model).

C. Future LF geoacoustics inversions at a shallower site are critical

The data-model comparisons have shown that both the effective LF sound speed and attenuation reported in this paper can be described well by the Biot–Stoll model.[5–11] (Comparisons to the Buckingham VGS model[14] or the Chotiros BICSQS model[15] can equally well be made using the LF data presented in this paper.) The Biot model also reasonably predicts the measurement variation for broadband speed dispersion and the frequency dependence of sound attenuation in sandy bottoms (see Figs. 9–13 for the data-model comparisons). The LF-field-inverted data satisfy the Kramers–Kronig relationship. However, the LF and HF combined broadband data do not. Comparing with the broadband data shown in Figs. 12 and 13, these models either underestimate dispersion or overestimate LF sound attenuation. Possible explanations include the following: (1) The LF data and the mid-frequency to HF data were not collected from the same site. Sediments at many LF inversion sites are fine sand or sand-silt mixture. The sediments at the SAX99, SAX04, and other HF experiment sites are coarser sand. Thus, using a unique set of the geophysical parameters to fit both mid-frequency to HF data and the LF-field-derived data might not be appropriate in general. (2) Sound propagation in sediments might need different physical models for the LF and HF regions. (3) Some of the Biot parameters might be dependent on frequency. (4) The sound speed dispersion in most of the sandy and sandy mixture bottoms might not be as strong as the SAX99/SAX04 (or TY) group reported. (5) A key problem is that we lack convincing experimental data, obtained from one site in a frequency band that covers a portion of the LF to HF speed/attenuation transition regions to firmly support these geoacoustic models. The SAX99 group had a difficulty obtaining attenuation data below 2600 Hz; the LF-field inversions did not exhibit speed dispersion. The LF inverted sound speed and attenuation data were less reliable with increasing frequency, as discussed below.

It is well known that for resolving sound propagation problems in SW, one may add a "hidden depth" ΔH to a real water depth H, then use a pressure-release boundary to replace a real seabottom.[27,75,76] In such a case, the effective water depth can be expressed by

$$H_{\mathrm{eff}} = H + \Delta H = H + \frac{P}{4\pi}\lambda = H + \frac{(\rho_b/\rho_w)}{[1-(c_w/c_b)^2]^{1/2}}\frac{\lambda}{2\pi}, \tag{40}$$

where c_b and c_w are sound speed in the bottom and water, respectively, and λ is the acoustic wavelength in the water. (P is the phase of equivalent seabed reflection coefficient.) If we take ρ_b/ρ_w=1.8, and let c_b/c_w vary in a range 1.03–1.12 (below 3 kHz), for most of sandy bottoms in SW, the hidden depth ΔH will vary only about half of the wavelength between 1.20λ and 0.64λ. Possible shear waves in sediments only modifies ΔH by a factor of $[1-2(c_s/c_w)^2]^2$, where c_s is the shear wave speed.[75,76] For most of sediments, the effect of the shear wave is negligible.

In the authors' opinion, almost all of seabed geoacoustic inversion methods are sensitive to the modal eigenvalues used to invert for values of the sound speed in the bottom (MFP, Hankel transform, mode dispersion, signal interference, etc.). The modal eigenvalues are mainly determined by the sound speed profile in the water column and the effective water depth (H_{eff}=H+ΔH). The geoacoustic inversion for sound speed in the bottom is to measure (determine) small variations of the modal eigenvalues that are caused by different hidden depth ΔH (i.e., different sound speed ratios). If a real water depth H is much larger than a hidden depth ΔH, the modal eigenvalues will be less sensitive to the variation of ΔH (sound speed in the bottom). The shallower the water depth (a smaller H) or the lower the frequency, the more reliable the geoacoustic inversion results. Besides, if a real depth H is smaller, for a given frequency (wavelength), there will be fewer modes (to be processed). In the wave number domain, wave numbers will have relatively larger separations, and in the time domain, arrival time differences for different modes will be relatively larger. In such a case, the eigenvalues (and decay rates) of normal modes will be more sensitive to the variation of effective water depth, i.e., more sensitive to the sound speed (attenuation) in the bottom. Unfortunately, most geoacoustic inversion experiments have been conducted in areas with relative larger water depths, which only allow one to successfully invert for sound speed and attenuation in the bottom for the lower frequencies.

Based on above discussions, it will be desirable to conduct a LF geoacoustic inversion experiment at a shallower area (such as the SAX99 site) where the water depth is less than 20 m. At a shallower area, experimenters may extensively measure the Biot parameters of the sediments as well as the sound speed and attenuation in the bottom at mid-frequencies to HFs. The LF inversion group may offer more reliable data for sound speed and attenuation in a frequency band that covers a portion of the LF to HF speed/attenuation transition regions that will be very critical to sediment acoustic modeling.[20] Such high-fidelity data might shed more light on the decades-lasting debate on seabed geoacoustics models.

VI. SUMMARY AND DISCUSSION

This paper has analyzed and reviewed SW field measurements at LFs that have revealed a nonlinear frequency dependence of sound attenuation in the effective medium equivalent to sandy and sandy mixture seabottoms. The measurements were conducted at 20 locations in different coastal zones around the world.

The sound speed and attenuation in the effective seabottom medium, inverted from the LF-field measurements with different methods, have been analyzed and summarized in a frequency range of 50–2000 Hz. The sound speed ratios at the water-sediment interface, inverted from the LF-field measurements at seven locations, are in 1.04–1.08 range. The averaged value is 1.061 ± 0.009 (see Fig. 8). Although all the inversion methods unavoidably involved their own uncertainties, the sound attenuations in sandy bottoms exhibit similar magnitudes and a similar nonlinear frequency dependence (see Figs. 7, 9, and 11). It can be expressed by $\alpha = (0.37 \pm 0.01) f^{(1.80 \pm 0.02)}$ dB/m in 50–1000 Hz range, where f is in units of kHz.

Both the LF-field-derived effective sound speed and attenuation in the sandy bottoms can be well described by the Biot–Stoll model with parameter values that are consistent with either theoretical considerations or historical experimental measurements for sandy seabed. (See Table X and Figs. 9–11 for the LF data comparison with the Biot model.) Comparisons to the Buckingham VGS model[14] or the Chotiros BICSQS model[15] can equally well be made using the LF data presented in this paper.

The possible speed-attenuation coupling problem in seabottom geoacoustic inversions was addressed. It shows that, if acoustic energy losses such as the incoherent TL or modal decay law are used to derive the sound attenuation in the seabed, one should first have an accurate value of the sound speed in the seabed (and density), independently derived from LF measurements, as a constraint. A possible error in dispersion-based inversion made at one distance, caused by the filter phase shift, is also discussed in the paper.

The LF effective sound speeds and attenuations of the sandy bottoms are smoothly joined with the SAX99/04 benchmark data for mid-frequencies and HFs and also match Hamilton's prediction[1–3] on sound attenuation very well around 1 kHz. A combination of the LF data with the SAX99/04 data as well as other mid-frequency to HF measurements offers a reference broadband data set of sound speed and attenuation for sandy bottoms in SW in a frequency range of 50–400 000 Hz. The resultant broadband data on sound speed and attenuation are within or close to theoretical curves calculated by two sets of parameters of the Biot–Stoll model. Numerical studies reveal that these models with one set of adjustable input parameters can perfectly match either the broadband speed dispersion or attenuation, but not both. If both the inverted LF data and the SAX data are deemed reliable, these models seem to underestimate the broadband sound speed dispersion or overestimate the LF sound attenuation of sandy bottoms. Possible reasons have been discussed.

In this paper the total sound TL in SW was assumed to be due to "effective" p-wave attenuation in seabottom, including the intrinsic absorption in the seabed material, sound scattering due to sediment spatial inhomogeneities, and so on. The possible shear wave effect in sandy sediments was assumed to be unimportant. Possible scattering losses due to sea surface or bottom roughness, biological life, or internal waves within the water column were not taken into account. Fortunately, most of the LF inversion measurements presented in this paper were made under downward refracting conditions in the water column with negligible sea surface effects for long-range sound propagation. The possible biological and internal wave effects with frequency selectivities can often be separated from normal sound propagation. Possible speed/attenuation gradient profiles in the bottom may change apparent frequency dependence of the LF-field-derived sound speed and attenuation values in the seabottom.[35,77] Most of the inverted LF data for sound speed and attenuation were obtained from long-range acoustic field data for which the surficial sediment layer with a thickness on the order of a few wavelengths plays the dominant role. Thus, it would be better to interpret the bottom sound speed and attenuation values as well as the Biot parameter values presented in this paper as "effective/equivalent" or "averaged" values in the effective medium of the top layer of the sandy bottoms. Chapman[78] indicated that in the case of geoacoustic inversion, the inversion techniques are divided roughly into two types: (1) those which estimate the geophysical properties of the seabed as precisely as our models will allow, and (2) those which estimate parameters of an effective seabed model that is adequate for predicting the acoustic field in the ocean. Inversions of type 1 are necessary to construct a "true" picture of the seabed composition (accounting for layering, gradients, etc.) while inversions of type 2 may turn out to be more useful for sonar performance prediction models. The inverted LF *effective* sound speed and attenuation presented in this paper may offer some reference data for predicting the sonar performance at long range in SW and for investigating the effective mechanism of LF sound interactions with the bottom.

Uncertainties of the LF-field-inverted acoustic sound speed and attenuation from different sites depend on their own measurements and methodologies. Hopefully, those uncertainties are close to random and follow a normal distribution of zero mean; an averaging over many measurements (at different locations) and many methods may smooth out or decrease some uncertainties. If so, the resultant LF-field-inverted sound speed and attenuation from many locations can reveal/infer some physics of sound propagation in marine sediments. The LF sound speed and attenuation in sandy bottoms, reported in this paper, are observed only through a "window" of the LF sound propagation in SW. This may or may not represent the actual physics of sound propagation within marine sediments.

As to the decades-long debate on the speed dispersion and the frequency dependence of sound attenuation in sediments, the question is still open. For a critical test, it is desirable to have quality data on sound speed and attenuation in sediments from one site for a frequency band that covers a portion of the LF to HF speed/attenuation transition regions. Thus, a LF inversion experiment together with mid- to high direct measurements at a shallower area ($H \sim 20$ m) is proposed.

ACKNOWLEDGMENTS

This work was supported by the ONR. The authors would like to express their great appreciation to Dr. Ellen Livingston and Dr. Benjamin Reeder for their support and helpful comments. They sincerely thank Dr. Kevin Williams who offered the SAX99 original data and plots, read the draft manuscript, and participated in several conversations about this research. J.-X. Z. and X.-Z. Z. appreciate their colleagues, Dr. Peter Rogers and James Martin, for their assistance and encouragement. Dr. Matthew Tattersall, Dr. Zhenglin Li, Dr. Fenghua Li, Dr. Zhaohui Peng, Mr. Lin Wan, Dr. Jason Holmes, Dr. Yongmin Jiang, and Dr. Gopu Potty are much appreciated for kindly offering the original data of seabed acoustic parameters at their experimental sites. Thanks go to Dr. Nicholas Chotiros, Dr. Allan Pierce, Dr. William Carey, Dr. Michael Buckingham, and Dr. Altan Turgut for helpful discussions on seabed geoacoustic models. Finally, the authors thank the associate editor and the two anonymous reviewers for their valuable suggestions on the earlier version of this manuscript.

[1] E. L. Hamilton, "Compression-wave attenuation in marine sediments," Geophysics **37**, 620–646 (1972).

[2] E. L. Hamilton, "Sound attenuation as a function of depth in the sea floor, Compression-wave attenuation in marine sediments," J. Acoust. Soc. Am. **59**, 528–535 (1976).

[3] E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

[4] E. L. Hamilton and R. T. Bachman, "Sound speed and related properties of marine sediments," J. Acoust. Soc. Am. **72**, 1891–1904 (1982).

[5] 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).

[6] 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).

[7] M. A. Biot, "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 1254–1264 (1962).

[8] R. D. Stoll, "Acoustic waves in ocean sediments," Geophysics **42**, 715–725 (1977).

[9] R. D. Stoll, "Theoretical aspects of sound transmission in sediments," J. Acoust. Soc. Am. **68**, 1341–1350 (1980).

[10] R. D. Stoll and T. K. Kan, "Reflection of acoustic waves at a water-sediment interface," J. Acoust. Soc. Am. **70**, 149–156 (1981).

[11] R. D. Stoll, "Marine sediment acoustics," J. Acoust. Soc. Am. **77**, 1789–1799 (1985).

[12] M. J. Buckingham, "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2796–2815 (2000).

[13] M. J. Buckingham, "Compressional and shear wave properties of marine sediments: Comparisons between theory and data," J. Acoust. Soc. Am. **117**, 137–152 (2005).

[14] M. J. Buckingham, "On pore-fluid viscosity and the wave properties of saturated granular materials including marine sediments," J. Acoust. Soc. Am. **122**, 1486–1501 (2007).

[15] N. P. Chotiros and M. J. Isakson, "A broadband model of sandy ocean sediments: Biot-Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am. **116**, 2011–2022 (2004).

[16] A. Turgut and T. Yamamoto, "Measurements of acoustic wave velocities and attenuation in marine sediments," J. Acoust. Soc. Am. **87**, 2376–2383 (1990).

[17] A. Maguer, E. Bovio, W. L. J. Fox, and H. Schmidt, "*In situ* estimation of sediment sound speed and critical angle," J. Acoust. Soc. Am. **108**, 987–996 (2000).

[18] R. D. Stoll, "Speed dispersion in water-saturated granular sediment," J. Acoust. Soc. Am. **111**, 785–793 (2002).

[19] K. L. Williams, D. R. Jackson, E. I. Thorsos, D. J. Tang, and S. G. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng. **27**, 413–428 (2002).

[20] M. J. Isakson and T. B. Neilsen, "The viability of reflection loss measurement inversion to predict broadband acoustic behavior," J. Acoust. Soc. Am. **120**, 135–144 (2006).

[21] J. L. Buchanan, "A comparison of broadband models for sand sediments," J. Acoust. Soc. Am. **120**, 3584–3598 (2006).

[22] A. C. Kibblewhite, "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," J. Acoust. Soc. Am. **86**, 716–738 (1989).

[23] J. X. Zhou and X. Z. Zhang, "Nonlinear frequency dependence of the effective sea bottom acoustic attenuation from low-frequency field measurements in shallow water (A)," J. Acoust. Soc. Am. **117**, 2494 (2005).

[24] J. X. Zhou and X. Z. Zhang, "Broadband sound speed and attenuation in sandy seabottoms in shallow water (A)," J. Acoust. Soc. Am. **119**, 3447 (2006).

[25] J. D. Holmes, W. M. Carey, S. M. Dediu, and W. L. Siegman, "Nonlinear frequency-dependence attenuation in sandy sediments," J. Acoust. Soc. Am. **121**, EL218–EL222 (2007).

[26] D. E. Weston, "Intensity-range relation in oceanographic acoustics," J. Sound Vib. **18**, 271–287 (1971).

[27] M. J. Buckingham, "Array gain of a broadside vertical line array in shallow water," J. Acoust. Soc. Am. **65**, 148–161 (1979).

[28] P. W. Smith, "The average impulse response of a shallow-water channel," J. Acoust. Soc. Am. **50**, 332–336 (1971).

[29] E. C. Shang, J. X. Zhou, and D. H. Guan, "Theoretical analysis of the boundary reflection loss and the parameter extraction at small grazing angles," Acta Acust. **6**, 308–313 (1981) [Chin. J. Acoust. **3**, 356–361 (1984)].

[30] E. C. Lo, J. X. Zhou, and E. C. Shang, "Normal mode filtering in shallow water," J. Acoust. Soc. Am. **74**, 1833–1836 (1983).

[31] E. C. Shang, X. Z. Zhang, and B. L. Ni, "Numerical sound field of normal modes in shallow water," Acta Oceanologica Sinica **7**, 212–224 (1985) (in Chinese).

[32] J. X. Zhou, "Inversion techniques for obtaining seabed low-frequency reflection loss at small grazing angles in shallow water," J. Acoust. Soc. Am. **79**, S68 (1986).

[33] J. X. Zhou, "Normal mode measurements and remote sensing of sea-bottom sound speed and attenuation in shallow water," J. Acoust. Soc. Am. **78**, 1003–1009 (1985).

[34] J. X. Zhou, X. Z. Zhang, and P. H. Rogers, "Effect of frequency dependence of sea-bottom attenuation on the optimum frequency for acoustic propagation in shallow water," J. Acoust. Soc. Am. **82**, 287–292 (1987).

[35] J. X. Zhou, X. Z. Zhang, P. H. Rogers, and J. Jarzynski, "Geoacoustic parameters in a stratified seabottom from shallow water acoustic propagation," J. Acoust. Soc. Am. **82**, 2068–2074 (1987).

[36] E. C. Shang, "Transition range of the average sound field in shallow water," Sci. Sin. **19**, 794–804 (1976).

[37] J. X. Zhou, "Vertical coherence of the sound field and boundary losses in shallow water," Acta Phys. Sin. **1**, 494–504 (1981).

[38] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960), Chap. 5.

[39] E. Murphy and O. V. Olesen, "Intensity-range relations for shallow-water sound propagation," J. Acoust. Soc. Am. **59**, 305–311 (1976).

[40] E. L. Murphy, A. Wasiljeff, and F. B. Jensen, "Frequency-dependent influence of the sea bottom on the near-surface sound field in shallow water," J. Acoust. Soc. Am. **59**, 839–845 (1976).

[41] F. Ingenito, "Measurements of mode attenuation coefficients in shallow water," J. Acoust. Soc. Am. **53**, 858–863 (1973).

[42] R. H. Ferris and W. A. Kuperman, "An experiment on acoustic reflection from the sea surface," NRL Report No. FR-7075 (1970).

[43] R. A. Koch, C. Penland, P. J. Vidmar, and K. E. Hawker, "On the calculation of normal mode group speed and attenuation," J. Acoust. Soc. Am. **73**, 820–825 (1983).

[44] F. H. Li and R. H. Zhang, "Bottom sound speed and attenuation inverted by using pulsed waveform and transmission loss," Acta Acust. **25**, 297–302 (2000).

[45] P. H. Rogers, J. X. Zhou, X. Z. Zhang, and F. H. Li, "Seabottom acoustic parameters from inversion of Yellow Sea experimental data," in *Experimental Acoustic Inversion Methods for Exploration of the Shallow Water Environment*, edited by A. Caiti, J.-P. Hermand, S. M. Jesus, and M. B. Porter (Kluwer, Dordrecht, 2000), pp. 219–234.

[46] L. Wan, J. X. Zhou, and P. H. Rogers, "Comparison of transmission loss data with predictions based on inverted seabed parameters from the Yellow Sea '96 experiment (A)," J. Acoust. Soc. Am. **119**, 3226 (2006).

[47] S. N. Wolf, D. K. Copper, and B. J. Orchard, "Environmentally adaptive signal processing in shallow water," Oceans'93 (1993), Vol. **1**, pp. 99–104.

[48] Z. L. Li and R. H. Zhang, "A broadband geoacoustic inversion scheme," Chin. Phys. Lett. **24**, 1100–1103 (2004).

[49] P. H. Dahl, R. H. Zhang, J. H. Miller, L. R. Bartek, Z. Peng, S. R. Ramp, J. X. Zhou, C. S. Chiu, J. F. Lynch, J. A. Simmen, and R. C. Spindel, "Overview of results from the Asian Sea International Experiment in the East China Sea," IEEE J. Ocean. Eng. **29**, 920–928 (2004).

[50] Z. H. Peng, J. X. Zhou, P. H. Dahl, and R. H. Zhang, "Sea-bed acoustic parameters from dispersion analysis and transmission loss in the East China Sea," IEEE J. Ocean. Eng. **29**, 1038–1045 (2004).

[51] Z. L. Li, R. H. Zhang, J. Yan, F. H. Li, and J. J. Liu, "Geoacoustic inversion by matched-field processing combined with vertical reflection coefficients and vertical correlation," IEEE J. Ocean. Eng. **29**, 973–979 (2004).

[52] J. X. Zhou, X. Z. Zhang, P. H. Rogers, J. A. Simmen, P. H. Dahl, G. L. Jin, and Z. H. Peng, "Reverberation vertical coherence and sea-bottom geoacoustic inversion in shallow water," IEEE J. Ocean. Eng. **29**, 988–999 (2004).

[53] D. P. Knobles, T. W. Yudichak, R. A. Koch, P. G. Cable, J. H. Miller, and G. R. Potty, "Inferences on seabed acoustics in the East China Sea from distributed acoustic measurements," IEEE J. Ocean. Eng. **31**, 129–144 (2006).

[54] D. P. Knobles, R. A. Koch, T. Udagawa, and E. K. Wstwood, "The inversion of ocean waveguide parameters using a nonlinear least squares approach," J. Comput. Acoust. **6**, 1–15 (1998).

[55] P. A. Barakos, "Experimental determination of compressional speed for the bottom layer by the dispersion method," J. Acoust. Soc. Am. **34**, 1919–1925 (1962).

[56] J. M. Tattersall, D. Chizhik, B. F. Cole, and F. R. DiNapoli, "The effect of frequency-dependent bottom reflectivity on transmission loss in shallow water over a sandy bottom," J. Acoust. Soc. Am. **93**, 2395 (1993).

[57] B. F. Cole and E. M. Podeszwa, "Shallow-water propagation under down-refraction conditions," J. Acoust. Soc. Am. **41**, 1479–1484 (1967).

[58] J. S. Cohen and B. F. Cole, "Shallow-water propagation under down-refraction conditions. II," J. Acoust. Soc. Am. **61**, 213–217 (1977).

[59] R. B. Evans and W. M. Carey, "Frequency dependence of sediment at-

tenuation in two-frequency shallow-water acoustic experimental data sets," IEEE J. Ocean. Eng. **23**, 439–447 (1998).

[60] M. Badiey, A. H.-D. Chaeng, and Y. Mu, "From geology to geoacoustics-evaluation of Biot-Stoll sound speed and attenuation for shallow water acoustics," J. Acoust. Soc. Am. **103**, 309–320 (1998).

[61] S. M. Dediu, W. L. Siegmann, and W. M. Carey, "Statistical analysis of sound transmission results obtained on the New Jersey continental shelf," J. Acoust. Soc. Am. **122**, EL23–EL28 (2007).

[62] I. Rozenfeld, W. M. Carey, P. G. Cable, and W. L. Siegmann, "Modeling and analysis of sound transmission in the Strait of Korea," IEEE J. Ocean. Eng. **26**, 809–819 (2001).

[63] J. D. Holmes, W. M. Carey, and J. F. Lynch, "Shallow-water waveguide characterization using an autonomous underwater vehicle-towed hydrophone array (A)," J. Acoust. Soc. Am. **119**, 3346 (2006).

[64] D. P. Knobles, P. S. Wilson, J. A. Goff, and S. E. Cho, "Seabed acoustics of a sand ridge on the New Jersey continental shelf," J. Acoust. Soc. Am. **124**, EL151–EL156 (2008).

[65] J. H. Beebe, S. T. McDaniel, and L. A. Rubano, "Shallow-water transmission loss prediction using the Biot sediment model," J. Acoust. Soc. Am. **71**, 1417–1426 (1982).

[66] P. C. Hines, J. C. Osler, J. Scrutton, and A. P. Lyons, "Time-of-flight measurements of acoustic wave speed in sandy sediments from 0.6–20 kHz," *Boundary Influences in High Frequency, Shallow Water Acoustics* (University of Bath, UK, 2005), pp. 49–56.

[67] G. R. Potty, J. H. Miller, and J. F. Lynch, "Inversion for sediment geoacoustic properties at the New England Bight," J. Acoust. Soc. Am. **114**, 1874–1887 (2003).

[68] Y. M. Jiang, N. R. Chapman, and M. Badiey, "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data," J. Acoust. Soc. Am. **121**, 1879–1894 (2007).

[69] S. D. Rajan and G. V. Frisk, "Seasonal variations of the sediment compressional wave-speed profile in the Gulf of Mexico," J. Acoust. Soc. Am. **91**, 127–135 (1992).

[70] T. P. Kumar, "Seasonal variation of relaxation time and attenuation sediment at the sea bottom interface," Acta Acust. **83**, 461–466 (1997).

[71] R. D. Stoll, "Comments on Biot model of sound propagation in water-saturated sand [JASA 97, 199–214 (1995)]," J. Acoust. Soc. Am. **103**, 2723–2725 (1998).

[72] E. I. Thorsos, K. L. Williams, N. P. Chotiros, J. T. Christoff, K. W. Commander, C. F. Greenlaw, D. V. Holliday, D. R. Jackson, J. L. Lopes, D. E. McGehee, J. E. Piper, M. D. Richardson, and D. J. Tang, "An overview of SAX99: Acoustic measurements," IEEE J. Ocean. Eng. **26**, 4–25 (2001).

[73] M. D. Richardson, K. B. Briggs, L. D. Bibee, P. A. Jumars *et al.*, "An overview of SAX99: Environmental considerations," IEEE J. Ocean. Eng. **26**, 26–53 (2001).

[74] H. J. Simpson, B. H. Houston, S. W. Liskey, P. A. Frank, A. R. Berdoz, I. A. Kraus, C. K. Frederickson, and S. Stanic, "At-sea measurements of sound penetration into sediments using a buried vertical synthetic array," J. Acoust. Soc. Am. **114**, 1281–1290 (2003).

[75] D. M. F. Chapman, P. D. Ward, and D. D. Ellis, "The effective depth of a Pekeris ocean waveguide, including shear wave effects," J. Acoust. Soc. Am. **85**, 648–653 (1989).

[76] Z. Y. Zhang and C. T. Tindle, "Complex effective depth of the ocean bottom," J. Acoust. Soc. Am. **93**, 205–213 (1993).

[77] S. K. Mitchell and K. C. Focke, "The role of sea-bottom attenuation profile in shallow-water acoustic propagation," J. Acoust. Soc. Am. **73**, 465–473 (1983).

[78] D. M. F. Chapman, "What are we inverting for," in *Inverse Problems in Underwater Acoustics*, edited by M. I. Taroudakis and G. Makrakis (Springer, New York, 2001), pp. 1–14.

Bayesian inversion of reverberation and propagation data for geoacoustic and scattering parameters

Stan E. Dosso[a)]
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

Peter L. Nielsen and Christopher H. Harrison
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19138 La Spezia, Italy

(Received 10 December 2008; revised 26 February 2009; accepted 26 February 2009)

This paper applies nonlinear Bayesian inference theory to quantify the information content of reverberation and short-range propagation data, both individually and in joint inversion, to resolve seabed geoacoustic and scattering properties. The inversion of reverberation data alone is shown to poorly resolve seabed properties because of strong multi-dimensional correlations between parameters. Inversion of propagation data alone is limited by different correlations, but better constrains the geoacoustic parameters. However, propagation data are insensitive to scattering parameters such as Lambert's scattering coefficient. In each case the parameter correlations are inherent in the physics of the forward problem (reverberation and propagation) and cannot be overcome by processing or inversion techniques; rather, the inversion of more informative data is required. This is accomplished here by joint inversion of reverberation and propagation data, weighted according to their respective maximum-likelihood error estimates. Joint inversion of reverberation and propagation data collected on the Malta Plateau (Strait of Sicily) resolves both geoacoustic and scattering properties and achieves smaller uncertainties for all parameters than obtained by the inversion of either data set alone.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106524]

PACS number(s): 43.30.Pc, 43.30.Wi, 43.60.Pt [AIT] Pages: 2867–2880

I. INTRODUCTION

Seabed geoacoustic and scattering properties (bottom layering, sound speed, density, attenuation, and scattering strength) represent key environmental inputs for sonar prediction tools. However, these properties are difficult and expensive to measure by direct methods (e.g., coring). Geoacoustic inversion represents a practical alternative based on estimating *in situ* seabed parameters from measured acoustic data. However, geoacoustic inversion represents a nonlinear problem that is inherently nonunique, and hence it is important to quantify the uncertainties of seabed parameter estimates, which can represent the limiting factor in sonar processing.

In recent years there has been growing interest in "through-the-sensor" inversion approaches which are based on using the same sonar system for data acquisition as intended for performance predictions. Through-the-sensor methods allow efficient surveying of wide regions and provide parameter sensitivities that are suited to the application. Through-the-sensor methods include inversion of short-range propagation data measured with a towed source-receiver configuration[1–5] and inversion of long-range reverberation data.[6–14] Both methods have been considered individually in the past; however, neither short-range propagation data nor reverberation data contain sufficient information about all seabed geoacoustic and scattering parameters. In particular, the propagation data are insensitive to scattering strength and are relatively insensitive to seabed attenuation because of the short propagation range. Further, the dependence of reverberation data on scattering and geoacoustic parameters is complicated by inter-parameter correlations (discussed below). One of the goals of this work is to quantify the improvement in geoacoustic and scattering parameter estimation achieved by joint inversion of reverberation and short-range propagation data measured on the same sonar system.

To date, work on reverberation inversion has concentrated on the ability to fit the data with a particular seabed model, often obtained via optimization, without rigorously quantifying the uncertainties of the estimated parameters (i.e., the range of parameter values which acceptably fit the data). This is of paramount importance since without some measure of uncertainties, no confidence can be placed in the model parameters obtained by inversion. A rigorous approach to uncertainty estimation is particularly important for problems in which the data depend on combinations of the unknown parameters. This limits the ability to resolve individual parameters, leading to correlated parameter estimates and increased uncertainties via inversion. Harrison[8] and Ainslie[12] derived approximate analytic expressions illustrating that reverberation intensity depends on combinations of parameters representing scattering strength and seabed reflection loss (which itself depends on combinations of geoacoustic parameters). However, the actual effects of these correlations in terms of parameter uncertainties via inversion are not readily apparent from the analytic treatment and require quantitative inversion analysis. It is important to note that

[a)] Author to whom correspondence should be addressed. Electronic mail: sdosso@uvic.ca correlation effects are inherent in the physics of reverberation and cannot be addressed by data processing or inversion techniques. To overcome correlations, information must be added to the problem, either in the form of additional data or prior information.

This paper applies a nonlinear Bayesian formulation[15–17] to the inversion of reverberation and/or short-range propagation data. Bayesian inversion treats the model parameters as random variables constrained by measured data and prior information, with the goal of computing properties of the posterior probability density (PPD). In addition to optimal parameter estimation, the rigorous computation of parameter variances, correlations, and one- and two-dimensional marginal probability distributions quantifies parameter uncertainties and inter-parameter relationships and provides an understanding of the geoacoustic information content of the inverse problem. Numerical reverberation modeling required in the inversion is carried out efficiently using the formulation of Harrison,[8] but substituting a numerical angle integral that incorporates Lambert's law scattering and reflection from layered sediments.

The Bayesian inversion approach is first applied to examine inter-parameter correlations in reverberation inversion using simple simulations. Bayesian inversion is also applied to measured reverberation and short-range propagation data recorded on the Malta Plateau south of Sicily in the BASE'04 experiment, with the data sets considered both separately and in joint inversion. Joint inversion employs a maximum-likelihood weighting of the two data sets based on their estimated error statistics (including effects of both theory and measurement errors). The assumptions of random, Gaussian-distributed errors, which define the likelihood function, are validated using residual statistical tests. The results indicate that inversion of single-frequency reverberation data cannot usefully resolve geoacoustic or scattering parameters due to strong correlation effects. However, joint inversion of reverberation and short-range propagation data provides good estimates of Lambert's scattering coefficient and seabed sound-speed structure, with significantly better parameter resolution than achieved by inversion of either data set alone.

II. THEORY AND IMPLEMENTATION

A. Bayesian inversion

This section summarizes a Bayesian approach to geoacoustic inversion;[15–17] more complete treatments of Bayesian inference theory can be found elsewhere.[18–21] Let **d** and **m** represent random data and model-parameter vectors related by Bayes rule

$$\mathbf{P}(\mathbf{m}|\mathbf{d})P(\mathbf{d}) = P(\mathbf{d}|\mathbf{m})P(\mathbf{m}). \tag{1}$$

For measured (fixed) data, $P(\mathbf{d})$ is a constant factor, and the conditional data probability $P(\mathbf{d}|\mathbf{m})$ is interpreted as a function of **m**, known as the likelihood function, $L(\mathbf{m}) \propto \exp[-E(\mathbf{m})]$, where E is the data misfit function (considered in Sec. II B). Combining the likelihood and prior distribution $P(\mathbf{m})$ to define a generalized misfit

$$\phi(\mathbf{m}) \equiv E(\mathbf{m}) - \log_e P(\mathbf{m}), \tag{2}$$

the PPD can be written as

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m})]}{\int \exp[-\phi(\mathbf{m}')]d\mathbf{m}'}, \tag{3}$$

where the domain of integration spans the parameter space. The multi-dimensional PPD is typically interpreted in terms of properties defining parameter estimates, uncertainties, and inter-relationships, such as the maximum *a posteriori* (MAP) estimate, mean model, model covariance matrix, and marginal probability distributions defined, respectively, as

$$\hat{\mathbf{m}} = \arg_{\max}\{P(\mathbf{m}|\mathbf{d})\} = \arg_{\min}\{\phi(\mathbf{m})\}, \tag{4}$$

$$\bar{\mathbf{m}} = \int \mathbf{m}' P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{5}$$

$$\mathbf{C_m} = \int (\mathbf{m}' - \bar{\mathbf{m}})(\mathbf{m}' - \bar{\mathbf{m}})^T P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{6}$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i' - m_i) P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{7}$$

where δ is the Dirac delta function and two-dimensional (joint) marginal distributions are defined in a manner similar to Eq. (7). Inter-parameter correlations are quantified by normalizing the covariance matrix to produce the correlation matrix

$$R_{ij} = C_{m_{ij}} / \sqrt{C_{m_{ii}} C_{m_{jj}}}. \tag{8}$$

Elements R_{ij} are within $[-1,1]$, with a value of 1 indicating perfect correlation between parameters m_i and m_j, -1 indicating perfect anti-correlation, and near-zero values indicating uncorrelated parameters.

For nonlinear problems, such as geoacoustic inversion, numerical solutions to the optimization and integrations in Eqs. (4)–(7) are required. In this paper, optimization is carried out using adaptive simplex simulated annealing,[22] a hybrid optimization that combines elements of very fast simulated annealing and the downhill simplex method. Integration is carried out by sampling the PPD using Metropolis–Hastings sampling (MHS) (sometimes referred to as Metropolis–Gibbs sampling), a Markov-chain Monte Carlo importance-sampling method.[20,21] In MHS, the parameters of the model are perturbed repeatedly, with perturbations accepted if a random number ξ drawn from a uniform distribution on [0, 1] satisfies the Metropolis criterion

$$\xi \leqslant \exp(-\Delta\phi), \tag{9}$$

where $\Delta\phi$ represents the change in ϕ due to the perturbation. The parameter perturbations in MHS can be drawn from any proposal distribution; the choice of proposal distribution does not affect the integral estimates, but can strongly affect the sampling efficiency. Here, perturbations are applied in a principal-component parameter space where the axes align with the dominant correlation directions.[15,16] The orthogonal transformation (rotation) between physical parameters **m** and rotated parameters **m**′ is given by

$$\mathbf{m}' = \mathbf{U}^T\mathbf{m}, \quad \mathbf{m} = \mathbf{U}\mathbf{m}', \tag{10}$$

where $\mathbf{U}$ is the column-eigenvector matrix of the model covariance $\mathbf{C_m}$,

$$\mathbf{C_m} = \mathbf{U}\mathbf{W}\mathbf{U}^T, \tag{11}$$

and $\mathbf{W}=\text{diag}[w_i]$ is the eigenvalue matrix, with w_i representing the parameter variances projected along eigenvector $\mathbf{u}_i$. Rotated parameters are perturbed individually, and the perturbed models rotated back to the physical space for misfit evaluation. The covariance matrix also provides representative length scales for perturbations in rotated space. For a linear problem, the optimal proposal distribution is Gaussian-distributed with variances w_i from Eq. (11). However, for nonlinear problems, a more appropriate proposal distribution is based on the heavy-tailed Cauchy distribution, scaled according to the estimated rotated variances.[17]

The above procedure can be initiated efficiently using the model covariance matrix for a linearized approximation to the nonlinear inverse problem[17,23]

$$\mathbf{C_m} = [\mathbf{J}^T\mathbf{C_d}^{-1}\mathbf{J} + \mathbf{C}_{\hat{\mathbf{m}}}^{-1}]^{-1}, \tag{12}$$

where $\mathbf{J}$ represents the Jacobian matrix of partial derivatives evaluated at a MAP starting model $\hat{\mathbf{m}}$ determined by optimization

$$J_{ij} = \frac{\partial d_i(\hat{\mathbf{m}})}{\partial m_j}, \tag{13}$$

and $\mathbf{C}_{\hat{\mathbf{m}}}$ is the covariance matrix of a Gaussian prior distribution about $\hat{\mathbf{m}}$. (In the common case where the actual parameter priors consist of bounded uniform distributions, $m_i^- < m_i < m_i^+$, the covariance $\mathbf{C}_{\hat{\mathbf{m}}}$ is taken to be a diagonal matrix with variances equal to those of the uniform distributions, $[m_i^+ - m_i^-]^2/12$.) As sampling progresses, the initial linearized estimate for $\mathbf{C_m}$, Eq. (12), is adaptively replaced with the nonlinear estimate, Eq. (6), based on the MHS solution to that point, which better represents the overall covariance of the parameter space and the actual prior distribution.[17] For inverse problems involving strongly correlated parameters (such as reverberation inversion), initiating the sampling from a linearized estimate as opposed to applying initial unrotated sampling to build the covariance estimate for rotation can increase efficiency by orders of magnitude.

Numerical integration and optimization require solving the forward problem a large number of times [i.e., applying reverberation and/or propagation modeling to $\mathbf{m}$ to compute the misfit $E(\mathbf{m})$]. Hence, it is important that this be carried out as efficiently as possible. Propagation modeling is carried out here using ray theory [GAMARAY (Ref. 24)], including the tracing of bottom-penetrating rays, which represents a highly efficient approach to computing acoustic fields at a large number of frequencies. Reverberation modeling is based on a two-way propagation and scattering formulation derived by Harrison[8] as an angle integral over a continuum of incoherent modes or rays for the outward and return paths (see also Refs. 12, 25, and 26). An effective complex half-space reflection coefficient is efficiently calculated for arbitrary layered sediments using the technique specified in Ref. 27. The power reflection coefficient, raised to the power of range divided by ray-cycle distance, is then substituted for the exponential terms in the angle integrals [e.g., Eq. (27) of Ref. 8].

To be consistent with the effective half-space reflection, the same half-space boundary is assumed to be the source of scattering, and all volume and sub-layer scattering is ignored. The precise meaning of "incident field" in the context of distant scatterers is discussed in Ref. 8 [paragraph after Eq. (25) on page 2748]. Lambert's law is applied here as a scattering function

$$S = \mu' \sin\theta_{\text{in}} \sin\theta_{\text{out}}, \tag{14}$$

where S is the linear scattering strength, θ_{in} and θ_{out} are the in- and out-going scattering angles, respectively, and μ' is an empirical constant which typically has a value of roughly $\mu' = 10^{-2.7}$ or $\mu = -27$ dB re 1 μPa, where $\mu = 10 \log \mu'$ represents Lambert's scattering coefficient. Comparisons of reverberation predictions of Harrison's[8] analytic solution with those of a numerical reverberation model based on ray tracing and Lambert's law scattering indicate differences of a fraction of decibel for the model types considered in this paper, which is generally insignificant compared to the level of data noise.[28,29]

B. Likelihood and misfit

This section derives the likelihood-based data misfit function for combined reverberation and propagation inversion; the misfit for either data set alone is easily extracted from the result. The case considered here involves reverberation data (in decibels) as a function of range at a single-frequency, and propagation data as a function of either space or frequency. The extension to multi-frequency reverberation data and/or propagation data as a function of both space and frequency is straightforward, but is not considered in this paper.

Consider N_r real reverberation data $\mathbf{d_r}$ with Gaussian-distributed random errors with covariance matrix $\mathbf{C_r}$, and N_p complex propagation data $\mathbf{d_p}$ with circularly-symmetric, complex Gaussian errors with covariance matrix $\mathbf{C_p}$. Assuming that errors on the two data sets are independent, the likelihood function for the combined data is the product of the reverberation likelihood $L_r(\mathbf{m})$ and the propagation likelihood $L_p(\mathbf{m})$:

$$L(\mathbf{m}) = L_r(\mathbf{m})L_p(\mathbf{m}), \tag{15}$$

where

$$L_r(\mathbf{m}) = \frac{1}{(2\pi)^{N_r/2}|\mathbf{C_r}|^{1/2}} \exp\{-[\mathbf{d_r} - \mathbf{d_r}(\mathbf{m})]^T\mathbf{C_r}^{-1}[\mathbf{d_r} - \mathbf{d_r}(\mathbf{m})]/2\}, \tag{16}$$

$$L_p(\mathbf{m}) = \frac{1}{\pi^{N_p}|\mathbf{C_p}|} \exp\{-[\mathbf{d_p} - Ae^{i\theta}\mathbf{d_p}(\mathbf{m})]^\dagger\mathbf{C_p}^{-1}[\mathbf{d_p} - Ae^{i\theta}\mathbf{d_p}(\mathbf{m})]\}. \tag{17}$$

In Eqs. (16) and (17), $\mathbf{d_r}(\mathbf{m})$ and $\mathbf{d_p}(\mathbf{m})$ represent replica (modeled) reverberation and propagation data, respectively, and A and θ define the complex source magnitude and phase (considered unknown for the propagation data). If it can be assumed that both the reverberation errors and the propagation errors are uncorrelated, the error covariance matrices become $\mathbf{C_r}=\nu_r\mathbf{I}_{N_r}$ and $\mathbf{C_p}=\nu_p\mathbf{I}_{N_p}$, where ν_r and ν_p are the respective variances and $\mathbf{I}_{N_r}$ and $\mathbf{I}_{N_p}$ are identity matrices of dimension N_r and N_p. The unknown variances and source factors can be treated by maximizing the likelihood over ν_r, ν_p, A, and θ (i.e., setting $\partial L/\partial\nu_r=\partial L/\partial\nu_p=\partial L/\partial A=\partial L/\partial\theta=0$).[30,31] After some algebra, the resulting maximum-likelihood (ML) misfit function may be expressed as

$$E(\mathbf{m}) = \frac{N_r}{2\tilde{\nu}_r}|\mathbf{d_r}-\mathbf{d_r}(\mathbf{m})|^2 + \frac{N_p}{\tilde{\nu}_p}\left(|\mathbf{d_p}|^2 - \frac{|\mathbf{d_p}(\mathbf{m})^\dagger\mathbf{d_p}|^2}{|\mathbf{d_p}(\mathbf{m})|^2}\right), \tag{18}$$

where the ML variance estimates $\tilde{\nu}_r$ and $\tilde{\nu}_p$ are given by

$$\hat{\nu}_r = \frac{1}{N_r}|\mathbf{d_r}-\mathbf{d_r}(\tilde{\mathbf{m}})|^2, \tag{19}$$

$$\hat{\nu}_p = \frac{1}{N_p}\left(|\mathbf{d_p}|^2 - \frac{|\mathbf{d_p}(\tilde{\mathbf{m}})^\dagger\mathbf{d_p}|^2}{|\mathbf{d_p}(\tilde{\mathbf{m}})|^2}\right), \tag{20}$$

evaluated at the model estimated by (numerically) minimizing the misfit function

$$\tilde{E}(\mathbf{m}) = \frac{N_r}{2}\log_e|\mathbf{d_r}-\mathbf{d_r}(\mathbf{m})|^2 + N_p\log_e\left(|\mathbf{d_p}|^2 - \frac{|\mathbf{d_p}(\mathbf{m})^\dagger\mathbf{d_p}|^2}{|\mathbf{d_p}(\mathbf{m})|^2}\right). \tag{21}$$

Note that in the combined ML misfit function, Eq. (18), the reverberation misfit (first term) is measured by an L_2 (magnitude-squared) norm while the propagation misfit (second term) is measured by a Bartlett correlator; each misfit term is weighted according to the number of data and the estimated data variance (the factor of 2 difference arises because the complex propagation data include real and imaginary parts). The ML variance estimates in Eqs. (19) and (20) quantify the total data uncertainty of the inverse problem, including both measurement and theory errors.

The misfit function derived here is based on the assumptions of uncorrelated, Gaussian-distributed random errors on both the reverberation and propagation data. If these assumptions are not valid, the inversion results (particularly uncertainty estimates) may not be reliable. Hence, the assumptions should be examined *a posteriori* by considering standardized residuals for reverberation and propagation data defined[32] as

$$\mathbf{n_r} = [\mathbf{d_r}-\mathbf{d_r}(\hat{\mathbf{m}})]/\nu_r^{1/2}, \tag{22}$$

$$\mathbf{n_p} = \left[\mathbf{d_p} - \frac{\mathbf{d_p}(\hat{\mathbf{m}})^\dagger\mathbf{d_p}}{|\mathbf{d_p}(\hat{\mathbf{m}})|^2}\mathbf{d_p}(\hat{\mathbf{m}})\right] \Bigg/ \nu_p^{1/2}, \tag{23}$$

where $\hat{\mathbf{m}}$ is the MAP model obtained by minimizing Eq. (2) with $E(\mathbf{m})$ given by Eq. (18). The assumption of Gaussian-distributed errors can be considered qualitatively by comparing a histogram of the residuals to the standard normal distribution. Quantitative statistical tests, such as the Kolmogorov–Smirnov (KS) test,[32,33] can also be applied to provide a p-value indicating the level of evidence against the null hypothesis of Gaussian-distributed errors (commonly, $p\geq 0.05$ is considered to provide no significant evidence against the null hypothesis). The assumption of uncorrelated errors can be considered qualitatively by plotting the residual autocorrelation: uncorrelated errors lead to a narrow correlation peak at zero lag, while serial correlations widen the peak. Statistical tests, such as the runs test, can be applied to quantify the level of evidence against the null hypothesis of uncorrelated errors.[32,33] If either of these tests are failed, it may be necessary to estimate and include full covariance matrices or consider an alternative data error distribution, although this complicates the analysis.[32]

In this paper, the reverberation data errors are assumed to be Gaussian distributed in decibels. This is an example of the proportional effect,[34] whereby the magnitude of the uncertainties scales directly with the magnitude of the data (not uncommon for theory errors). In particular, consider reverberation data d_i in linear units with error given by ηd_i, where η is a random variable (i.e., the error scales with the data). Converting to decibels

$$10\log_{10}(d_i+\eta d_i)^2 = 10\log_{10}d_i^2 + 10\log_{10}(1+\eta)^2, \tag{24}$$

i.e., the data and error terms can both be expressed in decibels. The validity of this assumption can be examined using the *a posteriori* residual statistical tests described above.

III. SIMULATION: INTER-PARAMETER CORRELATIONS IN REVERBERATION INVERSION

This section considers Bayesian reverberation inversion for a simplified synthetic test case, prior to inverting measured data in Sec. IV. Advantages of considering simulations include the facts that the true model is known for comparison and the data error statistics are known and controlled. Hence, the ideal information content for an inversion scenario can be examined independent of complicating factors which often arise in at-sea experiments. Specific aspects of an inverse problem can be isolated and studied with simulation in a manner that is not generally possible with measured data. Here, inter-parameter correlations are studied via simulation.

Before considering the simulation results, it is instructive to consider a simple analysis based on an approximate analytic formulation for a uniform ocean of depth D over a layered seabed. Within the critical angle, θ_c, the reflection loss in decibels, $R_{\mathrm{dB}}=-10\log|R|^2$ (where R is the reflection coefficient), is well approximated as a linear function of grazing angle θ as $R_{\mathrm{dB}}=\alpha_{\mathrm{dB}}\theta$ or

$$|R| = \exp[-(\alpha/2)\theta], \tag{25}$$

where $\alpha=\alpha_{\mathrm{dB}}/(10\log e)$. Assuming Lambert's law scattering, an approximate relationship for reverberation intensity I as a function of range r is[8,11]

$$I = \frac{\mu\Phi\tau}{\alpha^2 r^3}\left\{1-\exp\left(\frac{-\alpha\theta_c^2}{2D}r\right)\right\}^2, \tag{26}$$

where Φ is the horizontal beam-width and τ is the spatial pulse length. At long range Eq. (26) can be approximated by[8,11]

$$I = \frac{\mu\Phi\tau}{\alpha^2 r^3}, \tag{27}$$

and at short range by

$$I = \frac{\mu\Phi\tau\theta_c^4}{2rD^2}, \tag{28}$$

with the transition range given by

$$r_0 = \frac{2D}{\alpha\theta_c^2}. \tag{29}$$

As noted by several authors,[8,11,12] at long range μ and α are inseparable as I depends on μ/α^2. To seek a simple relationship involving μ and geoacoustic parameters, consider the case of a water column with sound speed c_w and density ρ_w overlying a uniform seabed with sound speed c, density ρ, and attenuation coefficient which can be expressed either as a_λ in dB/wavelength or a in dB/m kHz, with $a_\lambda = ac/1000$. In this case the reflection loss can be related to the geoacoustic properties as[11,35]

$$R_{\mathrm{dB}} = \alpha_{\mathrm{dB}}\theta = \frac{a_\lambda \rho}{c^2 \sin^3 \theta_c}\frac{c_w^2}{\pi\rho_w}\theta. \tag{30}$$

Hence,

$$\alpha = \frac{a\rho}{c \sin^3 \theta_c}\frac{c_w^2}{\pi\rho_w 10\,000 \log e} \propto \frac{a\rho}{c[1-(c_w/c)^2]^{3/2}}. \tag{31}$$

Equation (31) and the dependence of long-range reverberation on μ/α^2 suggest a rough proportionality

$$I \propto \frac{\mu c^2}{a^2\rho^2}. \tag{32}$$

Equation (32) suggests that in reverberation inversion for seabed parameters, μ will be positively correlated with a since I is unchanged by increasing or decreasing both μ and a in a suitable manner. Likewise, Eq. (32) suggests that μ is also positively correlated with ρ but negatively correlated with c, that c is positively correlated with both a and ρ, and that a is negatively correlated with ρ. However, the following simulation shows that actual inversion results are not so straightforward.

The synthetic test case consists of monostatic reverberation for a source and receiver at 70-m depth in a uniform 100-m water column with sound speed 1500 m/s, over a uniform seabed with sound speed 1580 m/s, density 1.6 g/cm^3, attenuation 0.32 dB/m kHz, and Lambert's scattering coefficient −27 dB. For simplicity, only the seabed sound speed (c), attenuation (a), and scattering coefficient (μ) are treated as unknown parameters; other environmental and geometric parameters are considered known exactly. Eighty reverberation data were generated for scattering ranges of 1–10 km and contaminated with zero-mean, Gaussian-distributed random errors of standard deviation 3 dB (Fig. 1).

Figure 2 shows marginal posterior probability distributions for c, a, and μ computed for four inversion cases which are identical except that the uniform prior distribution for c increases (from top to bottom in each panel) as follows: [1579, 1581], [1570, 1590], [1540, 1620], and [1500, 1750] m/s. Wide uniform priors of [0, 1] dB/m kHz and [−20, −40] dB are employed for a and μ in all inversions. Figure 2 shows that the marginal distributions for all three parameters widen substantially as the prior bounds for c increase. In particular, for the narrowest prior on c, good inversion estimates (narrow marginal distributions) are obtained for a and μ, while for the widest prior, a and μ are poorly resolved (wide marginals). This behavior is indicative of correlated parameters.

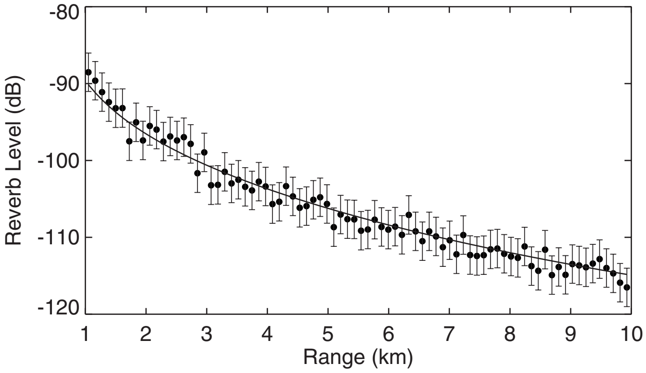


FIG. 1. Data for the synthetic reverberation example. Noisy data are shown as filled circles with one standard-deviation error bars; solid line indicates noise-free data.

Inter-parameter correlations for the four inversions are quantified in Fig. 3 in terms of correlation matrices. For the narrowest prior on c, Fig. 3(a) shows that c is uncorrelated with a and μ, likely because the small range of allowable values for c is inconsequential to the other parameters, given the noise on the data. However, a and μ exhibit a strong positive correlation, as expected from the analytic analysis

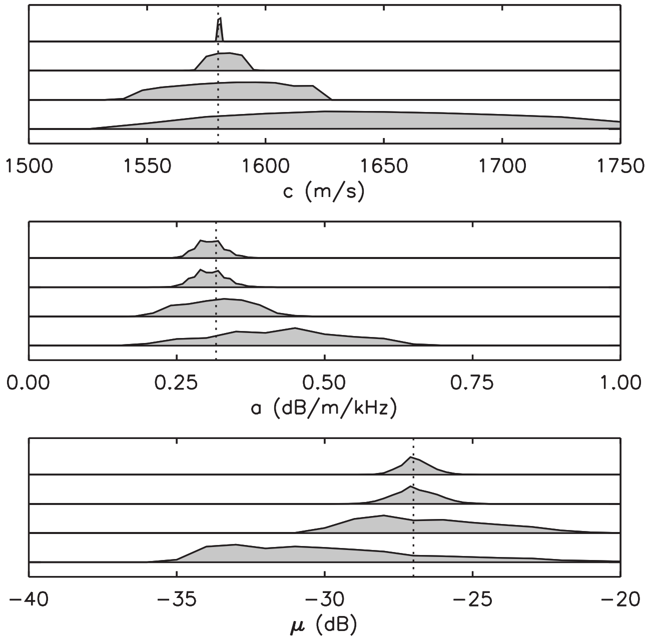


FIG. 2. Marginal probability distributions for the synthetic reverberation inversion example. The four distributions in each panel represent results of different inversions of the same noisy data, with the prior bounds on sound speed c increasing from top to bottom. Dotted lines indicate true parameter values.

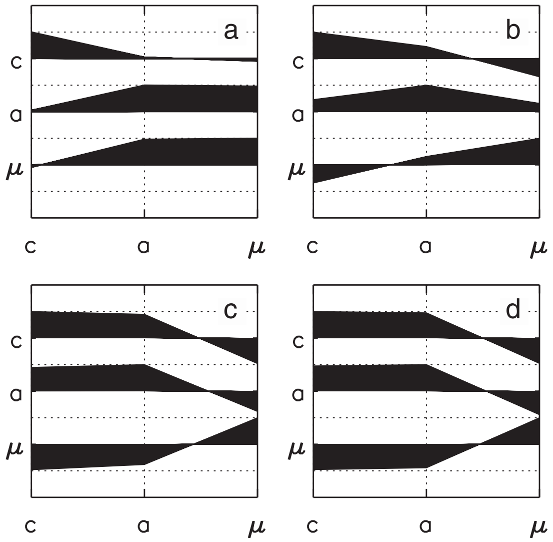


FIG. 3. Correlation matrices for the synthetic reverberation inversion example; (a)–(d) indicate results for increasing prior bounds for sound speed *c*.

above. In progressing through successively wider priors on *c* in Figs. 3(a)–3(d), a strong positive correlation develops between *c* and *a*, and a strong negative correlation develops between *c* and μ, as expected from the analysis. Perhaps surprisingly, however, the relationship between *a* and μ changes from a strong positive correlation for a narrow prior on *c* [Fig. 3(a)] to a strong negative correlation for a wide prior [Figs. 3(c) and 3(d)], which was not expected from the approximate analytic analysis.

The simulation results are examined further in Fig. 4, which shows joint marginal probability distributions for the four inversion cases arranged in four columns, with the prior bound width for *c* increasing from left to right. Figure 4 shows that interparameter correlations result in ridges of probability that run obliquely through the parameter space. The two top rows of Fig. 4 illustrate the development of the positive and negative correlations between *c* and *a* and between *c* and μ, respectively, as the prior bound width for *c* increases. The bottom row illustrates the relationship between *a* and μ with increasing prior bound width. For the narrowest prior bound on *c*, Fig. 4(i) shows that the joint marginal distribution over *a* and μ is oriented with a positive slope (i.e., a positive correlation, as expected from the analytic analysis). In Fig. 4(j), the effect of increasing the bound width for *c* is that the marginal distribution elongates in a direction roughly perpendicular to the original correlation direction [i.e., perpendicular to the distribution orientation in Fig. 4(i)]. This effect becomes increasingly pronounced in Figs. 4(k) and 4(l), with the marginal distribution elongating to the extent that it acquires an overall negative slope, resulting in a negative correlation between *a* and μ. This transition from positive to negative correlation results from multidimensional effects. For a fixed *c* value, the PPD cross-section in *a* and μ is oriented with a positive slope. However, the PPD extends in the *c* dimension such that its projection on the *a*-μ plane has a negative slope (this follows from the fact that *c* is positively correlated with *a* but negatively correlated with μ). Hence, integrating over a small range for *c* produces a joint marginal with a positive slope/correlation, while integrating over a sufficiently large range produces a joint marginal with a negative slope/correlation.

Figure 4 provides a simple example illustrating that strong inter-parameter correlations can preclude meaningful parameter estimation in reverberation inversion. In effect,

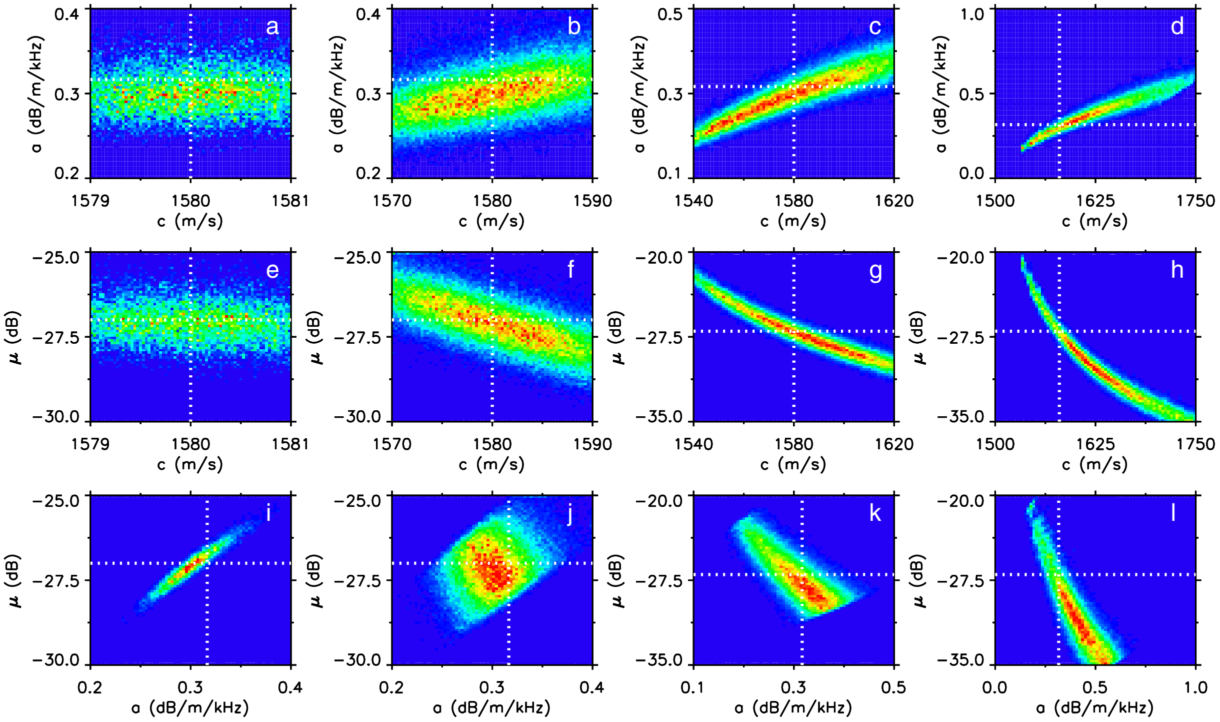


FIG. 4. (Color online) Joint marginal probability distributions for the synthetic reverberation inversion example. The four columns represent results for increasing prior bounds for sound speed *c* (increasing from left to right), as given by the plot range for *c*. Dotted lines indicate true parameter values.

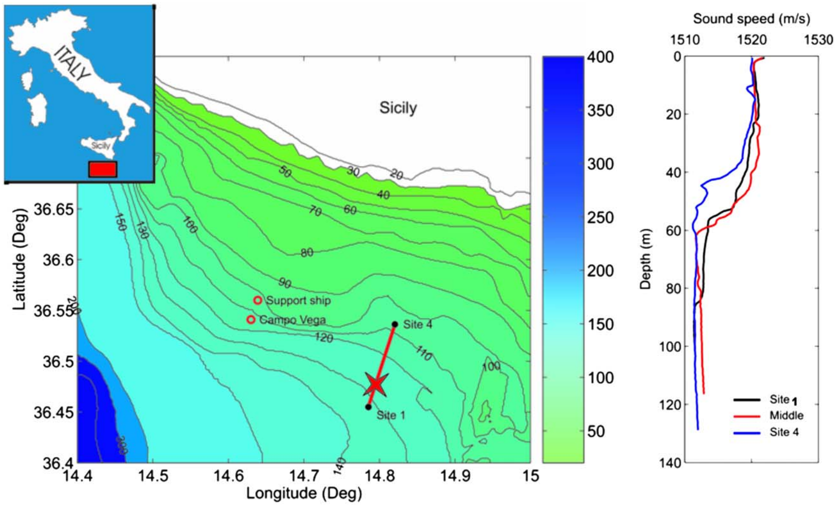


FIG. 5. (Color online) Location and bathymetry of experiment site. The heavy line indicates the ship track and the star indicates the ship location where the data considered in this paper were recorded. Sound-speed profiles measured along track are shown at right.

correlations result from a lack of information to resolve parameters individually. To overcome correlation effects, the problem must be augmented with additional data or prior information. The approach taken in this paper is to combine reverberation and short-range propagation data acquired on the same sonar system, as described in Sec. IV.

IV. INVERSION OF MALTA-PLATEAU DATA

A. Experiment, data, and model

In 2004, the NATO Undersea Research Centre conducted the BASE'04 experiment on the Malta Plateau in the Strait of Sicily (see Fig. 5). Part of this experiment was dedicated to through-the-sensor environmental assessment using a towed source-receiver configuration (Fig. 6). A particular 10-km tow track was chosen where some knowledge of the seabed properties existed from ground-truth measurements and geoacoustic inversion of data acquired during previous sea trials. The BASE'04 data considered here are based on transmissions of a 1-s linear-frequency-modulated signal in the band 850–1750 Hz, with a 1-min repetition rate along the track. The receiver array consisted of 84 hydrophones spaced at 0.42-m intervals, for a total array length of 34.86 m, towed approximately 370 m behind the source. The nominal source and receiver depths were 70 and 71 m, respectively. The water-column sound-speed profile was measured at three positions along the track (Fig. 5) and was relatively constant with range, with some variation over the thermocline from 40 to 60-m depth. The sound-speed profile used in inversion represented an average of the three measured profiles. The signals received at the array were recorded at a sampling frequency of 12.8 kHz for a 25-s time duration which includes short-range propagation data (with no clipping of the signal at the hydrophones) and long-range reverberation data out to scattering ranges of ~19 km (assuming an average water-column sound speed of 1515 m/s). One recording close to the south end of the track is considered for inversion in this paper.

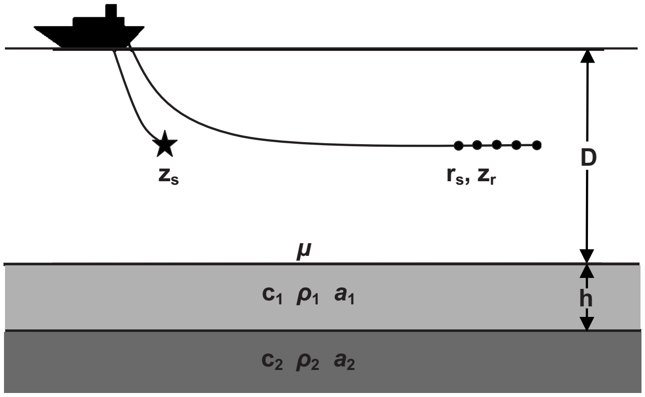


FIG. 6. Schematic of towed source/array configuration, indicating environmental and geometric unknowns (see text).

The recorded acoustic-pressure time series were matched-filtered with a synthetic source replica to yield the calibrated impulse response of the underwater waveguide. These matched-filtered array data were subsequently plane-wave beamformed and corrected for the estimated pulse length and beam-width to represent the long-range calibrated reverberation intensity resulting from an omni-directional source with a level of 1 μPa at 1 m and pulse length of 1 s (these source parameters are used in the numerical reverberation modeling). Only the broadside reverberation data are utilized in the inversion here, as these represent scattering in a direction roughly parallel to the bathymetric contours providing an approximately range-independent water depth (see Fig. 5). The measured reverberation level (in decibels) over the entire 0–19-km scattering range is shown in Fig. 7(a), averaged (in linear units) over the frequency band 900–1100 Hz. The data inverted here consist of 80 reverberation measurements for ranges of 1–10 km (i.e., ~112-m range interval), as shown in Fig. 7(b). For ranges less than 1 km the modeling assumption of a monostatic reverberation geometry is not justified, and beyond 10-km range Fig. 7(a) shows strong returns from three-dimensional features, such as the Ragusa Ridge and the Campo Vega oil platform and tethered support ship (Fig. 5), which cannot be treated with the numerical model. The short-range propagation data consist of complex acoustic fields obtained by fast-Fourier transforming the measured acoustic-pressure time series. Data are considered at 161 frequencies from 900 to 1700 Hz (i.e., 5-Hz frequency spacing, no frequency averaging), as measured at a hydrophone near the center of the towed array.

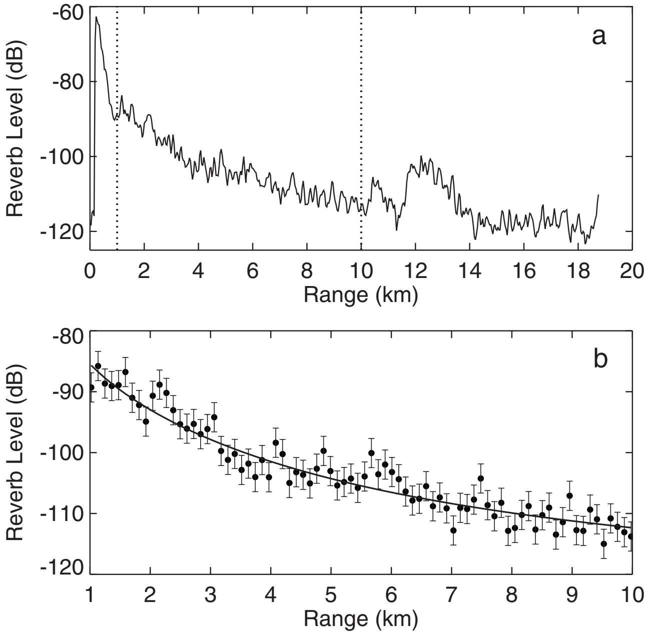


FIG. 7. Malta-Plateau reverberation data. (a) Data for all ranges (dotted lines indicate range interval used for inversion). (b) Data used for inversion with error bars corresponding to ML standard-deviation estimate; solid line indicates data fit of MAP model estimate.

The geoacoustic model (Fig. 6) adopted for inversion consists of an upper sediment layer of thickness h, sound speed c_1, density ρ_1, and attenuation a_1, overlying a semi-infinite basement with properties c_2, ρ_2, and a_2; Lambert's scattering coefficient is μ. Also included as unknowns are small corrections to the nominal geometric parameter values for the experiment corresponding to the water depth D, source depth z_s, receiver depth z_r, and source-receiver range r_s. Uniform bounded priors are assumed for all parameters. Wide bounds are assumed for the seabed parameters, consisting of 1500–1900 m/s for sound speeds, 1.2–2.2 g/cm^3 for densities, 0–1.5 dB/m kHz for attenuations, and −10 to −40 dB for μ. The bounds for water, source, and receiver depths are ±2 m about the nominal values, and the bounds for range extend from 360 to 390 km. Prior bounds for all parameters are summarized in Table I.

B. Reverberation inversion

In this section, the reverberation data described in Sec. IV A are inverted using the Bayesian methodology outlined in Sec. II. The fit to the measured data achieved by the MAP model estimate is shown in Fig. 7(b). Figure 8 shows the marginal probability distributions computed for all parameters. There is some sensitivity to scattering coefficient μ; however, the marginal distribution for this parameter varies over a large range of values from approximately −15 to −35 dB. The marginal distributions for all other parameters are essentially flat, although there is slight sensitivity for sediment parameters c_1, ρ_1, and a_1. Values for the MAP and mean parameter estimates, with two standard-deviation uncertainty estimates, are given in Table I (note that for nonlinear problems the mean and MAP estimates do not generally coincide, and the standard deviation about the mean is always less-than-or-equal to that about the MAP).

TABLE I. Prior bounds and parameter estimates with two standard-deviation uncertainties (upper value—MAP; lower value—mean) via inversion of reverberation (reverb), propagation (prop), and combined (reverb+prop) Malta-Plateau data sets.

Parameter and units	Prior bounds	Reverb inversion	Prop inversion	Reverb+prop inversion
μ (dB)	[−10, −40]	−19.9 ± 17	⋯	−24.6 ± 4.5
		−27.5 ± 8	⋯	−23.8 ± 4.1
h (m)	[0, 30]	9.5 ± 20	12.5 ± 4.0	12.4 ± 1.0
		13.9 ± 17	14.0 ± 3.0	12.3 ± 1.0
c_1 (m/s)	[1500, 1900]	1570 ± 370	1520 ± 50	1526 ± 10
		1720 ± 200	1540 ± 30	1524 ± 9
ρ_1 (g/cm^3)	[1.2, 2.2]	2.0 ± 0.6	1.5 ± 0.4	1.7 ± 0.4
		1.7 ± 0.6	1.5 ± 0.4	1.7 ± 0.4
a_1 (dB/m kHz)	[0, 1.5]	0.51 ± 1.1	0.08 ± 0.10	0.02 ± 0.02
		0.97 ± 0.7	0.03 ± 0.04	0.02 ± 0.02
c_2 (m/s)	[1500, 1900]	1510 ± 440	1770 ± 180	1750 ± 180
		1700 ± 220	1690 ± 100	1680 ± 90
ρ_2 (g/cm^3)	[1.2, 2.2]	1.8 ± 0.6	1.4 ± 0.5	1.4 ± 0.4
		1.7 ± 0.6	1.4 ± 0.5	1.5 ± 0.4
a_2 (dB/m kHz)	[0, 1.5]	0.64 ± 0.9	0.80 ± 0.9	0.74 ± 0.8
		0.74 ± 0.9	0.62 ± 0.8	0.61 ± 0.7
D (m)	[125, 129]	126.7 ± 1.1	128.9 ± 0.4	128.3 ± 1.0
		127.4 ± 0.9	128.3 ± 0.1	128.7 ± 0.5
z_r (m)	[69, 73]	72.3 ± 1.7	70.0 ± 0.7	69.4 ± 0.2
		72.0 ± 1.7	69.4 ± 0.2	69.3 ± 0.2
z_s (m)	[68, 72]	68.1 ± 2.3	68.1 ± 0.5	68.9 ± 0.4
		70.0 ± 1.2	68.5 ± 0.2	68.6 ± 0.2
R (m)	[360, 390]	380.9 ± 10	386.0 ± 0.1	386.0 ± 0.1
		374.8 ± 9	386.0 ± 0.1	386.0 ± 0.1

Figure 8 and Table I indicate that inversion of reverberation data alone cannot usefully resolve scattering and geoacoustic parameters in this case, even though the modeled

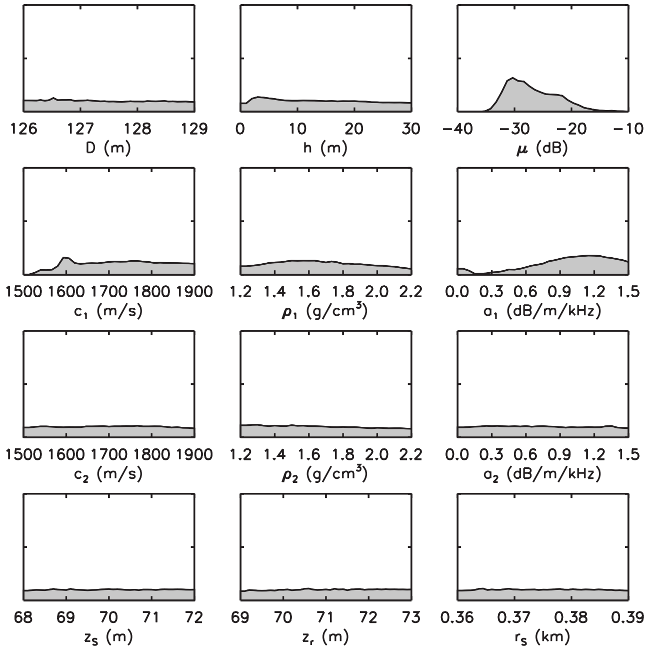


FIG. 8. Marginal posterior probability distributions from inversion of Malta-Plateau reverberation data.

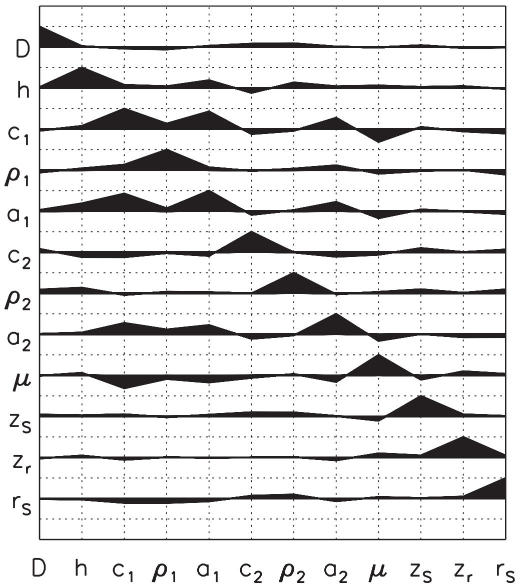


FIG. 9. Correlation matrix from inversion of Malta-Plateau reverberation data.

reverberation data depend strongly on these parameters. The reason appears to be inter-parameter correlations. The correlation matrix for reverberation inversion is shown in Fig. 9. This figure indicates a strong negative correlation of μ with c_1 and weaker negative correlations of μ with ρ_1, a_1, and a_2. Further, c_1 is strongly correlated with a_1, correlated less strongly with ρ_1 and a_2, and negatively correlated with c_2 (other inter-parameter correlations also exist). While this correlation matrix is more complex due to a greater number of parameters, the correlations between c_1, a_1, and μ are similar to those for the synthetic test case in Fig. 3(c) or Fig. 3(d).

Selected inter-parameter relationships are illustrated in Fig. 10 in terms of joint marginal probability distributions. This figure shows oblique probability ridges for correlated parameters, precluding precise parameter estimation. Hence, for example, while Fig. 10 shows that μ is reasonably well determined for any particular (fixed) value of c_1, the fact that c_1 is itself unknown precludes meaningful estimation of μ. Further, the uncertainty in c_1 is exacerbated by correlations with a_1, ρ_1, etc. As mentioned previously, overcoming inter-parameter correlations requires additional information. In Sec. IV D, combined reverberation and propagation data are inverted; however, first the propagation data are considered independently in Sec. IV C.

C. Propagation inversion

Inversion results for the short-range propagation data (alone) are shown in Fig. 11 (note changes in parameter plot ranges compared to Fig. 8). While the scattering coefficient μ is, of course, completely undetermined, the geometric parameters are highly resolved and reasonably good results (i.e., fairly narrow marginal distributions) are obtained for parameters defining the seabed sound-speed profile (h, c_1, and c_2). The sediment density ρ_1, although not well determined, is constrained to be less than approximately 1.8 g/cm^3, which is consistent with the relatively low sediment sound speed (1500–1575 m/s) over the 11–17-m layer. A low value for sediment attenuation a_1 is indicated. The basement sound speed c_2 is less well constrained than c_1, with values in the range 1600–1800 m/s. Basement attenuation a_2 is unresolved over the prior bounds. Basement density ρ_2 appears to be constrained to values less than about 2.0 g/cm^3; however, it is not clear if this is meaningful, as

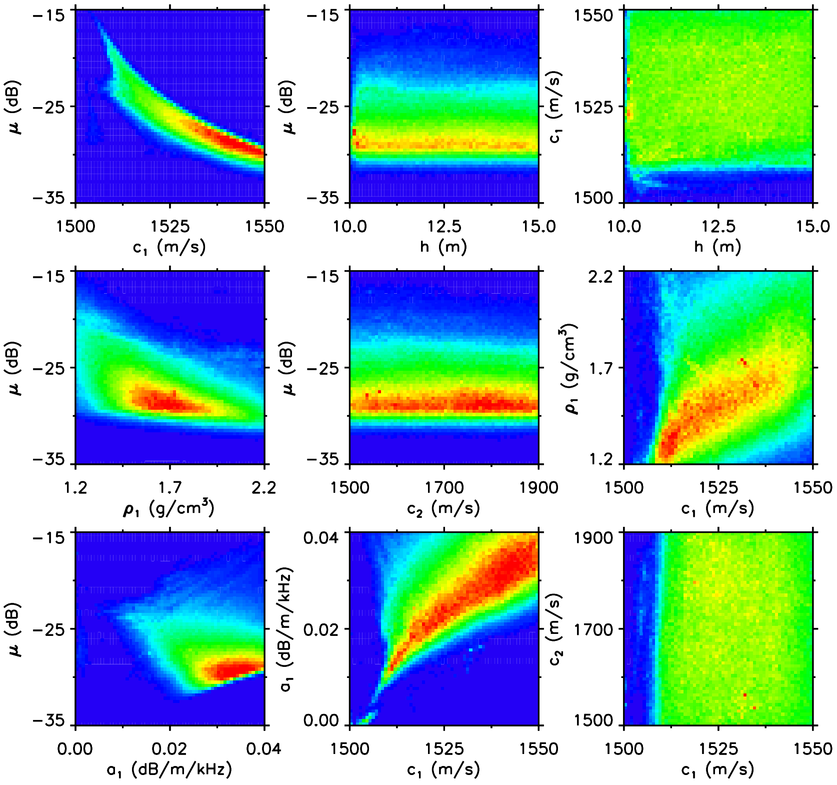


FIG. 10. (Color online) Selected joint marginal probability distributions from inversion of Malta-Plateau reverberation data.

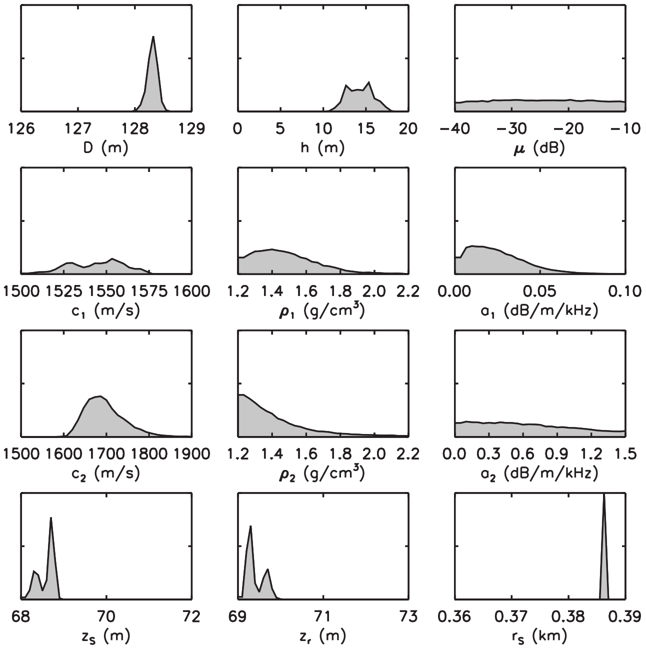


FIG. 11. Marginal posterior probability distributions from inversion of Malta-Plateau short-range propagation data.

basement density is expected to be an insensitive parameter for this combination of sediment thickness and data frequencies. Figure 11 illustrates the nonlinearity of matched-field inversion, with strong multi-modal distributions for z_s and z_r, and weaker multi-modality indicated for h and c_1. Table I summarizes the MAP and mean parameter estimates and uncertainties.

The correlation matrix for propagation inversion, shown in Fig. 12, indicates very different inter-parameter relationships from reverberation inversion (Fig. 9). Figure 12 indicates a strong positive correlation between sediment thickness h and sound speed c_1 (likely related to the fact that acoustic transit time through the layer remains unchanged by increasing or decreasing both parameters). The figure also indicates a significant negative correlation between c_1 and ρ_1 and a weak positive correlation between c_1 and c_2 (likely related to the fact that reflection coefficient depends on the product $\rho_1 c_1$ and on the contrast between c_1 and c_2). Finally, a strong negative correlation is indicated between the source and receiver depths, z_s and z_r. Figure 13 illustrates joint marginal distributions for selected parameter pairs (same parameters and same bounds as Fig. 10). The correlation between c_1 and h leads to a particularly-narrow, curved ridge of high probability. Multi-modal behavior for h and c_1 is evident in several of the joint marginals (e.g., h-μ and c_1-c_2).

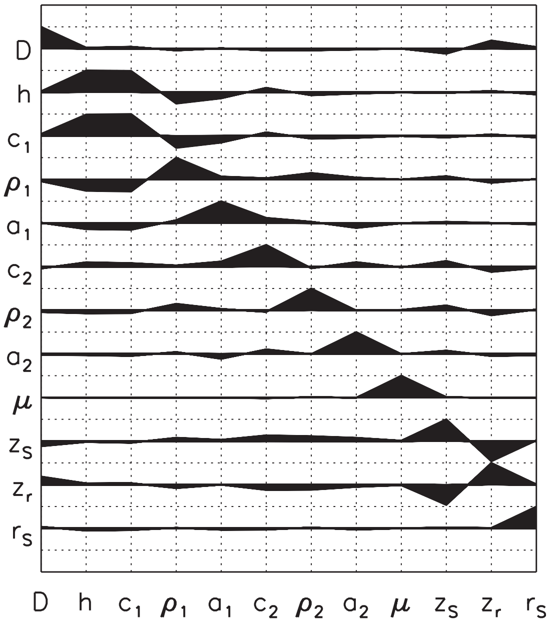


FIG. 12. Correlation matrix from inversion of Malta-Plateau short-range propagation data.

D. Joint reverberation/propagation inversion

One of the goals of this work is to investigate the efficacy of combined reverberation and propagation inversion to resolve geoacoustic and scattering parameters. To this end, the reverberation and propagation data were inverted together, as described in Sec. II B. Figure 14 shows marginal probability distributions computed for joint inversion. In comparison to the marginals obtained from separate inversions of reverberation and propagation data (Figs. 8 and 11, respectively), all environmental parameters are more highly resolved in Fig. 14 (with the exception of ρ_2, which is dubious in all cases). To more clearly compare inversion results for the environmental parameters, Fig. 15 plots the environmental marginal distributions for the three inversions together on the same scale, and Table I summarizes MAP and mean parameter estimates and uncertainties from all three inversions. In particular, Fig. 15 and Table I indicate that μ is much better resolved by joint inversion than by reverberation-only inversion, with a mean estimate of $\bar{\mu}=-24\pm 4$ dB. Further, sediment thickness and sound speed, h and c_1, are significantly better resolved by joint inversion than by propagation-only inversion.

The correlation matrix for joint inversion, shown in Fig. 16, indicates some of the inter-parameter relationships exhibited by reverberation and propagation inversions (Figs. 9 and 12, respectively). For example, for joint inversion c_1 is positively correlated with h and with c_2 (as in propagation inversion) and negatively correlated with a_1 and with μ (as in reverberation inversion). However, the correlation for c_1 and ρ_1, which was positive for reverberation inversion and negative for propagation inversion, is near-zero for joint inversion. Further, the negative correlations of μ with ρ_1 and with a_1 in reverberation inversion are not evident in joint inversion. The joint marginal distributions for combined reverberation/propagation inversion, given in Fig. 17, show that even in cases where correlation effects remain (e.g., c_1-μ, c_1-h, and c_1-a_1), the parameter uncertainty distributions are much better constrained than in inversion of reverberation or propagation data alone. Further, the multimodality appears to be significantly suppressed.

E. Residual analysis

The inversion results presented in Secs. IV B and IV D are based on the assumptions of uncorrelated, Gaussian-

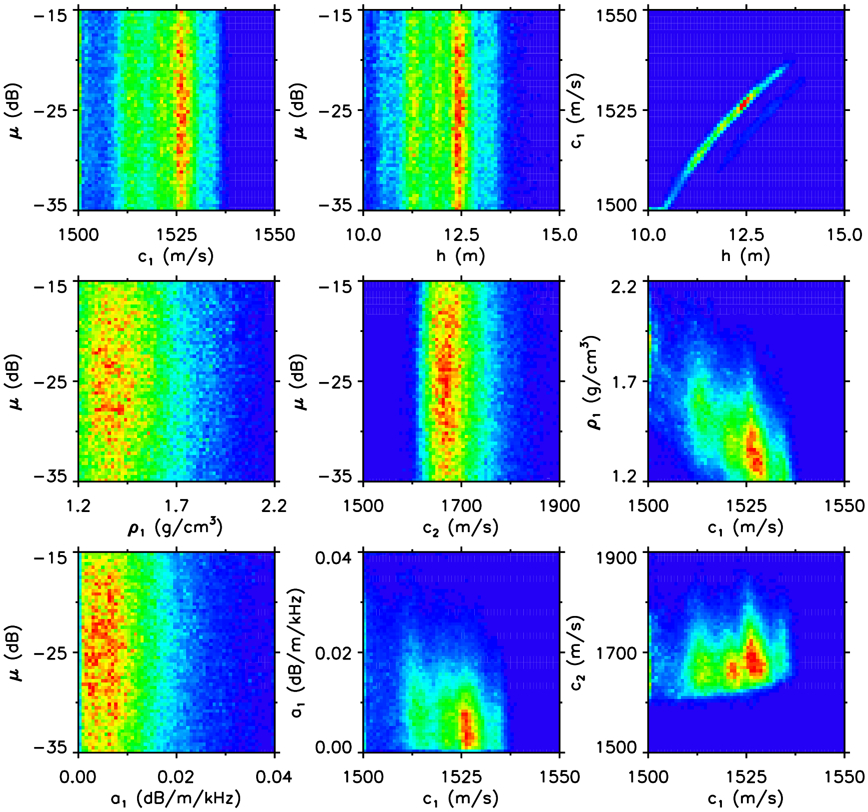


FIG. 13. (Color online) Selected joint marginal probability distributions from inversion of Malta-Plateau propagation data.

distributed errors for the measured reverberation and propagation data. As discussed in Sec. II B, for confidence in the results (particularly uncertainty estimates), these assumptions should be examined *a posteriori* using residual analysis. This analysis is presented here for the joint reverberation/propagation inversion; the results for individual inversions of the two data sets are similar and are not shown.

Qualitative tests for the Gaussianity and randomness of the reverberation data are shown in Fig. 18. Figure 18(a) shows that the histogram of standardized residuals is a good approximation to the standard Gaussian distribution. In particular, there is no evidence of large residual outliers, which are problematic in L_2-norm inversions. The KS test for Gaussianity yielded a p value of 0.3, indicating no evidence against the assumption of Gaussian-distributed errors. Figure 18(b) shows the residual autocorrelation: the narrow central peak suggests uncorrelated residuals. The runs test for randomness yielded $p=0.2$, indicating no evidence against the assumption of uncorrelated errors. Note, in particular, that this analysis supports the treatment of reverberation errors in decibels (i.e., proportional errors), as discussed in Sec. II B.

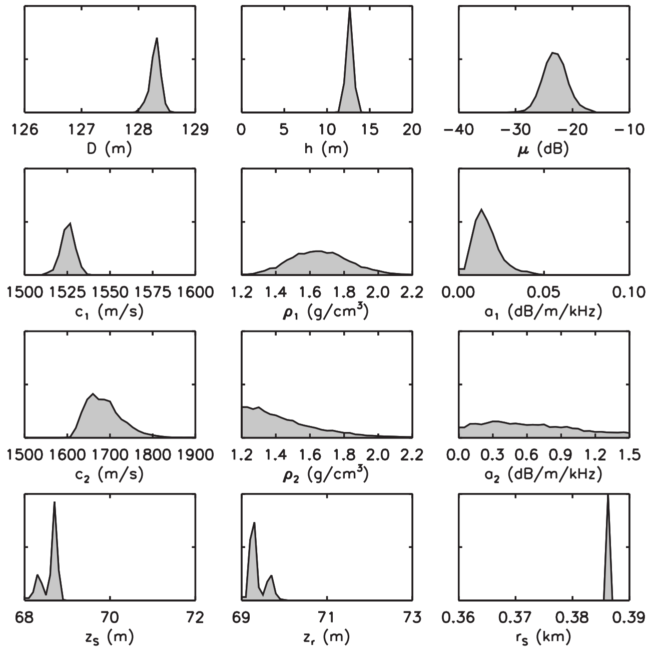


FIG. 14. Marginal posterior probability distributions from joint inversion of Malta-Plateau reverberation and short-range propagation data.

A similar analysis is shown in Fig. 19 for the (complex) propagation data. Figure 19(a) shows that the residual histogram (including real and imaginary parts) is a reasonably good approximation to the standard Gaussian; the KS test yielded $p=0.07$ indicating no significant evidence against the Gaussian assumption. Figure 19(b) shows that the residual autocorrelation (sum of autocorrelations of real and imaginary parts) has a narrow central peak; the runs test provided no evidence of serial correlations with $p=0.9$.

V. SUMMARY AND DISCUSSION

This paper considered Bayesian inversion of through-the-sensor reverberation and short-range propagation data measured on the Malta Plateau, with the goal of estimating seabed geoacoustic and scattering parameters and understanding their uncertainties. The inversion employed hybrid optimization and MHS in a principal-component parameter space to compute properties of the PPD, including MAP and

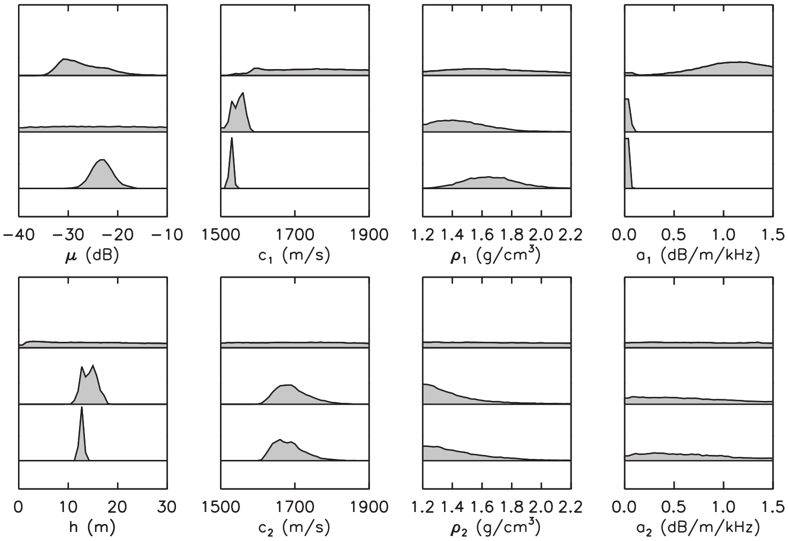


FIG. 15. Comparison of marginal probability distributions for environmental parameters from inversion of reverberation data (top distributions in each panel), short-range propagation data (middle distributions), and joint reverberation plus propagation data (bottom distributions).

mean parameter estimates, parameter variances, marginal and joint marginal probability distributions, and interparameter correlations. The results indicated that (single-frequency, from a 200-Hz band average) reverberation data alone cannot resolve seabed parameters due to strong interparameter correlations inherent in the physics of reverberation. Inversion of frequency-coherent propagation data provided improved resolution of geoacoustic parameters, but is insensitive to scattering strength. However, joint inversion of reverberation and short-range propagation data, with each data set weighted according to its ML variance estimate, provided good resolution of scattering strength and seabed sound-speed structure, and some indication of sediment attenuation and density. The combined inversion provided significantly better resolution for some parameters than either data set alone due largely to the different multi-dimensional parameter correlations imposed by the differing physics for the two data types. The assumptions of uncorrelated, Gaussian-distributed random errors (in decibels for reverberation data) were validated by *a posteriori* residual statistical tests. These inversion results, while representative of only a small data set, illustrate the deficiencies of reverberation-only inversion but the significant promise for combined reverberation and propagation inversion.

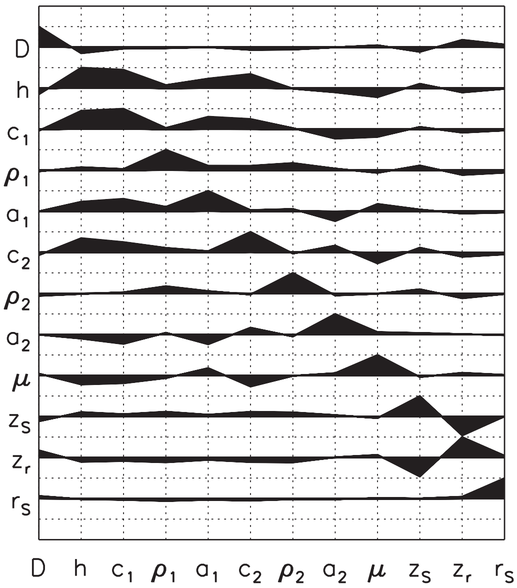


FIG. 16. Correlation matrix from joint inversion of Malta-Plateau reverberation and propagation data.

The seabed sound-speed profile obtained here is in reasonably good agreement with previous inversion results near the measurement site, taking into account the different frequency bands of the various measurements. Figure 20 compares the seabed sound-speed profile obtained in this study to matched-field and reflection-scattering inversion results. The joint reverberation-propagation inversion is represented by the mean profile (Table I) since the mean is generally a good estimator for unimodal distributions. The matched-field results were obtained by Siderius *et al.*[3] and Fallat *et al.*,[4] who inverted identical data from the MAPEX experiment based on a 250–850-Hz signal recorded on a 64-element, 252-m towed array. The reflection-scattering inversion was carried out by Holland[36] for the Boundary 2004 experiment at 1–6 kHz. This latter method has a significantly smaller seafloor footprint and hence is sensitive to local structure that may not be present or may be averaged out in the other inversions (e.g., the thin, high-speed layer at about 1-m depth in this result). This is particularly true for the joint inversion developed in this paper, which assumes range independence to 10 km for the reverberation modeling. Of the four results in Fig. 20, only the reverberation-propagation inversion provides uncertainty estimates (indicated by shaded region in Fig. 20). The value for Lambert's scattering coefficient estimated by the joint reverberation-propagation inversion is $\mu = -24 \pm 4$ dB.

In interpreting the results in Fig. 20, it is important to keep in mind that the homogeneous-layer structure is imposed by the parametrization and that the model provides the best representation, for the particular frequency band, of the actual depth-dependent sound-speed function of the seabed. The (mid-frequency) reverberation-propagation inversion will be more sensitive to the shallow sound-speed structure than the (low-frequency) matched-field inversion, but not as

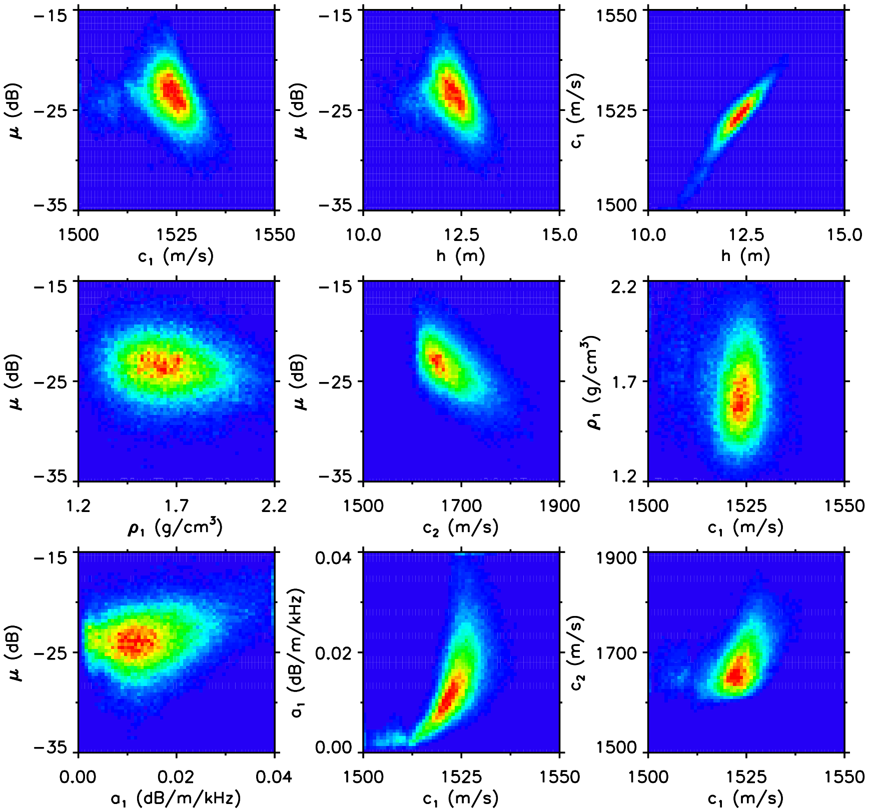


FIG. 17. (Color online) Selected joint marginal probability distributions from joint inversion of Malta-Plateau reverberation and propagation data.

sensitive as the (high-frequency) reflection-scattering inversion. For instance, the reverberation-propagation result for the uppermost sound speed is intermediate between the reflection-scattering result and the matched-field result; this is consistent with frequency-dependent sensing of a low surficial sound speed that increases rapidly over the near-surface sediments. Further, the mid-frequency inversion reacts more quickly to a sound-speed increase with depth than the low-frequency inversion, but not as quickly as the high-frequency inversion. The sound speeds at depth of all methods are similar.

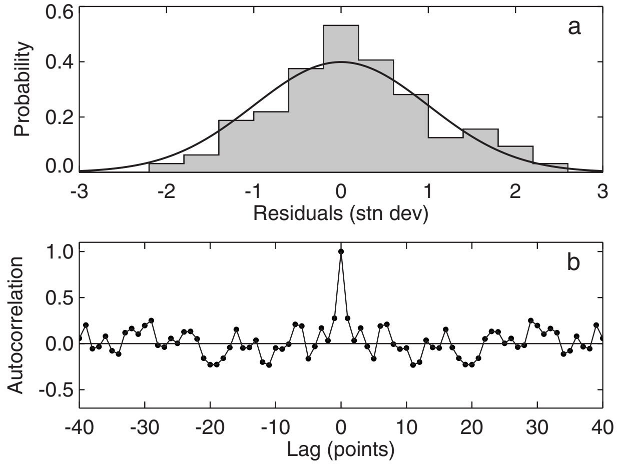


FIG. 18. (a) Histogram of standardized reverberation residuals compared to Gaussian distribution (solid line). (b) Autocorrelation of reverberation residuals (for clarity, correlations for only the central ±40 lag points are shown).

Finally, it should be emphasized that the results in this paper were obtained under the assumption of Lambert's law scattering, which is consistent with previous reverberation inversion work.[6,7,10,13] However, Holland[9,14] demonstrated that reverberation results can differ significantly under other choices for the scattering kernel, and there is not, at present, consensus on this choice. Further, this initial work has concentrated on the inversion of frequency-averaged reverberation data for a single scattering coefficient, and not considered possible frequency and depth dependencies. These additional sources of uncertainty are beyond the scope of the present study, but represent important issues in reverberation inversion.

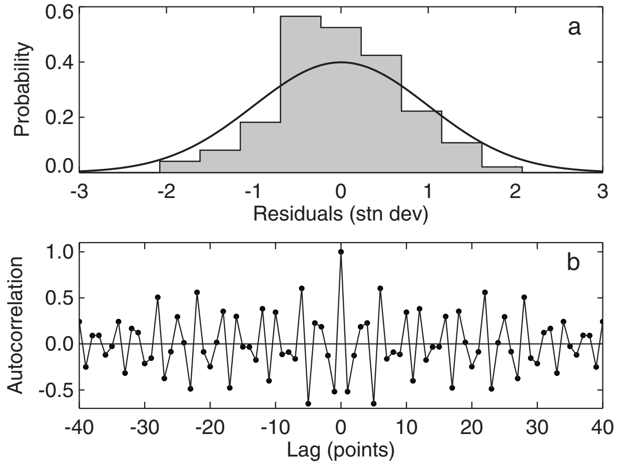


FIG. 19. (a) Histogram of standardized propagation residuals compared to Gaussian distribution (solid line). (b) Autocorrelation of propagation residuals (for clarity, correlations for only the central ±40 lag points are shown).

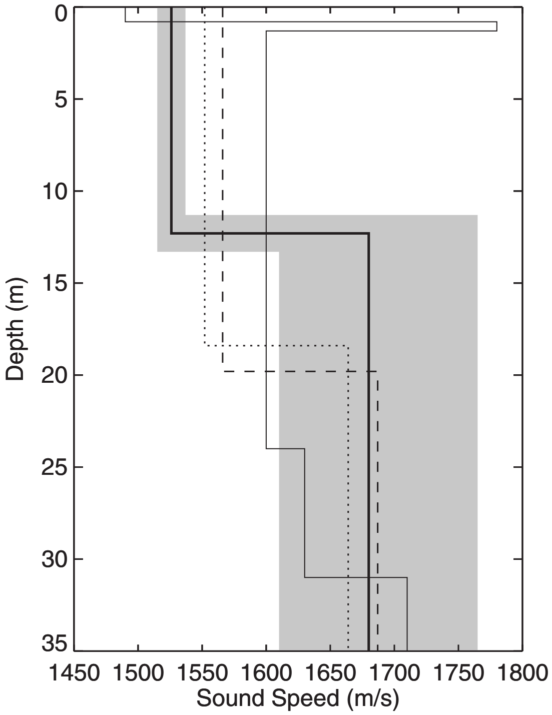


FIG. 20. Comparison of seabed sound-speed profile estimated in this paper at 900–1700 Hz (heavy solid line with shaded uncertainty estimate) and those estimated via matched-field inversion at 250–850 Hz by Siderius *et al.* (Ref. 3) (dotted line) and Fallat *et al.* (Ref. 4) (dashed line) and via reflection-scattering inversion at 1–6 kHz by Holland (Ref. 36) (light solid line).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Georgios Haralabus (Scientist-in-Charge), Piero Boni (Data Acquisition), and Luigi Troiano (Engineering Coordinator) during the BASE'04 trial.

[1] A. Caiti, S. M. Jesus, and A. Kristensen, "Geoacoustic seafloor exploration with a towed array in an shallow water area of the Strait of Sicily," IEEE J. Ocean. Eng. **21**, 355–366 (1996).

[2] S. M. Jesus and A. Caiti, "Range-dependent seabed characterization by inversion of acoustic data from a towed array," J. Comput. Acoust. **4**, 273–290 (1996).

[3] M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).

[4] M. R. Fallat, S. E. Dosso, and P. L. Nielsen, "An investigation of algorithm-induced variability in geoacoustic inversion," IEEE J. Ocean. Eng. **29**, 78–87 (2004).

[5] 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**, 198–199 (2005).

[6] J. D. Bishop, M. T. Sundvik, and F. W. Grande, "Inverting seabed parameters from reverberation data," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachok, A. Carini, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, 1995), pp. 401–406.

[7] D. D. Ellis and P. Gerstoft, "Using inversion techniques to extract bottom scattering strengths and sound speeds from shallow water reverberation data," in Third European Conference on Underwater Acoustics, edited by J. S. Papadakis, Heraklion, Greece (1996), pp. 557–562.

[8] C. H. Harrison, "Closed-form expressions for ocean reverberation and signal excess with mode stripping and Lambert's law," J. Acoust. Soc. Am. **114**, 2744–2756 (2003).

[9] C. W. Holland, "On errors in estimating seabed scattering strength from long-range reverberation (L)," J. Acoust. Soc. Am. **118**, 2787–2790 (2005).

[10] J. R. Preston, D. D. Ellis, and R. C. Gauss, "Geoacoustic parameter extraction using reverberation data from the 2000 Boundary Characterization Experiment on the Malta Plateau," IEEE J. Ocean. Eng. **30**, 709–732 (2005).

[11] C. H. Harrison and P. L. Nielsen, "Separability of seabed reflection and scattering properties in reverberation inversion," J. Acoust. Soc. Am. **121**, 108–110 (2007).

[12] M. A. Ainslie, "Observable parameters from multipath bottom reverberation in shallow water," J. Acoust. Soc. Am. **121**, 3363–3376 (2007).

[13] J. R. Preston, "Using triplet arrays for broadband reverberation analysis and inversions," IEEE J. Ocean. Eng. **32**, 879–896 (2007).

[14] C. W. Holland, "Fitting data, but poor predictions: Reverberation prediction uncertainty when seabed parameters are derived from reverberation measurements," J. Acoust. Soc. Am. **123**, 2553–2562 (2008).

[15] S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion I: A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).

[16] S. E. Dosso and P. L. Nielsen, "Quantifying uncertainty in geoacoustic inversion II: Application to broadband, shallow-water data," J. Acoust. Soc. Am. **111**, 143–159 (2002).

[17] S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2007).

[18] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).

[19] M. K. Sen and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion* (Elsevier, Amsterdam, 1995).

[20] W. R. Gilks, S. Richardson, and G. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

[21] J. J. K. O Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer-Verlag, New York, 1996).

[22] S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).

[23] D. J. 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).

[24] E. K. Westwood and P. J. Vidmar, "Eigenray finding and time series simulation in a layered-bottom ocean," J. Acoust. Soc. Am. **81**, 912–924 (1987).

[25] X. Lurton and J. Marchal, "Long-range propagation losses and reverberation levels in shallow water using an averaged intensity model," in Proceedings of the Third European Conference on Underwater Acoustics, edited by J. S. Papadakis, Heraklion, Greece (1996), pp. 569–574.

[26] J.-X. Zhou and X.-Z. Zhang, "Shallow-water reverberation and small-angle bottom scattering," in Proceedings of the International Conference on Shallow-Water Acoustics, edited by X.-Z. Zhang and J.-X. Zhou, Beijing, China (1997), pp. 315–322.

[27] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994), pp. 53.

[28] E. I. Thoros and J. S. Perkins, "Overview of the reverberation modelling workshop," in Proceedings of the International Symposium on Underwater Reverberation and Clutter, edited by P. L. Nielsen, C. Harrison, and J.-C. Le Gac, La Spezia, Italy (2008), pp. 3–14.

[29] M. Ainslie, W. Boek, P. L. Nielsen, and C. Harrison, "The role of benchmarks in the inversion of low frequency bottom reverberation," in Proceedings of the Second International Conference on Underwater Acoustic Measurements: Technologies and Results, edited by P. L. Nielsen, J. S. Papadakis, and L. Buørnø, Heraklion, Greece (2007), pp. 15–22.

[30] C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **6**, 1–28 (2000).

[31] S. E. Dosso and M. J. Wilmut, "Estimating data uncertainty in matched-field geoacoustic inversion," IEEE J. Ocean. Eng. **31**, 470–479 (2006).

[32] S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).

[33] D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis* (Wiley, New York, 1992).

[34] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier, New York, 2005).

[35] D. E. Weston, "Intensity-range relations in oceanographic acoustics," J. Sound Vib. **18**, 271–287 (1971).

[36] C. W. Holland, "Coupled scattering and reflection measurements in shallow water," IEEE J. Ocean. Eng. **27**, 454–470 (2002).

The impact of ocean sound speed variability on the uncertainty of geoacoustic parameter estimates

Yong-Min Jiang[a] and N. Ross Chapman
School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055, Victoria, British Columbia V8W 3P6, Canada

(Received 12 November 2008; revised 19 February 2009; accepted 19 February 2009)

This paper investigates the influence of water column variability on the estimates of geoacoustic model parameters obtained from matched field inversions. The acoustic data were collected on the New Jersey continental shelf during shallow water experiments in August 2006. The oceanographic variability was evident when the data were recorded. To quantify the uncertainties of the geoacoustic parameter estimates in this environment, Bayesian matched field geoacoustic inversion was applied to multi-tonal continuous wave data. The spatially and temporally varying water column sound speed is parametrized in terms of empirical orthogonal functions and included in the inversion. Its impact on the geometric and geoacoustic parameter estimates is then analyzed by the inter-parameter correlations. Two different approaches were used to obtain information about the variation of the water sound speed. One used only the profiles collected along the experimental track during the experiment, and the other also included observations collected over a larger area. The geoacoustic estimates from both the large and small sample sets are consistent. However, due to the diversity of the oceanic sound speed, more empirical orthogonal functions are needed in the inversion when more sound speed profile samples are used.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097770]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 2881–2895

I. INTRODUCTION

The prediction of sound propagation in the ocean waveguide depends on our knowledge of the ocean environment. The sea surface, water column, and sea bottom may have different influences on different underwater acoustic applications. For low frequency (LF), long range propagation in shallow water, seabed geoacoustic properties usually play a dominant role due to the multiple interactions of sound with the sea bottom. Nevertheless, it is also true that water column inhomogeneities can influence the sound propagation. Hence, determining geoacoustic properties in a variable ocean environment is a practical issue for applications such as sound propagation interpretation and sonar performance prediction. Studies of the effect of the water column variability on shallow water acoustics and geoacoustic inversion can be found in Refs. 1–8 and the works referred therein.

Matched field geoacoustic inversion (MFI) is a widely adopted technique to infer seabed properties. It is a model based method that makes use of the sensitivity of the acoustic field to the ocean environment and the seabed properties. MFI has been successfully applied to experimental data as well as benchmark synthetic problems, both in range independent and range dependent waveguides.[9,10] Most of the inversion studies to date were carried out by assuming that the water column sound speed profile (SSP) is both range independent and time invariant. In this paper, we investigate the impact of temporal and spatial variations of the water column SSP on geoacoustic inversion. We assume that the SSP is represented by an effective range independent profile that is generated from information from different oceanographic observations in the vicinity. Results are reported for MFI in a temporally and spatially varying ocean waveguide at different ranges. Both the acoustic and the oceanographic data analyzed here were collected during the shallow water experiments 2006 (SW06) carried out on the continental shelf off New Jersey.

Spatial and temporal variations of the oceanic SSP along the propagation path are important factors that cause acoustic field fluctuations and reduce the signal coherence at the receiver. Consequently, these affect sonar performance in applications for underwater communication and source localization. Spatial and temporal variations of the oceanic SSP will also contaminate geoacoustic inversion results. Siderius *et al.*[4,5] examined geoacoustic inversion uncertainties with a vertical line array (VLA) at different ranges in a time varying environment. They applied an optimization inversion approach to data collected in the Strait of Sicily over a time period of several hours to tens of hours with a fixed experimental geometry. Their study showed that the sound field fluctuations over the observation time increased the standard deviations of the geometric/geacoustic estimates over the range. Strong influence of a dynamic ocean SSP on geoacoustic inversion was also observed when conventional MFI was applied to the data recorded during SW06. Huang *et al.*[7] and Jiang and Chapman[8] reported results for matched field inversions of multi-tonal continuous wave (CW) data from the same experiment in SW06. Both groups observed that the conventional approach of using a single SSP measured at a specific site and time in the experiment failed to generate the correct experimental geometry in the inversions, and con-

[a] Author to whom correspondence should be addressed. Electronic mail: minj@uvic.ca

cluded that the performance degradation was due to the dynamically varying SSP. To mitigate the impact of the dynamic SSP on the geoacoustic inversions, the ocean SSP was parametrized in terms of empirical orthogonal functions (EOFs) and included as a search parameter in MFI of the data. The estimated geoacoustic profiles from the two inversions were consistent, and both groups concluded that inverting the ocean SSP along with the geoacoustic/geometric parameters improved the inversion performance. However, Jiang and Chapman[8] assumed that the ocean SSP in the variable ocean could be modeled adequately with a single SSP for the entire propagation range.

In this paper, we investigate the limits of using a single SSP that represents the temporally and spatially varying SSP along the track for the inversions of the SW06 data. We compare the effect of two approaches for generating EOFs for constructing the oceanic SSP in MFIs. The first approach used a limited set of SSPs that were taken from the source ship at the time of the signal transmission. This approach includes the relevant variation of the ocean SSP during the experiment. It also models only the variations in the thermocline, which was the portion of the SSP where the greatest variation occurred. Inversions based on this approach were compared with inversions for which the EOFs were constructed from a more extensive set of SSPs from other environmental moorings nearby to capture a wider degree of sound speed variation. The inversions were carried out at different ranges to investigate the range dependence of the geoacoustic model.

The remainder of this paper is organized as follows. Section II describes the experimental site, the source and the receiver geometry, and the ocean environment. Section III briefly reviews Bayesian matched field inversion framework and estimation of data covariance matrices by the use of multiple data segments. Section IV describes Bayesian inversion of SW06 data, including data processing procedure, seabed, and environment parametrization. Section V compares Bayesian inversion results by using different oceanographic observations, and discusses the impact of the temporally and spatially varying SSP on geoacoustic inversion results and the sensitivity of geoacoustic model parameter at different ranges. Section VI summarizes the results of this work.

II. DESCRIPTION OF THE EXPERIMENT AND THE ENVIRONMENT

SW06 was a series of multidisciplinary shallow water experiments carried out near the shelf break on the New Jersey continental shelf from mid-July to mid-September of 2006.[11] Physical oceanographic, ocean acoustic, and geoacoustic experiments were carried out simultaneously to obtain the necessary acoustic and environmental data to investigate the impact of dynamic variations in the water column SSP on sound propagation and geoacoustic inversion.

This paper focuses on the LF CW tonal data transmitted at different ranges and measured at a bottom moored VLA deployed by the Marine Physical Laboratory (MPL). Figure 1 depicts the experimental site where the data for this study were collected. The data were transmitted from way points (WPs) 21, 22, and 23 and recorded at MPL VLA1 on JD239 (Julian Day). The locations of MPL-VLA1 and the source stations are listed in Table I. At each station, a 5-min LF multi-tonal set at 53, 103, 203, and 253 Hz were transmitted first, followed by a 5-min mid-frequency (MF) multi-tonal set at 303, 403, 503, 703, and 953 Hz. The time period between the first transmission at WP21 and last transmission at WP23 was Greenwich mean time (GMT) 21:54–23:35. During the transmissions, the source ship R/V Knorr was maintained at pre-scheduled positions by use of the dynamic positioning system. The ranges from WP21 to VLA1 were maintained at 1.007 ± 0.001 km, from WP22 at 3.011 ± 0.001 km and from WP23 at 5.018 ± 0.001 km according to the navigation system on board.[12] The bathymetry along the track between VLA1 and WP23 was generally flat with slight water depth variation. The statistics of water depths measured by the 12 kHz echo sounder on R/V Knorr are shown in Figs. 2(a)–2(c). Figure 2(a) shows the histogram of the water depths between WP19 (230 m away from VLA1) and WP21; Fig. 2(b) shows the water depth histogram between WP21 and WP22 and Fig. 2(c) shows the histogram for water depths between WP22 and WP23. The mean values indicate that the water depth becomes slightly deeper from WP21 to WP23, with an increase of about 1.6 m over 5 km. The water depth at VLA1 measured on JD237 (2 days before the data analyzed here collected) was 78.8 ± 0.2 m. The maximum fluctuation of water depth at the same location is 1.5 m if the diurnal tidal effect is considered. The source depth was monitored continuously[12]

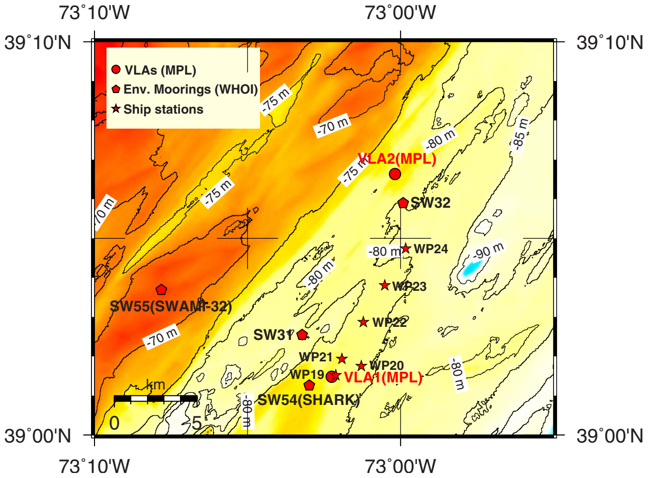


FIG. 1. (Color online) Bathymetry of the experimental site and the positions of the source, acoustic arrays, and environmental array moorings.

TABLE I. Information of MPL-VLA1 and the source stations.

Station name	Latitude (N)	Longitude (W)	Distance to MPL-VLA1 (km)	Transmission start time (GMT)
MPL-VLA1	39°01.477′	73°02.256′	0.000	
WP19	39°01.520′	73°02.109′	0.230	19:16
WP21	39°01.932′	73°01.914′	1.000	21:54
WP22	39°02.868′	73°01.218′	3.000	22:47
WP23	39°03.804′	73°00.522′	5.000	23:29

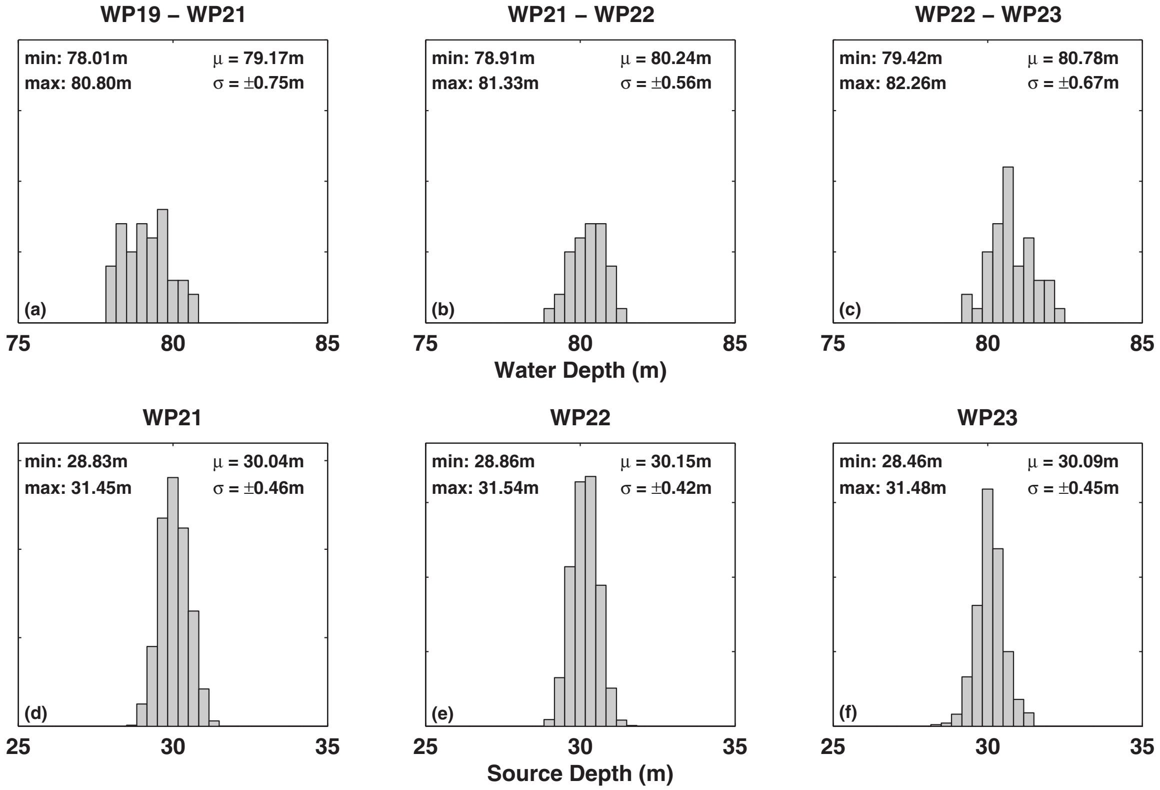


FIG. 2. Histograms of the water depth and source depth measurements. [(a)–(c)] Water depth measured by the shipborne echo sounder between the ship stations; [(d)–(f)] Source depth measured at the ship stations during the source transmission.

(two samples per minute) and the statistics at the three WPs during the source transmission are shown in Figs. 2(d)–2(f), respectively.

Figure 3 is a schematic of the experimental geometry. Sixteen hydrophones were equally spaced at 3.75 m apart on VLA1. The distance from the bottom most hydrophone to the sea bottom was 8.2 m. A tilt/pressure/temperature sensor was located 0.5 m above the top most hydrophone. The data from this sensor indicate that the tilt of VLA was approximately from 0° to 5° during the source transmissions.

The spatial and temporal variations of the water column SSP are evident in Fig. 3 from the measurements made at the WPs. Figure 4(a) is an overlay plot of all the profiles. The times of the SSP being measured were aligned with respect to the SSP measured earlier at 19:17 GMT on JD239 at WP19. Substantial SSP variations are evident at the depths

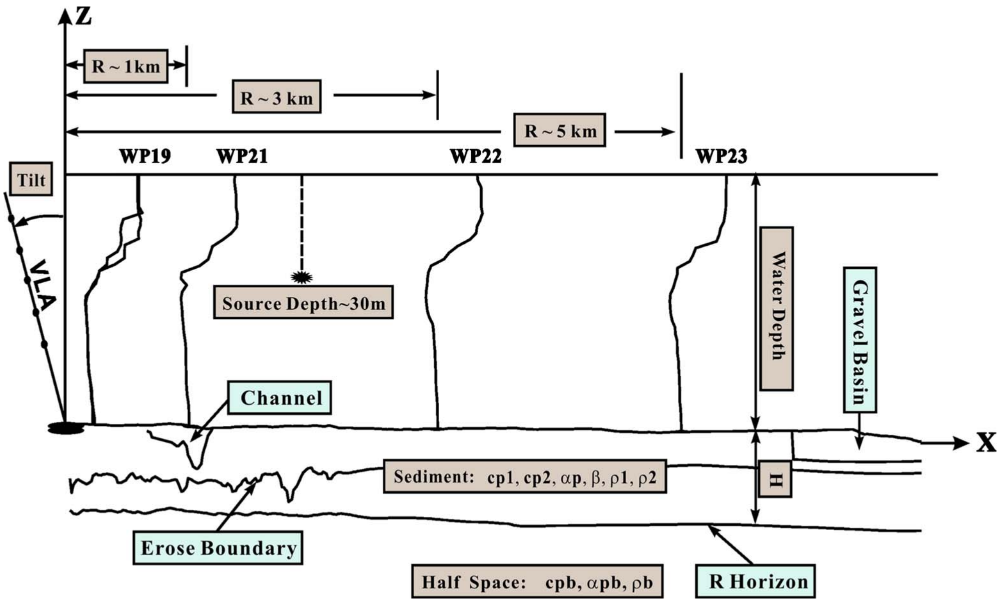


FIG. 3. (Color online) Sketch of the experimental geometry and geoacoustic model. The sediment structure is plotted according to the interpolated seismic data along the source track provided by the geophysical seismic survey in the vicinity (Ref. 29).

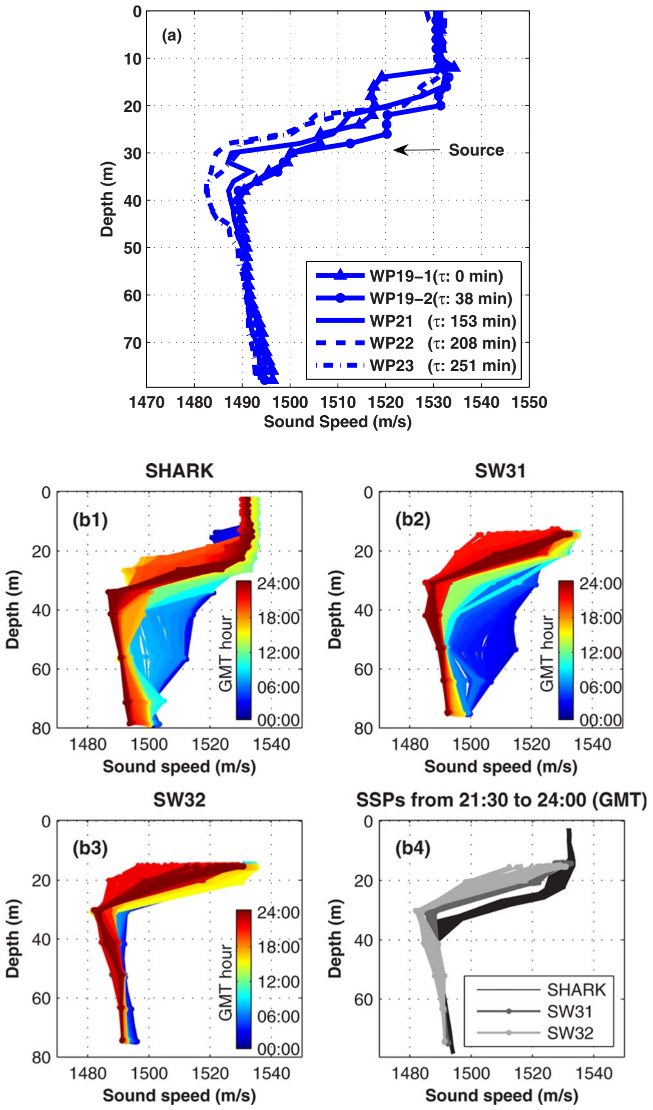


FIG. 4. (Color online) (a) CTD casts measured at the source stations during the experiment. (b) Oceanographic observations recorded on WHOI environmental arrays on JD239.

from 10 to 50 m over 4 h in time and 5 km in range. The source depth was around 30 m, in the middle of the thermocline. The water column sound speed changed considerably at this depth over the observation time.

There were a large number of Woods Hole Oceanographic Institution (WHOI) environmental array moorings deployed over the experimental region during SW06. The pentagons in Fig. 1 are the environmental array moorings (SW31, SW32, and SW54) that were closest to the source track. SSPs recorded at those three locations on JD239 are shown in Fig. 4(b)(1) to Fig. 4(b)(3). SSPs at SW54 (SHARK) were derived from the environmental array moorings nearby.[13] The SSPs at SW31 and SW32 were determined from the measured conductivity-temperature-depth (CTD) data[13] and the latitude of the experimental site according to Refs. 14–18. The SSPs were sampled in time at a rate of 30 s. In Fig. 4, the sample time was color coded to represent the SSPs measured at different times. It is clear that the SSPs at the three locations varied significantly. Figure 4(b)(4) shows all of the SSPs collected at the three locations during the CW tonal transmission from 21:30 to 24:00 GMT on JD239. Notable differences in the thermocline are seen from the SPPs measured at SW54 and SW31/SW32. The differences can be characterized in terms of the sound speed gradient and the depth of the thermocline.

III. BAYESIAN MATCHED-FIELD GEOACOUSTIC INVERSION THEORY

A. Bayes' rule

For the completeness of the paper, this section outlines the Bayesian approach to MFI. More detailed treatments of Bayesian theory and its applications to MFI can be found in Refs. 19–21 and the references therein. Let $\mathbf{m}$ be the vector of model with M parameters to be estimated, and $\mathbf{d}$ be the data vector. The elements of both $\mathbf{m}$ and $\mathbf{d}$ are considered to be random variables. Bayes' rule can be written as

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \tag{1}$$

where $P(\mathbf{d}|\mathbf{m})$, the conditional probability density function (PDF) of $\mathbf{d}$ given $\mathbf{m}$, represents the data information; $P(\mathbf{m})$, the PDF of $\mathbf{m}$, represents model prior information and is independent of $\mathbf{d}$; and $P(\mathbf{d})$ is the PDF of $\mathbf{d}$ and is a fixed constant when the data are measured. $P(\mathbf{m}|\mathbf{d})$ is the conditional PDF of $\mathbf{m}$ given $\mathbf{d}$, and it represents the model information that incorporates data [i.e., $P(\mathbf{d}|\mathbf{m})$] and model prior [i.e., $P(\mathbf{m})$] information. $P(\mathbf{m}|\mathbf{d})$ is also called the posterior probability density (PPD), which is the general solution to the inversion problem in the Bayesian formulation.

Interpreting $P(\mathbf{d}|\mathbf{m})$ in terms of observed data as a function of model parameter $\mathbf{m}$ defines the likelihood function $L(\mathbf{m})$, which can generally be written as $L(\mathbf{m}) \propto \exp[-E(\mathbf{m})]$, where $E(\mathbf{m})$ is an appropriate data misfit function. If data and model prior information are combined as a generalized misfit $\Phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m})$, then Eq. (1) can be written as

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\Phi(\mathbf{m})]}{\int_M \exp[-\Phi(\mathbf{m}')]d\mathbf{m}'}. \tag{2}$$

There is no restriction on prior information being used in the computation of PPD. Prior information may be represented by any distribution. In this approach, uniform distributions of the model parameter values are assumed.

Due to its multidimensional nature, the full solution to the inverse problem, PPD, is generally interpreted in terms of properties such as maximum *a posteriori* (MAP), posterior mean estimate, marginal probability distribution of model parameter m_i, where $i = 1, \ldots, M$, and model covariance matrix, which are defined as follows:

$$\hat{\mathbf{m}}_{\text{MAP}} = \text{arg}_{\max}\{P(\mathbf{m}|\mathbf{d})\}, \tag{3}$$

$$\overline{\mathbf{m}}_{\text{mean}} = \int \mathbf{m}P(\mathbf{m}|\mathbf{d})d\mathbf{m}, \tag{4}$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i' - m_i)P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{5}$$

$$\mathbf{C_m} = \int (\mathbf{m} - \overline{\mathbf{m}})(\mathbf{m} - \overline{\mathbf{m}})^T P(\mathbf{m}|\mathbf{d}) d\mathbf{m}, \quad (6)$$

where δ in Eq. (5) is the Dirac delta function, and the superscript T in Eq. (6) (and in the equations in the remainder of this paper) represents transpose operation. While Eqs. (3) and (4) represent the "point" estimates of the model, Eq. (5) reveals the uncertainty estimate of the model parameter. One may use the interval that contains a certain percentage of the area of the marginal distribution (credibility intervals) to quantify the uncertainties of the model parameter estimates. Higher dimensional (joint) marginal distributions can be defined in a similar manner to Eq. (5). Inter-parameter correlations are obtained from the correlation matrix, whose elements are defined as

$$R_{ij} = c_{ij}^m / \sqrt{c_{ii}^m c_{jj}^m}, \quad (7)$$

where R_{ij}'s are unity for the diagonal elements and have the values within $[-1, 1]$ for the off-diagonal elements. c_{ij}^m are the elements of the model covariance matrix $\mathbf{C_m}$.

For linear inversion problems, analytic solutions of Eqs. (3)–(5) exist. However, for nonlinear inversion problems such as MFI, analytic solutions do not exist and numerical integration must be applied.

B. Likelihood and misfit function for MFI

In matched field inversion, the data vector usually is the complex acoustic pressure at frequency f on N_H hydrophones of an array, $\mathbf{d}_f = [d_1, d_2, \ldots, d_{N_H}]^T$; the model vector $\mathbf{m} = [m_1, m_2, \ldots, m_M]^T$ can be any geoacoustic/geometric/water column sound speed parameter to be inverted, and the modeled acoustic pressure computed for model $\mathbf{m}$ at frequency f on an array of hydrophones can be written as $\mathbf{d}_f(\mathbf{m}) = [d_1(\mathbf{m}), d_2(\mathbf{m}), \ldots, d_{N_H}(\mathbf{m})]^T$. If the errors between the measured and modeled data are assumed to be zero mean, Gaussian distributed complex random variables that are uncorrelated from frequency to frequency, the likelihood function for frequency incoherent MFI can be written as

$$L(\mathbf{m}) = \prod_{f=1}^{N_F} \frac{1}{\pi^{N_H} |\mathbf{C}_f|} \exp\{-[\mathbf{d}_f - \mathbf{d}_f(\mathbf{m})]^\dagger \mathbf{C}_f^{-1} [\mathbf{d}_f - \mathbf{d}_f(\mathbf{m})]\}, \quad (8)$$

where "†" represents conjugate transpose, N_F is the number of the frequency components used, and $\mathbf{C}_f$ is the data covariance matrix at frequency f, which is introduced in the likelihood function to account for the spatial correlation between the hydrophones on an array.

For MFI with limited source spectrum information (i.e., amplitude and phase at different frequencies),[22,23] the modeled complex acoustic pressure at frequency f is generally written as

$$\mathbf{d}_f(\mathbf{m}) = A_f e^{i\theta_f} \mathbf{P}_f(\mathbf{m}), \quad (9)$$

where $\mathbf{P}_f(\mathbf{m}) = [P_1(\mathbf{m}), P_2(\mathbf{m}), \ldots, P_{N_H}(\mathbf{m})]^T$ is the complex acoustic pressure on N_H hydrophones of an array calculated by an acoustic propagation model; A_f and θ_f are unknown magnitude and phase information. The complex factor $A_f e^{i\theta_f}$ can be obtained by substituting Eq. (9) into Eq. (8) and maximizing the likelihood function by setting $\partial L / \partial A_f = 0$ and $\partial L / \partial \theta_f = 0$. Substituting the expression of $A_f e^{i\theta_f}$ back into Eq. (8) leads to the likelihood function for MFI with unknown source spectrum

$$L(\mathbf{m}) = \prod_{f=1}^{N_F} \frac{1}{\pi^{N_H}} \exp\left\{ - \left[\mathbf{d}_f^\dagger \mathbf{C}_f^{-1} \mathbf{d}_f - \frac{|\mathbf{P}_f^\dagger(\mathbf{m}) \mathbf{C}_f^{-1} \mathbf{d}_f|^2}{\mathbf{P}_f^\dagger(\mathbf{m}) \mathbf{C}_f^{-1} \mathbf{P}_f(\mathbf{m})} \right] \right\}. \quad (10)$$

The corresponding misfit function is

$$E(\mathbf{m}) = \sum_{f=1}^{N_F} \left[\mathbf{d}_f^\dagger \mathbf{C}_f^{-1} \mathbf{d}_f - \frac{|\mathbf{P}_f^\dagger(\mathbf{m}) \mathbf{C}_f^{-1} \mathbf{d}_f|^2}{\mathbf{P}_f^\dagger(\mathbf{m}) \mathbf{C}_f^{-1} \mathbf{P}_f(\mathbf{m})} \right]. \quad (11)$$

C. Data error covariance matrix estimation

Data error information is one of the most important factors in Bayesian inversion since wrong assumptions or knowledge of the data errors may bias the uncertainty estimation of the model parameters being inverted. Data errors include measurement errors and theory errors. Theory errors arise from the reasons such as inappropriate model parametrization and inaccurate acoustic forward theory. The data errors on the hydrophones of an array may have different means and variances, although the error distribution could be Gaussian. Furthermore, theory errors on the adjacent hydrophones could be correlated. Hence, it is not generally true that the data error covariance matrix has identical independent distribution (i.e., identical values at diagonal elements and zeros on the off-diagonal elements), especially for the cases when the frequencies are low and the hydrophone separations are relatively small.

While the variance of the measurement errors can be reduced by averaging over multiple data samples, we cannot assume that the theory errors can be reduced by simply averaging over multiple transmissions (or segments) of data. However, information about the theory errors may be extracted by changing the experimental geometry (as a result, changing the acoustic signal propagation conditions), such as varying the source-receiver range[24] (the amount of range change should be relatively small compared to the distance between the source and the receiver to ensure the geoacoustic model is valid for all ranges), changing the source depth or receiver depth, or simply maintaining the source and receiver geometry in a highly variable ocean environment. The dominant parameters in the proposed geoacoustic model should produce the acoustic replica that fits the data with small uncertainties under different sound propagating conditions. The differences between the measured and the modeled data from different sound propagation conditions will provide theory error information, so that the ensemble average of these differences may reveal the statistics of the theory errors.

The data analyzed here were collected in rough sea state conditions. The source was moving up and down along with the source ship, and the upper part of the VLA was rotating over the measurement period because of the effect of tidal current and wind on the subsurface torpedo float that was attached to the top of MPL-VLA1. The source depth variations within the strong thermocline caused changes in the sound propagation conditions in the waveguide that resulted in variations of the acoustic field. This variation provided a means for accounting for theory errors in MFI by using multiple data samples.

The data error covariance matrix at frequency f was estimated according to its formal definition, i.e., the ensemble average over multiple optimization realizations,

$$\mathbf{C}_f = \langle(\mathbf{r}_f(\mathbf{m}) - \langle\mathbf{r}_f(\mathbf{m})\rangle)(\mathbf{r}_f(\mathbf{m}) - \langle\mathbf{r}_f(\mathbf{m})\rangle)^\dagger\rangle, \tag{12}$$

where $\mathbf{r}_f$ are the data residuals (the difference between the measured and modeled data) at frequency f given by

$$\mathbf{r}_f(\mathbf{m}) = \mathbf{d}_f - \frac{\mathbf{P}_f^\dagger(\mathbf{m})\mathbf{d}_f}{|\mathbf{P}_f(\mathbf{m})|^2}\mathbf{P}_f(\mathbf{m}). \tag{13}$$

If the model $\mathbf{m}$ is appropriate for representing the data within the observation time, then the residuals at each hydrophone from multiple transmissions or segments can be considered as a stationary process. Therefore, the ensemble average in Eq. (12) can be replaced with the arithmetic average of the residuals on each hydrophone over N_{segt} data segments (data segmentation will be described later in data pre-processing section)

$$\mathbf{C}_f = \frac{1}{N_{\text{segt}}}\sum_{i=1}^{N_{\text{segt}}}\left\{\left[\mathbf{r}_{fi}(\mathbf{m}) - \frac{1}{N_{\text{segt}}}\sum_{j=1}^{N_{\text{segt}}}\mathbf{r}_{fj}(\mathbf{m})\right]\right.$$
$$\left.\times\left[\mathbf{r}_{fi}(\mathbf{m}) - \frac{1}{N_{\text{segt}}}\sum_{j=1}^{N_{\text{segt}}}\mathbf{r}_{fj}(\mathbf{m})\right]^\dagger\right\}. \tag{14}$$

It should be noted that the assumption of stationarity of the residuals at each hydrophone over the observation time should be examined.[24]

IV. BAYESIAN GEOACOUSTIC INVERSION ON SW06 TONAL DATA

Since the bathymetry between the source and MPL-VLA1 is almost flat, range independence is assumed in the inversion. As a result, the objective of inverting the ocean SSP is to search for a range independent effective SSP between the source and the receiver by using the statistical information from a number of ocean sound speed observations. The range independent acoustic propagation model ORCA (Ref. 25) is employed in the inversion. Since shear wave effects are not considered in the inversion, based on the earlier study in the vicinity of this experiment,[24] the real axis option of ORCA is used.

A. Data pre-processing

The time series signal of each VLA hydrophone was first windowed into 2.62-s segments and then a fast Fourier transform was applied to each data segment. The quality of the data was examined at each frequency by checking the signal to noise ratio (SNR). The correlation time of the background noise was around 1.6 s (at $1/e$ power point), which was determined by the time series with the tonal components removed. This correlation time was used to determine the separation of two consecutive data segments in time to ensure the background noise was independent from segment to segment. Since the data of the upper four hydrophones were contaminated by the noise due to the high sea state and the clipping of the recorded signal, only the time segments that had high SNR on the lower 12 hydrophones were chosen. The Bartlett mismatch of the spectral components with respect to the ones in a reference time segment was also considered as one of the criteria to choose the data.

The signals finally used in the inversion span 12 hydrophones over an aperture of 41.25 m, and 8 frequency components: 53, 103, 203, 253, 303, 403, 503, and 703 Hz. The reason for combining the LF and MF frequency components in the inversion is to take the advantage of the depth penetration ability of the LF signal and the sensitivity to the water column SSP of the MF signal, as suggested by the inversions carried out at the earlier stage of the work reported here by using only LF or MF separately.

B. Geoacoustic parametrization

The sea bottom parametrization in this study is based on the geophysical surveys in the vicinity of the track and the analysis of the parameter sensitivity. The track between the WPs and MPL-VLA1 is a track of common focus for the geoacoustic experiments in SW06. Extensive geological and geophysical surveys such as shallow cores, deep drills, *in-situ* sediment probes, grab samples, and high resolution chirp sonar sub-bottom surveys were carried out in the vicinity of the track.[26,27] The sea bottom structure shown in Fig. 3 was provided by geophysical/geological seismic survey[28] along the track in terms of two-way travel time (TWT). The figure shows that the bathymetry is weakly range dependent—the water depth slightly increases when the range increases. The seafloor is at TWT of 103–106 ms along the track. There exists an R reflector at TWT of 130–135 ms throughout the track. A shallower "erose" interface at TWT of 115–128 ms is also found between the sea floor and the R reflector. Besides, WP21 was directly above a filled channel. A 13 m core collected in 2002 indicated that the channel is filled by higher velocity unconsolidated sands, whereas the upper unit of the rest of the track is believed to have the same lower-velocity clay as most of the region of the outer shelf wedge.[27]

Several optimization trials were carried out to study the sensitivity of each parameter at three ranges. At the range of 1 km, the layer thickness is relatively sensitive when a two-layer over half space model was used, while at 3 and 5 km ranges, the layer thickness is not as sensitive even when a one-layer over half space model was used. Based on the geophysical surveys and the parameter sensitivity study, a simple geoacoustic model of one layer over a half space was chosen to represent the sea bottom for all the three sites. The objectives of the inversions were to investigate the resolvability of the R reflector at the sediment and basement interface, and the evolution of the parameter sensitivities over the range in the experiment. To investigate the frequency dependency of the sediment p-wave attenuation, it was parametrized in terms of $\alpha_p(f) = \alpha(f/f_0)^\beta$, where the unit of frequency f is kHz, f_0 is 1 kHz, and α_p is in dB/m. The constant factor α and the exponent β were inverted directly. α_p was computed from each α and β pair and then used in the acoustic model. Similar to the parametrization of attenuation in the earlier inversion approach,[8] α is considered to be homogeneous in the sediment, and the frequency dependence factor β applies to the top of the sediment.

Due to the spatial variability of the geoacoustic parameters over the range and the sediment depth, it is expected that the parameter estimates from the inversion are the effective ones that average over the propagation range and the depth.

C. Water column sound speed parametrization

As mentioned earlier, small and large SSP sample sets were used to compare the effect of the statistics of SSP observations on the estimates uncertainties. To isolate the problem, the small SSP sample set contained only the SSPs collected from CTD measurements on the source ship along the source-receiver track and within or close to the source transmission times. The large SSP sample set contained the SSPs recorded on the SW31, SW32, and SW54 environmental array moorings within the total source transmitting time (21:30–24:00 GMT on JD239) and all of the samples in the small SSP sample set.

First, the SSPs collected from the source stations were downsampled in depth yet dense enough to capture the characteristics of the SSPs. Next, the SSPs from WHOI environmental array moorings were interpolated and extrapolated to match the sample points in depth of SSPs acquired at source stations. Since the SSPs in the small sample set were almost identical from 0–10 m and from 50–79 m [see Fig. 4(a)], only the middle parts (10–50 m) were used in the small sample set analysis. This process essentially added some reasonable constraints in the search of the effective ocean SSP. There were 5 SSPs in the small sample set and 905 SSPs in the large sample set in the EOF analysis described as follows. The number of the depth data points in the small SSP sample set was 23 and in the large SSP sample set was 27.

The EOF representation of the ocean SSP is summarized in the Appendix. One can determine the number of eigenvectors, L, that is statistically significant to match a prescribed degree of fit to the total energy from the relationship

$$I_L^{\Lambda} = \sum_{j=1}^{j\le L\le N_J} \lambda_j/\mathrm{tr}(\mathbf{R}), \tag{15}$$

where λ is the eigenvalue.

Similar to Eq. (15), a misfit ratio is defined in determining P, which is the number of EOFs needed to reconstruct an individual SSP to a prescribed degree

$$I_P^A = [s_P^T(N_Z)s_P(N_Z)]/[(s_k(N_Z)-\bar{s}(N_Z))^T(s_k(N_Z)-\bar{s}(N_Z))], \tag{16}$$

where $s_P(N_Z)=\sum_{j=1}^{j\le P\le N_J} a_{j,k}v_j(N_Z)$. I_P^A is used as an criterion in this study to determine the number of EOFs to be used to reconstruct an individual SSP.

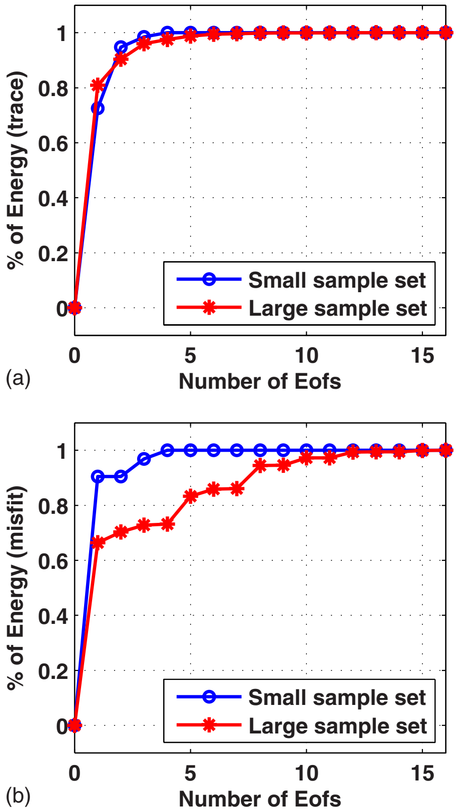


FIG. 5. (Color online) Comparison of the percentage of energy fit versus the number of EOF coefficients used between the small (circled line) and the large (star line) SSP sample sets. (a) Percentage of the energy versus the number of EOFs and (b) percentage of misfit energy over total misfit versus the number of EOFs used to recover individual SSP (CTD of WP19-1 is used as an example).

It should be mentioned that the values of L and P are not necessarily the same because they have different definitions. I_L^{Λ} indicates the number of modes that have statistically significant contributions to the total energy. If the SSPs in $\mathbf{S}(N_Z,N_S)$ are highly correlated, L could be only 2 or 3 in order to get 95% fit to the total energy. On the other hand, I_P^A gives the measure of the difference from one individual SSP $s_k(N_Z)$ is to the background profile $\bar{s}(N_Z)$. If $s_k(N_Z)$ is greatly different from $\bar{s}(N_Z)$, P is usually larger than L. Each individual SSP may generate different P and I_P^A. In other words, if the effective SSP that is suitable for the propagation problem is greatly different from the background SSP, it is expected that more EOFs are required. Examples of the differences between I_L^{Λ} and I_P^A are shown in Figs. 5(a) and 5(b); the SSP measured at WP19 on JD239 19:17 GMT is used as an example SSP to be recovered. For the small sample set, it requires two modes to get over 95% of the total energy fit; while it needs three to reconstruct $s_k(N_Z)$ to meet 95% level of the misfit energy. For the large sample set, three modes represent over 95% of the total energy, but at least eight should be used to reconstruct $s_k(N_Z)$ to meet the requirement of 95% of the misfit energy. Clearly, the method of choosing the number of EOF coefficients to be inverted according to I_P^A rather than I_L^{Λ} is more conservative.

The number of EOFs to be inverted in the inversion is determined by checking the I_P^A versus the number of EOFs curve of all the SSPs in $\mathbf{S}(N_Z,N_S)$. For the large sample set, the greatest number of EOFs needed to reach over 90% of the misfit energy, which is 8, is chosen to be inverted. For the small sample set, the number of EOFs is 4 to reach over 95% of the misfit energy. The search bounds for each EOF coefficient are found by examining the values in the coefficient (weighting) matrix **A**. The following steps are taken to make sure sufficient large bounds are used in the inversion algorithm: (1) find the maximum absolute value in each row of **A** (corresponding to one mode), (2) expand this value to about to 5%, (3) reflect this value to the other direction with respect to 0, and (4) collect the values in steps (2) and (3) to form the lower and higher search bounds of that coefficient. For example, if the values of the first mode's coefficient $a_{1,k} \in [-51.5, 25]$, where $k=1,\ldots,Ns$, then the final search bounds for the coefficient are set to $[-55, 55]$.

D. Parameters to be inverted

There are three groups of parameters to be inverted:

- four geometric parameters: water depth, source depth, range, and array tilt;
- nine geoacoustic parameters: sediment depth, sediment p-wave sound speed and density at top and bottom, the constant and the exponent of frequency dependence of the attenuation of the sediment, and p-wave sound speed and density of the lower half space; and
- 4/8 EOF coefficients to describe the water column SSP, where 4 is the number of EOF coefficient to be inverted for the small SSP sample set and 8 is the number for the large SSP sample set.

The total number of parameters to be inverted for the small SSP sample set is 17, and for the large SSP sample set is 21. The p-wave attenuation of the lower half space was fixed to 0.3 dB/m at 1 kHz according to the canonical model derived from the previous studies on the New Jersey Shelf.[24,26,27,29–33]

The search bounds for the water depth and the source depth are determined according to the measurements shown in Fig. 2. The search bounds for the array tilt were extended to [−6°, 6°] since there is no indication of the array tilt direction, although the VLA tilts measured by the tilt-meter on the upper potion of the VLA during the signal transmissions were within [0°, 5°]. The search bounds for source range were extended ±500 m from the nominal value to accommodate the effect of ocean SSP on source localization. Geoacoustic parameters were given adequate search bounds to account for the possible effect of their variability over the range. The search bounds for all of the parameters to be inverted are summarized in Table II.

E. Data error covariance matrix estimation from multiple data segments

The procedure of estimating the data error covariance matrices is similar to the one introduced in Ref. 24. Multitonal optimization inversions were carried out on multiple 2.62-s-long segments of the data at each of the different ranges. The energy function of the optimization was

TABLE II. The search bounds of the model parameters to be inverted. In the geometric parameters column, the three pairs of search bounds for the range are for the inversions of 1 km, 3 km, and 5 km data, respectively. In the EOF column, the top four are for the small SSP sample set and the lower eight are for the large SSP sample set.

Geoacoustic parameters	
H (m)	[10 30]
c_{p1} (m/s)	[1560 1800]
c_{p2} (m/s)	[1560 1800]
c_{pb} (m/s)	[1650 2000]
α	[0 1.0]
β	[1.0 2.1]
ρ_1 (g/cm^3)	[1.6 2.2]
ρ_2 (g/cm^3)	[1.6 2.5]
ρ_b (g/cm^3)	[1.6 3.0]
Geometric parameters	
WD (m)	[78 83]
SD (m)	[28 32]
Range (km)	[0.95 1.05]
	[2.95 3.05]
	[4.95 5.05]
Array tilt (deg)	[–6.0 6.0]
EOF coefficients	
EOF 1	[−30 30]
EOF 2	[−20 0]
EOF 3	[−10 10]
EOF 4	[−10 10]
EOF 1	[−55 55]
EOF 2	[−25 25]
EOF 3	[−15 15]
EOF 4	[−15 15]
EOF 5	[−10 10]
EOF 6	[−15 15]
EOF 7	[−15 15]
EOF 8	[−10 10]

$$E(\mathbf{m}) = N_H \sum_{i=1}^{N_f} \log_e |\mathbf{d}_f - \mathbf{d}_f(\mathbf{m})|^2. \tag{17}$$

The optimization algorithm was adaptive simplex differential evolution, which combines the local optimization downhill simplex algorithm[34] and the global optimization algorithm differential evolution.[35] The population size[35] was ten times the number of the parameters to be inverted, mutation factor was 0.8, crossover factor was 0.8, and the perturbation number in the downhill process was 5.

The residuals on the hydrophones at eight frequencies were then collected to compute the covariance matrices at each range according to Eq. (14). The number of segments used to estimate data error covariance matrices $\mathbf{C}_f$ for WP21 was 60, for WP22 was 49, and for WP23 was 36. Figure 6 shows the estimated data error covariance matrices of the eight frequency components used later in Bayesian MFI of WP21 data. It is clear that the matrices are not identical independent distributed, and the data errors between the hydrophones have different correlations from frequency to frequency.

Besides constructing the data error covariance matrices, the results of the optimization inversions of the multiple data

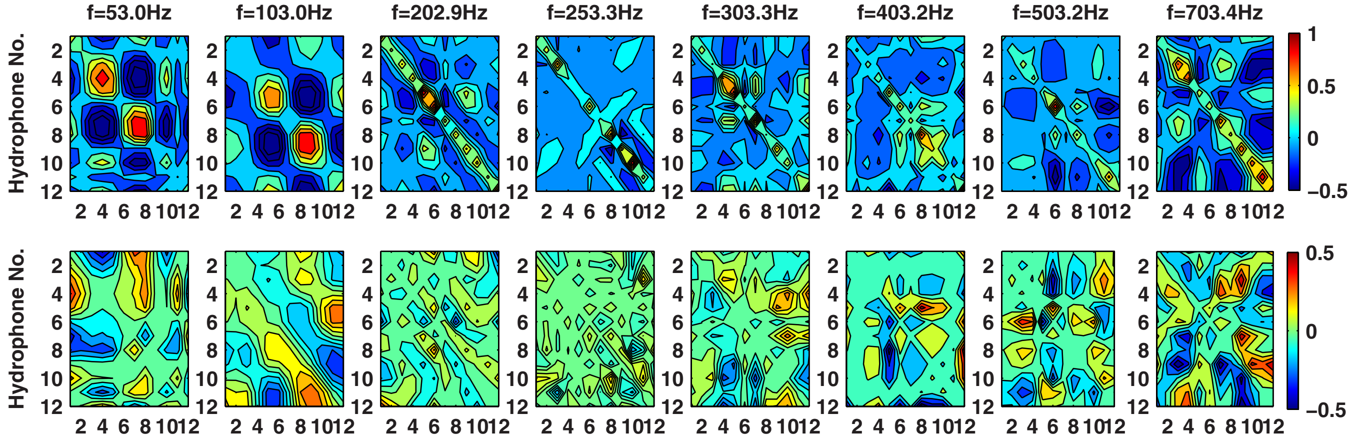


FIG. 6. (Color online) Estimated data error covariance matrices for WP21 data. The upper row displays the real parts of the covariance matrices at different frequencies and the lower row displays the imaginary parts.

samples (or segments) are used to examine the consistency of the geoacoustic parameter estimates. Since the seabed parameters such as sediment sound speed and layer thickness will not dramatically change during the time of the experiment, it should be expected that the dominant geoacoustic parameter estimates have small variances, even though the ocean environment and the experimental geometry may vary. The consistency of the estimates is checked quantitatively in terms of the means and variances of the optimization results, and qualitatively by examining the histogram[24,36] of the optimal parameter estimates. The results indicate that the sediment sound speed and layer thickness estimates from the optimization of multiple data segments are consistent at each range. The sediment sound speed and layer thickness estimates from the optimization inversion of the three ranges' data are also consistent.

F. Statistical validation of the assumptions made on the data errors in the inversion

The assumptions made in the Bayesian inversion approach were tested by checking the weighted residuals (the differences of the measured and modeled data, weighted by data error covariance). The assumption that the data errors are uncorrelated from frequency to frequency was tested qualitatively by checking the cross correlation of the weighted residuals between the frequencies, on both real and imaginary parts, respectively. There is no strong evidence against the assumption that data errors are uncorrelated from frequency to frequency. The assumption of the stationarity of data errors on each hydrophone over the observation time was checked by using Kolmogorov–Smirnov two-distribution test.[34,24] There was no strong evidence against the assumption of stationarity of the weighted residuals at a significance level α of 0.05 at each hydrophone and frequency. This is a sufficient condition that the ensemble average could be replaced with arithmetic average in estimating data error covariance matrices.

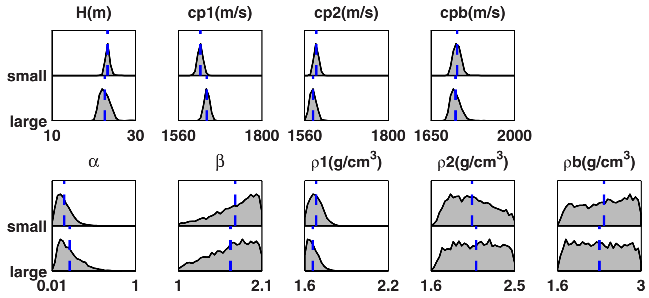


FIG. 7. (Color online) Comparison of the geoacoustic parameter estimates obtained by using small and large ocean SSP observation samples. The dashed lines indicate the mean values of the estimates.

As a necessary condition, the assumption that the weighted residuals are Gaussian distributed and spatially uncorrelated across the array was examined by using Kolmogorov–Smirnov test for normality and runs test for the randomness.[22] Both significant levels were 0.05. The tests results showed no strong evidence that the assumptions that the data errors were Gaussian distributed and spatially uncorrelated were violated.

V. BAYESIAN GEOACOUSTIC INVERSION RESULTS

A. Comparison of Bayesian MFI results between large and small SSP sample sets

To examine the inversion results obtained for the two different sets of SSP observations, one dimensional (1D) marginal distributions of geoacoustic parameter estimates obtained for the range of 1 km are shown in Fig. 7, and the geometric parameter estimates are shown in Fig. 8. The maximum values of the marginal distributions in all of the panels in both figures are set to the same value for the convenience of comparing the credibility intervals visually. The corresponding 95% credibility intervals and the MAP estimate of the parameters are listed in Table III. In terms of average values in the sediment layer, sediment sound speed

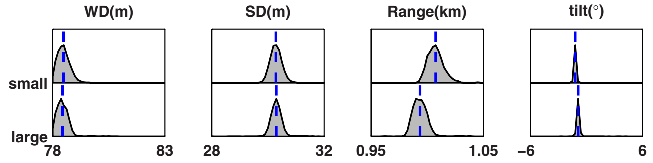


FIG. 8. (Color online) Comparison of the geometric parameter estimates obtained by using small and large ocean SSP observation samples. The dashed lines indicate the mean values of the estimates.

TABLE III. Comparison of 95% credibility intervals and MAP estimates of the parameters from Bayesian inversion by using different SSP sample information. The estimates are listed as (left bound MAP right bound).

	SSP sample set	
Parameters	Small	Large
H (m)	[22.3 23.0 24.3]	[20.6 22.0 25.0]
c_{p1} (m/s)	[1609.1 1626.0 1637.0]	[1626.0 1641.8 1656.6]
c_{p2} (m/s)	[1581.0 1590.9 1605.8]	[1567.2 1580.1 1601.7]
c_{pb} (m/s)	[1730.1 1758.6 1786.4]	[1717.6 1735.2 1795.2]
α	[0.016 0.154 0.321]	[0.028 0.220 0.481]
β	[1.230 1.383 2.099]	[1.164 1.712 2.099]
ρ_1 (g/cm^3)	[1.60 1.68 1.77]	[1.60 1.62 1.74]
ρ_2 (g/cm^3)	[1.68 1.82 2.43]	[1.70 1.94 2.48]
ρ_b (g/cm^3)	[1.75 2.34 2.99]	[1.68 2.10 2.95]
WD (m)	[78.0 78.5 79.0]	[78.0 78.2 78.4]
R (km)	[0.994 1.005 1.022]	[0.981 0.989 0.997]
SD (m)	[29.8 30.3 30.7]	[30.0 30.1 30.6]
Tilt (deg)	[−1.37 −1.17 −1.06]	[−1.14 −0.93 −0.81]

estimates are quite consistent in both approaches, with the average MAP value at 1608.3 m/s for the small sample set and 1610.4 m/s for the large sample set. In both approaches, sediment depth and half space sound speed are very well resolved. The p-wave attenuation constant and density at the top of the sediment are less sensitive. There is weak sensitivity of the attenuation in the sediment, but no information about the density at the bottom of the sediment and the half space is extracted in both approaches.

The geometric parameters are generally very sensitive. Estimates of water depth, source depth, and array tilt from both approaches are consistent with each other. Range estimates are not as consistent. The range estimated from the small SSP sample set approach is closer to the value of 1.007 km, which is the average range derived from global positioning system measurements. The very narrow 95% credibility interval of the array tilt estimation indicates the VLA was leaning in a specific direction in the experiment, which would be expected from the wind conditions during the data collection.

EOF coefficients estimates are plotted in Fig. 9 to show the sensitivity of the coefficients at 1 km range. The 1D marginal distributions in the figure are normalized in a way that the integral of each individual marginal PPD equals 1. The creditability intervals can be used as an indicator of the weight of a specific EOF. Figures 10(a) and 10(b) display the SSP marginal distributions derived from the EOF coefficient estimates shown in Figs. 9(a) and 9(b). It is seen from the rather broad minimum-maximum boundary of the two plots that the inversions were given large search bounds in searching for the water column SSP. The very tight 95% credibility bounds indicate that the water column SSP is very sensitive in both of the inversions. Figure 10(c) displays the MAP SSP estimates from the two approaches on top of the SSPs measured at the source ship stations. It is evident from Fig. 10(c) that the effective SSP for the 1 km data is very close to CTD-WP19-2 (CTD measured 2 h earlier than the time of source transmission at WP19). The SSP estimated from the small sample set approach follows CTD-WP19-2 closely. It is obvious that the top 12 m of the SSP estimated from the large sample set approach shifts away from the measured SSPs, which are almost identical. However, the lower part of the SSP (from 25 m and down to the sea bottom) is consistent with the one estimated from the small SSP sample set approach. Given the fact that the inverted geoacoustic and geometric parameters from the large SSP sample set approach are consistent with the ones from the small SSP sample set approach, it suggests that the SSP gradient around the source depth is important.

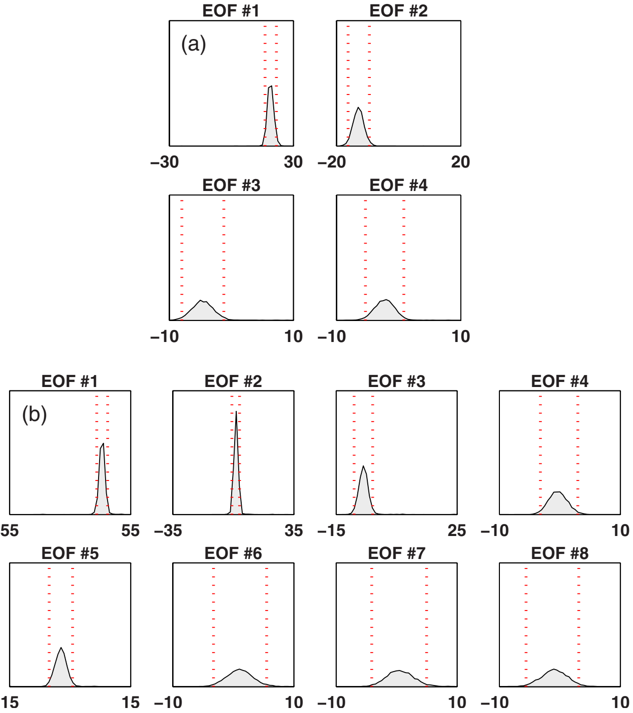


FIG. 9. (Color online) Comparison of the EOF coefficient estimates obtained by using different ocean SSP observation samples. (a) The small SSP sample set and (b) the large SSP sample set. The dotted lines indicate the bounds of the 95% credibility intervals.

The sensitivity of the EOF coefficients shown in Fig. 9(b) confirms that the number of EOF coefficients to be inverted is appropriate. The inversion demonstrates that the number of EOF coefficients to be inverted should be determined by the number of modes needed to recover an individual SSP to a certain degree of fitness, rather than by simply using the number of the most dominant modes (the modes that have the largest eigenvalues). Because the small SSP sample set has the most relevant information about the SSP between the source and the receiver, the inversion works very well even though the number of SSPs used in the analysis is not large statistically. Moreover, since the upper and lower portions are constrained to the average measured profile values, this approach requires fewer EOFs to construct the effective SSP in the thermocline. More EOF coefficients are necessary in the inversion when using the large SSP sample set because we are fitting the entire profile, and the effective SSP for the 1 km data is considerably different from the average SSP of the sample set. Since more EOFs are required, the downside of using the large SSP sample set is the increased computational effort in the inversion to estimate the geoacoustic parameters.

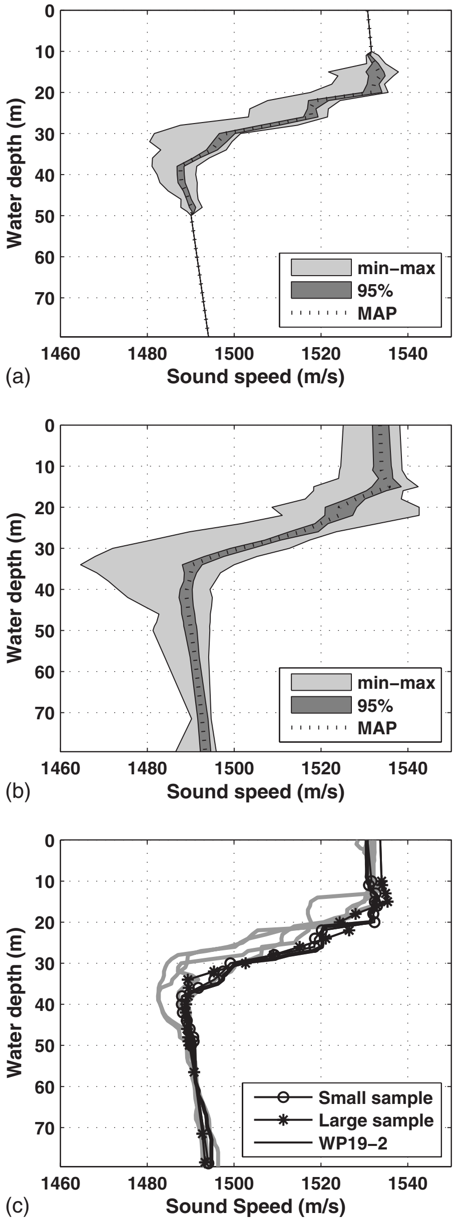


FIG. 10. Comparison of the inverted SSP estimates obtained by using different ocean SSP statistics: (a) marginal distribution of SSP using the small SSP sample set, (b) marginal distribution of SSP using the large SSP sample set, and (c) MAP estimates of SSPs from the small and the large SSP sample sets. The lighter curves are the SSPs measured at different source stations.

B. The sensitivity of the geoacoustic parameter versus range

In order to observe the evolution of the parameter sensitivity over the range, Bayesian MFI is applied on the multitonal data collected at 1, 3, and 5 km, using the small SSP sample set for water column SSP parametrization. A comparison of 1D marginal distributions of the geoacoustic parameter estimates at three ranges is shown in Fig. 11. The corresponding 95% credibility intervals and MAP estimates are listed in Table IV.

Compared to the inversion results of 1 and 5 km data, only sediment *p*-wave sound speeds are resolvable at 3 km. Although there is some sensitivity of layer thickness, attenuation constant, and half space sound speed and density, the relatively poor estimation of water depth (which was pushed to the lower search bound) makes the inversion results of 3 km data not very convincing. A Bayesian MFI of 3 km data using the large SSP sample set gives similar results. It is likely that the ocean environment variability is too severe to be represented by a single effective range independent SSP for this case.

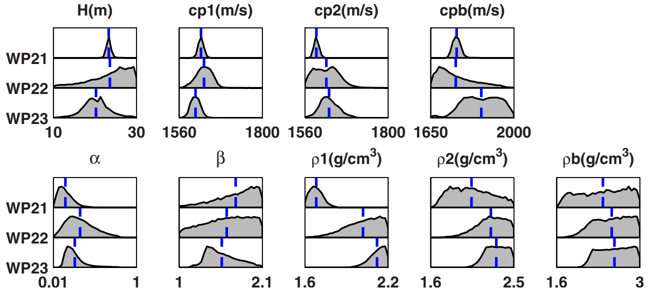


FIG. 11. (Color online) Comparison of geoacoustic estimates from different WPs obtained by using the small SSP sample set. The dashed lines indicate the mean values of the estimates.

The inversion results at 1 and 5 km are consistent with each other. The average sediment sound speeds are 1608.3 and 1610.0 m/s, respectively. Negative gradient of the sediment sound speed can also be observed if MAP values are considered. The sediment layer thickness estimates are 23.3 ± 0.5 and 20.3 ± 3.2 m, respectively. The sensitivities of the sediment layer thickness, *p*-wave sound speeds in the sediment, and half space decrease when the range increases. The sensitivity of *p*-wave attenuation constant stays almost the same but the sensitivity of the exponent of the frequency dependence of the attenuation increases over the range. It is expected that the effect of the frequency dependence of the attenuation would be more evident at longer ranges. It should be noticed that the attenuation discussed here is a general loss factor that integrates other mechanisms of energy loss process caused by seafloor roughness, sediment spatial inhomogeneities (variability in range and in depth), and absorption in the water and sediment, etc.

C. The inter-parameter correlation of the ocean SSP parameters and geometric/geoacoustic parameters

The correlations between the water column SSP and geometric/geoacoustic parameters at the three ranges are displayed in Fig. 12 in terms of two dimensional (2D) marginal distributions and inter-parameter correlation matrices computed according to Eq. (7).

2D marginal distributions of the EOF coefficients versus selected geometric/geoacoustic parameters at the three ranges are shown in Figs. 12(a), 12(c), and 12(e). It is seen that the water column SSP has strong correlation with range and source depth at all ranges. The effect of the SSP on water depth estimation is greater at closer range. Dual modes of the PPDs of the EOF coefficients are found in Fig. 12(c). For these data, it is likely that the assumption of a single effective water column SSP is not adequate. It is also seen that the ocean SSP has a greater effect on water depth and range estimates at 3 km range. The impact of the oceanic SSP on the source range becomes more severe at 5 km range, which is shown in Fig. 12(e). Strong correlations between the EOF

TABLE IV. Summary of MAP estimates and 95% credibility interval of geoacoustic model estimates at different ranges listed as (left bound MAP right bound).

	Range		
Parameters	1 km	3 km	5 km
H (m)	[22.3 23.0 24.3]	[13.5 26.2 30.0]	[13.7 18.3 27.4]
c_{p1} (m/s)	[1609.1 1626.0 1637.0]	[1585.4 1642.5 1670.7]	[1581.0 1619.8 1632.8]
c_{p2} (m/s)	[1581.0 1590.9 1605.8]	[1560.0 1585.4 1679.6]	[1578.5 1600.4 1686.3]
c_{pb} (m/s)	[1730.1 1758.6 1786.4]	[1650.0 1848.8 1910.1]	[1744.1 1894.9 1997.6]
α	[0.016 0.154 0.321]	[0.029 0.049 0.654]	[0.104 0.136 0.541]
β	[1.230 1.383 2.099]	[1.013 1.106 2.099]	[1.252 1.342 1.999]
ρ_1 (g/cm^3)	[1.60 1.68 1.77]	[1.79 1.88 2.20]	[2.01 2.17 2.20]
ρ_2 (g/cm^3)	[1.68 1.82 2.43]	[1.97 2.46 2.50]	[2.12 2.18 2.50]
ρ_b (g/cm^3)	[1.75 2.34 2.99]	[2.03 2.96 3.00]	[2.17 2.82 3.00]
WD (m)	[78.0 78.5 79.0]	[78.0 78.3 78.8]	[78.4 78.7 79.1]
R (km)	[0.994 1.005 1.022]	[2.951 2.996 3.033]	[4.964 4.991 5.050]
SD (m)	[29.8 30.3 30.7]	[29.6 30.2 31.4]	[28.6 29.2 30.2]
Tilt (deg)	[−1.37 −1.17 −1.06]	[−1.77 −0.92 −0.62]	[−1.64 −0.86 −0.69]

coefficients and the range imply larger uncertainty in source range estimation in a dynamic ocean environment at longer range.

The inter-parameter correlations of all of the inverted parameters are shown in Figs. 12(b), 12(d), and 12(f). Strong positive correlations of water depth and range, and water depth and source depth are found at all three ranges, which is consistent with the findings in previous studies.[24,37–39] Negative correlation between the sound speed at the sediment top and bottom suggests that the inversion tries to adjust those two sound speeds to get an average sound speed in the layer. It also indicates a lower sound speed layer between the seafloor and the R reflector. Notable correlation between the EOF coefficients and geometric parameters (water depth, range source depth, and array tilt) are found at all three ranges. However, it is difficult to interpret the reason for these correlations because it is not intuitively clear what contribution each individual EOF coefficient makes to the ocean SSP characterization. One conclusion can be drawn that the ocean SSP has greater impact on geometric parameter estimates. Positive correlation between the constant and the exponent of the frequency dependence of the attenuation may suggest that there exist more than one frequency component that is sensitive to those two factors. The misfits of those frequency components dominate the misfit function used in the inversion. As a result, the inversion needs to find a proper factor to scale the attenuation to fit the data at different frequencies.

VI. SUMMARY

This paper applies Bayesian matched field inversion technique to multi-tonal data collected on the New Jersey continental shelf where strong water column sound speed variability caused by internal waves, local eddies, etc., was evident. The ocean SSP was parametrized in terms of EOFs and inverted along with the geoacoustic parameters to account for the dynamic water column environment in the inversion. Inverting ocean SSP simultaneously improved the estimates of the known geometric parameters, and provides qualitative confidence that the propagation was correctly modeled.[40] The inversions show that the most sensitive geoacoustic parameters are the sediment layer thickness, and the sediment sound speeds at the top and bottom of the layer. Notable negative gradient of the sediment sound speed is evident from the inversions at 1, 3, and 5 km.

Two sets of ocean SSP observations were employed in the EOF analysis to investigate the effects of different SSP statistics on the inversion results. The inversions of the 1 km data from the two approaches show that the impact of the dynamic ocean environment is mitigated, since the geometric parameter estimates such as water depth, range, and source depth are consistent with the known values, and the geoacoustic parameter estimates from the two approaches are also in excellent agreement. We can conclude that, for relatively weak variability in the ocean environment, using a limited set of SSPs that contain information about the variability during the experiment is an effective approach. Moreover, this approach may permit further simplifications such as using a portion of the observed SSP as was done effectively for the 1 km data. However, this approach is limited by more severe of the environmental variation along the propagation path that could cause mode coupling.

Inversions of multi-tonal data collected at different ranges reveal that the sensitivities of the geoacoustic parameters change over the range, as expected. The ability of resolving the layer information of the sediment decreases because the acoustic field becomes evanescent in the sediment at longer range. The sensitivity of the frequency dependency of p-wave attenuation increases over the range since the effects of different energy loss mechanisms on different frequencies build up over longer ranges. The 2D marginal distributions and the inter-parameter correlations derived from the PPD also indicate that the water column SSP has great impact on geometric/geoacoustic parameter estimation. Using an inappropriate SSP leads to incorrect water depth, range, and source depth estimation.

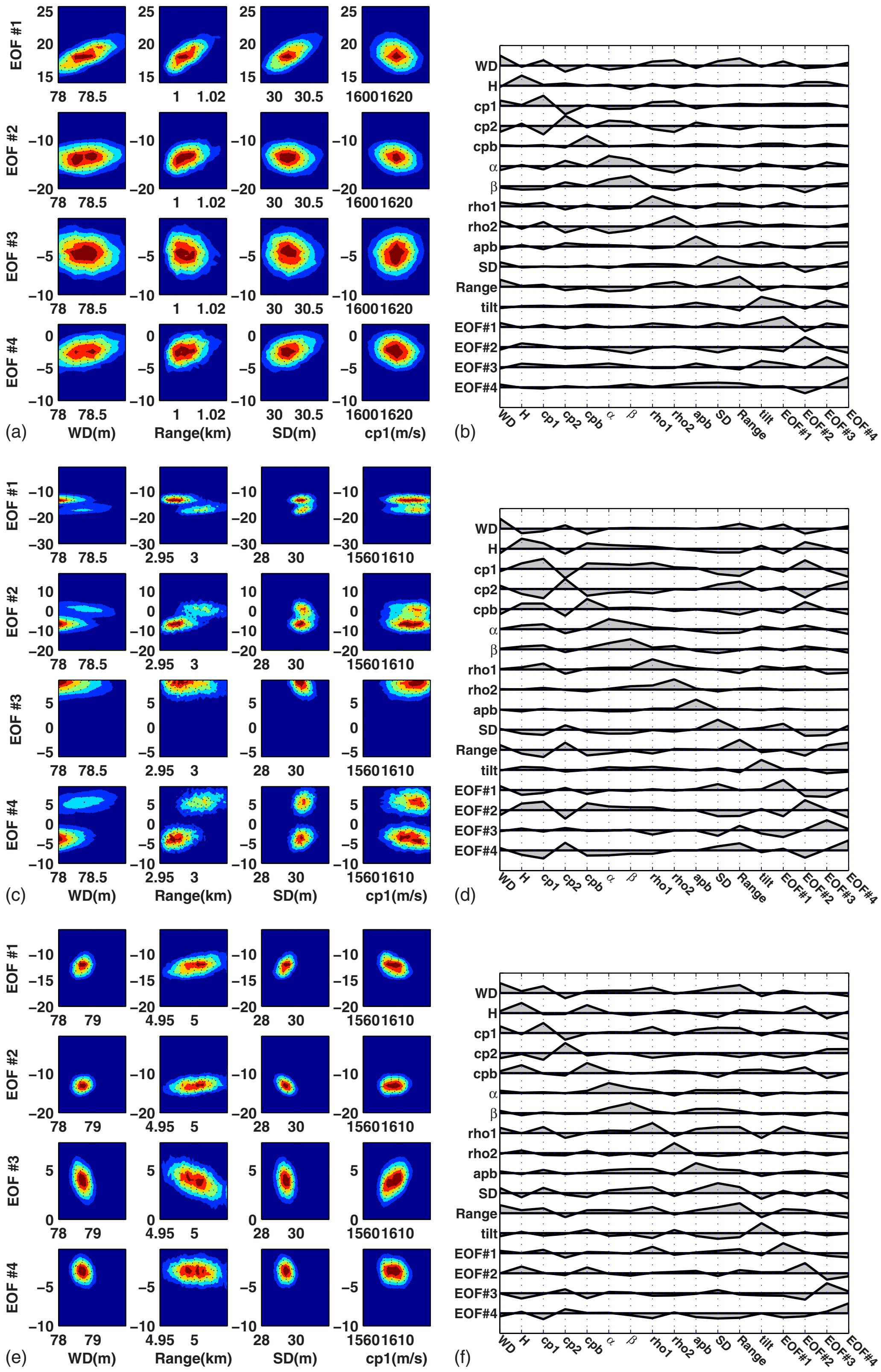


FIG. 12. (Color online) 2D marginal distributions of EOFs and water depth, range, source depth, and the sound speed at the sediment top at (a) WP21, (c) WP22, and (e) WP23, and inter-parameter correlations of the inverted parameters at different ranges: (b) WP21, (d) WP22, and (f) WP23.

ACKNOWLEDGMENTS

This work is supported by the Office of Naval Research Ocean Acoustic Team under Grant No. N00014-05-1-0264. The authors would like to thank Dr. William S. Hodgkiss and Dr. Peter Gerstoft, for the acoustic data, Dr. John Goff for the chirp seismic reflection data and high resolution bathymetry data of the experimental site, Dr. David Knobles for the navigation data and source depth measurements data during the source transmitting, and Mr. Arthur Newhall for the environmental mooring data.

APPENDIX: WATER COLUMN PARAMETRIZATION IN TERMS OF EOF

Let a $N_Z \times N_S$ matrix $\mathbf{S}(N_Z, N_S)$ be the ocean SSP observations collected at different locations, $Z = [Z_1, Z_2, \ldots, Z_{Nz}]^T$ be the vector of the depths of the SSP samples, where N_z is the number of SSP data points alone the depth, and N_s is the population of the SSP observed at different locations. $\mathbf{S}(N_Z, N_S)$ can be expressed in terms of the sum of the mean SSP (or background SSP), $\bar{s}(N_Z) = E[\mathbf{S}(N_Z, N_S)] = [\bar{s}(Z_1), \bar{s}(Z_2), \ldots, \bar{s}(Z_{N_Z})]^T$, and the residual SSPs $\mathbf{S_r}(N_Z, N_S)$

$$\mathbf{S}(N_Z, N_S) = \bar{s}(N_Z) + \mathbf{S_r}(N_Z, N_S), \tag{A1}$$

where $E[\cdot]$ is the mathematic expectation and $\mathbf{S_r}(N_Z, N_S)$ are the differences between $\mathbf{S}(N_Z, N_S)$ and $\bar{s}(N_Z)$. Apparently, $\mathbf{S_r}(N_Z, N_S)$ is also a $N_Z \times N_S$ matrix and can be written as a linear combination of a set of orthogonal functions

$$\mathbf{S_r}(N_Z, N_S) = \mathbf{HA}, \tag{A2}$$

where $\mathbf{H}$ is a $N_Z \times M$ matrix that contains M orthogonal functions and $\mathbf{A}$ is a $M \times N_S$ matrix, which represents the weights corresponding to the orthogonal functions to reconstruct the individual residual SSP. The weighting matrix $\mathbf{A}$ is found by projecting $\mathbf{H}^T$ on the SSP residual matrix $\mathbf{S_r}$, once the orthogonal function set is found,

$$\mathbf{A} = \mathbf{H}^T \mathbf{S_r}(N_Z, N_S). \tag{A3}$$

The statistical independent orthonormal basis set $\mathbf{V}_{\mathrm{EOF}}$ can be found by finding the eigenvectors of the SSP residual covariance matrix[41] $\mathbf{R}$ as follows:

$$\mathbf{R} = E[\mathbf{S_r}\mathbf{S_r}^T] = \mathbf{V}_{\mathrm{EOF}} \mathbf{\Lambda}_{\mathrm{EOF}} \mathbf{V}_{\mathrm{EOF}}^T. \tag{A4}$$

$\mathbf{V}_{\mathrm{EOF}}$ is a $N_Z \times N_J$ matrix that contains N_J significant eigenvectors, where $N_J \leq \min(N_Z, N_S)$. Each eigenvector represents one mode of the SSP variations in depth. $\mathbf{\Lambda}_{\mathrm{EOF}}$ is a $N_J \times N_J$ diagonal matrix, where the elements on the diagonal are the eigenvalues. Each eigenvalue represents the energy of individual mode.

$\mathbf{H}$ can be related to the orthonormal EOF basis by a unitary transform: $\mathbf{H} = \mathbf{V}_{\mathrm{EOF}}\mathbf{E}$, where $\mathbf{E}\mathbf{E}^T = \mathbf{E}^T\mathbf{E} = \mathbf{I}$, and $\mathbf{I}$ is the identity matrix. Consequently, Eq. (A1) can be written as

$$\mathbf{S}(N_Z, N_S) = \bar{s}(N_Z) + \mathbf{HA} = \bar{s}(N_Z) + \mathbf{V}_{\mathrm{EOF}} \cdot (\mathbf{V}_{\mathrm{EOF}}^T \cdot \mathbf{S_r}(N_Z, N_S)). \tag{A5}$$

Or an equivalent format for an individual profile $s_k(N_Z)$

$$s_k(N_Z) = \bar{s}(N_Z) + \sum_{m=1}^{N_J} a_{m,k} v_m(N_Z), \tag{A6}$$

where the weighting factor $a_{m,k}$ (an element of weighting matrix $\mathbf{A}$) is found by projecting $\mathbf{V}_{\mathrm{EOF}}$ on the corresponding SSP residual, v_m is the mth eigenvector, and $k \in [1, \ldots, N_S]$.

[1] J. C. Preisig and T. F. Duda, “Coupled acoustic mode propagation through continental-shelf internal solitary waves,” IEEE J. Ocean. Eng. **22**, 256–269 (1997).

[2] K. M. Becker and G. V. Frisk, “The impact of water column variability on horizontal wave number estimation and mode based geoacoustic inversion results,” J. Acoust. Soc. Am. **123**, 658–666 (2008).

[3] Y.-T. Lin, C.-F. Chen, and J. F. Lynch, “An equivalent transform method for evaluating the effect of water column mismatch on geoacoustic inversion,” IEEE J. Ocean. Eng. **31**, 284–298 (2006).

[4] M. Siderius, P. L. Nielsen, J. Sellschopp, M. Snellen, and D. Simons, “Experimental study of geo-acoustic inversion uncertainty due to ocean sound-speed fluctuations,” J. Acoust. Soc. Am. **110**, 769–781 (2001).

[5] M. Snellen, D. G. Simons, M. Siderius, J. Sellschopp, and P. L. Nielsen, “An evaluation of the accuracy of shallow water matched field inversion results,” J. Acoust. Soc. Am. **109**, 514–527 (2001).

[6] 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).

[7] C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, “Effect of ocean sound speed uncertainty on matched-field geoacoustic inversion,” J. Acoust. Soc. Am. **123**, EL162–EL168 (2008).

[8] Y.-M. Jiang and N. R. Chapman, “Bayesian geoacoustic inversion in a dynamic shallow water environment,” J. Acoust. Soc. Am. **123**, EL155–EL161 (2008).

[9] A. Tolstoy, N. R. Chapman, and G. E. Brooke, “Workshop’ 97: Benchmarking for geoacoustic inversion in shallow water,” J. Comput. Acoust. **6**, 1–28 (1998).

[10] N. R. Chapman, S. A. Chin-Bing, D. King, and R. Evans, “Benchmarking geoacoustic inversion methods for range-dependent waveguides,” IEEE J. Ocean. Eng. **28**, 320–330 (2003).

[11] D. J. Tang, J. Moum, J. Lynch, P. Abbot, R. Chapman, P. Dahl, T. Duda, G. Gawarkiewicz, S. Glenn, J. Goff, H. Graber, J. Kemp, A. Maffei, J. Nash, and A. Newhall, “Shallow Water ’06—A joint acoustic propagation/nonlinear internal wave physics experiment,” Oceanogr. **20**, 156–167 (2007).

[12] Navigation and source depth monitoring data were provided by Dr. David Knobles.

[13] SSPs for SW54 mooring and CTDs for SW31 and SW32 moorings were provided by Mr. Arthur Newhall.

[14] G. S. K. Wong and S. Zhu, “Speed of sound in seawater as a function of salinity temperature and pressure,” J. Acoust. Soc. Am. **97**, 1732–1736 (1995).

[15] C.-T. Chen and F. J. Millero, “Speed of sound in seawater at high pressures,” J. Acoust. Soc. Am. **62**, 1129–1135 (1977).

[16] C. C. Leroy and F. Parthiot, “Depth-pressure relationship in the oceans and seas,” J. Acoust. Soc. Am. **103**, 1346–1352 (1998).

[17] N. P. Fofonoff and R. C. Millard, Jr., “Algorithms for computation of fundamental properties of seawater,” UNESCO Technical Papers in Marine Science, No. 44, Division of Marine Sciences, UNESCO, Place de Fontenoy, Paris (1983).

[18] R. G. Perkin and E. L. Lewis, “The practical salinity scale 1978: Fitting the data,” IEEE J. Ocean. Eng. **5**, 9–16 (1980).

[19] M. K. Sen and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion* (Elsevier, Amsterdam, 1995).

[20] P. Gerstoft and C. F. Mecklenbräuker, “Ocean acoustic inversion with estimation of *a posteriori* probability distributions,” J. Acoust. Soc. Am. **104**, 808–819 (1998).

[21] S. E. Dosso, “Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach,” J. Acoust. Soc. Am. **111**, 129–142. (2002).

[22] S. E. Dosso, P. Nielsen, and M. J. Wilmut, “Data error covariance in matched-field geoacoustic inversion,” J. Acoust. Soc. Am. **119**, 208–219 (2006).

[23] C. F. Mecklenbräuker and P. Gerstoft, “Objective functions for ocean acoustic inversion derived by likelihood methods,” J. Comput. Acoust. **8**, 259–270 (2000).

[24] Y.-M. Jiang, N. R. Chapman, and M. Badiey, “Quantifying the uncertainty of geoacoustic model parameters for the New Jersey shelf by inverting air gun data,” J. Acoust. Soc. Am. **121**, 1879–1895 (2007).

[25] E. K. Westwood, C. T. Tindle, and N. R. Chapman, “A normal mode model for acousto-elastic ocean environments,” J. Acoust. Soc. Am. **100**, 3631–3645 (1996).

[26] A. Turgut, D. Lavoie, D. J. Walter, and W. B. Sawyer, “Measurements of bottom variability during SWAT New Jersey Shelf experiments,” *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer, Dordrecht, 2000), pp. 91–98.

[27] J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, “Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties,” Mar. Geol. **209**, pp. 147–172 (2004).

[28] High resolution geophysical survey data were provided by Dr. John Goff.

[29] B. J. Kraft, L. A. Mayer, P. Simpkin, P. Lavoie, E. Jabs, and J. A. Goff, “Calculation of in situ acoustic wave properties in marine sediments,”

Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance (Kluwer, Dordrecht, 2002), pp. 123–130.

[30] L. A. Mayer, B. J. Kraft, P. Simpkin, P. Lavoie, E. Jabs, and E. Lynskey, "In-situ determination of the variability of seafloor acoustic properties: An example from the ONR geoclutter area," *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer, Dordrecht, 2002), pp. 115–122.

[31] M. V. Trevorrow and T. Yamamoto, "Summary of marine sedimentary shear modules and acoustic speed profile results using a gravity wave inversion technique," J. Acoust. Soc. Am. **90**, 441–455 (1991).

[32] W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey Continental Shelf," IEEE J. Ocean. Eng. **20**, 321–336 (1995).

[33] M. D. Richardson and K. B. Briggs, "Relationships among sediment physical and acoustic properties in siliciclastic and calcareous sediment," ECUA2004, Delft (2004), Vol. **2**, pp. 659–664.

[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[35] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," J. Global Optim. **11**, 341–359 (1997).

[36] N. R. Chapman, Y.-M. Jiang, B. Hodgkiss, and P. Gerstoft, "Geoacoustic inversion in the SW06 shallow water experiments," Proceedings of Underwater Measurements Technology, Crete, Greece (2007).

[37] 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).

[38] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).

[39] C. H. Harrison and M. Siderius, "Effective parameters for matched field geoacoustic inversion in range-dependent environment," IEEE J. Ocean. Eng. **28**, 432–445 (2003).

[40] N. R. Chapman and Y.-M. Jiang, "Inference of ocean bottom properties by geoacoustic inversion," *Important Elements in: Geoacoustic Inversion, Signal Processing, and Reverberation in Underwater Acoustics* (Research Signpost, Kerala, India, 2008), pp. 55–78.

[41] L. R. LeBlanc and F. H. Middleton, "An underwater acoustic sound velocity data model," J. Acoust. Soc. Am. **67**, 2055–2062 (1980).

Passive acoustic detection of schools of herring

Thomas R. Hahn and Gary Thomas
RSMAS, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 28 May 2007; revised 13 February 2009; accepted 13 February 2009)

Herring (*Clupea pallasii* and *C. harengus*) have been observed to release gas from their bladders during vertical migration likely to adjust buoyancy and also when under strong predation pressure. Based on recently measured and modeled sound for individual fish, spectral levels are estimated for entire herring schools in the ocean for both scenarios, and the feasibility of passive detection is explored. For a typical school of migrating herring near-surface spectral levels of about 50 dB rel., 1 μPa/$\sqrt{\text{Hz}}$ at 3–7 kHz are predicted. If wind conditions are calm where migrating herring are found, such as for Pacific herring in Prince William Sound, Alaska, passive detection is very likely. For an exemplary 10 metric ton compact school, peak spectral source levels of about 80–90 dB rel. 1 μPa/$\sqrt{\text{Hz}}$ ref. 1 m are predicted, yielding a range of detection against calm wind background of about 1000 m. Field measurements of potential gas-release events agree with the predictions for the compact school scenario with regard to levels and spectral shape and indicate that passive acoustic monitoring is feasible and could be a prime tool to study predator-prey interactions.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097473]

PACS number(s): 43.30.Sf, 43.30.Nb, 43.80.Ka [KFG] Pages: 2896–2908

I. INTRODUCTION

It has been repeatedly observed,[1,2] both optically and acoustically, that schools of spawning herring release gas from their swim bladders when under attack from predators or during diurnal vertical migration. This collective gas release can result in bubble clouds large enough to form slick areas and white foam above the herring school when the rising bubbles break the surface. The reason for the gas release is believed to be twofold. During a vertical upward movement, herring release gas from their bladders to compensate for the pressure change and the corresponding increase in bladder volume. The vertical motion is either triggered by diel migration of schools from depths of up to 400 m to the surface at night or forced by predators. It is also speculated that compact herring schools may release gas to acoustically and optically confuse attacking predators or for communication purposes.[3] This hypothesis is based on the fact that bubble clouds have a strong impact on the active and passive acoustic environment and are also effective scatterers of light. This could lead to a decreased acoustic and optical "visibility" of the herring and also to the presence of "false targets" in both detection modes. Both effects not only potentially fool predators but also fisheries acousticians that must discriminate bubbles from fish when estimating biomass.

Wahlberg and Westerberg[4] found source levels of about 73 dB re μPa rms ref. 1 m of the chirps originating from gas release of individual herring (*Clupea harengus*). Hahn and Thomas[5] presented a detailed model for the source levels of individual herring that is based on the observations of Wahlberg and Westerberg[4] and Wilson *et al.*,[3] which generally agrees with their data. Based on this model and on echograms of gas release of herring schools obtained by Thorne and Thomas[2] for Pacific herring (*C. pallasii*) and by Nottestad[1] for Norwegian herring (*C. harengus*), this paper explores the feasibility of passive acoustic detection of entire herring schools by observing the sound that their collective gas release generates.

Passive acoustics devices at low frequencies are simple, very cost efficient, and have the potential to cover large areas. Passive acoustic monitoring of herring can complement ongoing active acoustic survey techniques that are used for stock assessment purposes. One practically and economically limiting factor of these surveys is the need to locate the herring aggregations within a large area before detailed biomass measurements can take place. Passive acoustics will help overcome these limitations. Besides making these surveys more efficient, understanding the acoustic impact of the gas release will also improve our understanding of predator-prey interactions.

Passive acoustic detection and monitoring techniques have been proposed and employed for several vocal fish species in the past. It is not appropriate to comprehensively review the body of literature on passive acoustic applications in fisheries here. We rather point to an excellent review paper by Rountree *et al.*,[6] as well as to the various contributions to the proceedings[7] of *Listening to Fish: An International Workshop on the Applications of Passive Acoustics in Fisheries*, which both also provide a rich bibliography. Because of their commercial importance and their strong vocal signature, special attention has been given to drum[8–10] (*sciaenids*) and cod[11–13] (*gadids*) fishes but even species that are not widely associated with vocalization, such as tuna,[14] have been studied from this point of view. Moreover, it is estimated that far more than 800 fish species are vocal[6] and, hence, are a potential target for passive acoustic monitoring. Although passive acoustic techniques in fisheries are a relatively new development, they have been commonly applied to the monitoring of marine mammals[15–17] for a much longer time.

All these studies are based on a thorough understanding of the spectral and temporal characteristics of the sound produced by the various species. As the sounds of many species are quite distinct, an acoustic discrimination can often be accomplished without deeper insight into the physics of sound production. The opposite is the case for herring. Here, the received sounds are not due to direct vocalizations of the fish but rather an acoustic consequence of the bubble release. One could argue that ringing bubbles are the most common and generic underwater sound in the frequency domain of interest. They are, for example, responsible for the wind-driven component of ambient noise in the ocean[18–20] emitted by breaking waves—due to collective modes even at very low frequencies[21–25]—and for the underwater sound of rain.[26,27] Hence, we aim to describe the sound of herring gas-release events from first physical principles to learn how to distinguish it from other bubble related sound sources. We think that it is fair to say that compared to traditional active acoustic assessments, passive techniques have not nearly been developed and used according to their true potential. We hope that this study will help to close this gap.

An outline of our approach is given in the first part of Sec. II. The presentation continues with a review of the acoustics of individual herring that was developed in our previous paper.[5] The core of Sec. II contains a theoretical estimate of the expected sound levels and a comparison with field data. We address two scenarios of collective gas release: a compact school under strong predation pressure and a vertically migrating school releasing gas at a narrow depth layer. A discussion of the results, as well as an outlook toward further developments, is given in Sec. III.

II. ESTIMATION OF SOUND LEVELS

A. General approach

Let $p_b(\boldsymbol{r},\boldsymbol{r}_b;t)$ be the waveform of an acoustic pulse received at $\boldsymbol{r}$ at time $t=0$ due to a single source event at $\boldsymbol{r}_b$. The entire signal at $\boldsymbol{r}$ is the superposition of all such individual pulses received at randomly occurring times t_k within the observation time T:

$$p(\boldsymbol{r},t)=\sum_{k=1}^{N} p_{b_k}(\boldsymbol{r},\boldsymbol{r}_{b_k};t-t_k). \qquad (1)$$

The pulses in question are the acoustic emissions from oscillating bubbles released by all individual fish in the school. These pulses have been documented *in situ* and in the laboratory[3,4] and are well understood theoretically. Hahn and Thomas[5] laid out the framework to determine the acoustic levels due to the sum of all such pulses originating in an "active" volume V_b that, depending on the behavior of the school, can be assumed to extend either over the entire school or only over a part of it. Following this approach, the spectral density observed at $\boldsymbol{r}$ is given by

$$\langle S_\omega\rangle = 8\pi^2\overline{\langle s_b{}^2\rangle|\tilde{f}_\omega|^2}\int \rho_t|G_\omega(\boldsymbol{r},\boldsymbol{r}_b)|^2 dV_b. \qquad (2)$$

In this expression, ρ_t is the number of events per unit time and volume, G_ω is the Green's function of the acoustic environment that includes attenuation effects within the school, s_b is the source strength, and $\tilde{f}_\omega$ is the spectral shape of the pulse.

The spectral shape $\tilde{f}_\omega$, that is, the normalized intensity of the source as a function of frequency f of the pulse, $\omega=2\pi f$, is in this model determined by standard low amplitude bubble dynamics,

$$|\tilde{f}|^2=\frac{4\beta}{\pi(\omega_0{}^2+5\beta^2)}\,\frac{\beta^2(\beta^2+4\omega^2+2\omega_0{}^2)+\omega_0{}^4}{[\beta^2+(\omega-\omega_0)^2][\beta^2+(\omega+\omega_0)^2]}, \qquad (3)$$

in terms of the bubble damping constant β and the bubble resonance frequency ω_0. Both are a function of the bubble radius a. $\tilde{f}_\omega$ is chosen to be a generic function of the bubble size only. If needed, the spectral density can later be averaged over the size distribution of all active bubbles. This is indicated by the over-bar in Eq. (2).

For completeness, we give a brief review[28] of the computation of β and ω_0: To a good approximation, for small amplitude variations,

$$\omega_0=\frac{P_{b0}}{\rho a^2}\left[3\gamma_{\text{eff}}-\frac{2\sigma}{aP_{b0}}\right],$$

$$P_{b0}=P_0(z)+\frac{2\sigma}{a}, \qquad (4)$$

where P_0 is the hydrostatic pressure at the locus of the bubble, ρ is the density of water, $\sigma=0.073$ N m^{-1} is the air-water surface tension, and γ_{eff} is the effective polytropic index, which depends on the bubble size and frequency. For millimeter-sized bubbles in the low-kilohertz frequency regime, $\gamma_{\text{eff}}=1.2-1.4$. The damping constant β has components due to viscous, thermal,[29,30] and radiation damping,

$$\beta=\frac{2\mu}{\rho a^2}+\frac{P_{b0}}{2\rho a^2}\,\text{Im}\,\phi+\frac{\omega^2 a}{2c},$$

$$\phi=\frac{3\gamma}{(1-3(\gamma-1)i\chi[(i/\chi)^{1/2}\coth(i/\chi)^{1/2}-1])},$$

$$\chi=\frac{D}{\omega a^2}, \qquad (5)$$

where $\mu=0.001$ Pa s is the dynamic viscosity of water, $D=2.08\times10^{-5}$ m^2 s^{-1} is the thermal diffusivity, and $\gamma=1.4$ is the ratio of the specific heats of the bubble gas. (Our observations were made in winter. Since herring do not feed much during this time of the year, we can exclude other intestinal gases and assume that the gas in the released bubbles is air.) The quantity P_{b0} used above is the equilibrium pressure in the bubble. The effective polytropic index in Eq. (4) is one-third of the real part of the function ϕ: $\gamma_{\text{eff}}=\text{Re}\,\phi/3$.

The source strength can approximately be related to the effective volume flux into the bubble at the point in time when the oscillation is initiated by interruptions of the gas-stream or by bubble separation:

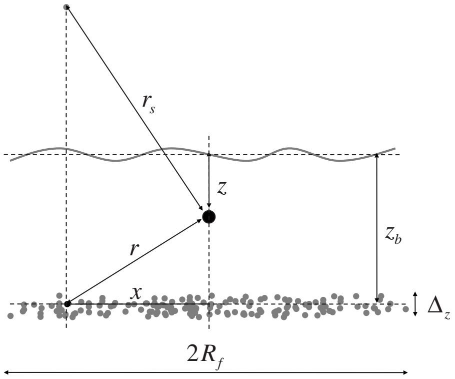


FIG. 1. Schematic diagram of the *vertically migrating fish school* geometry. The receiver depth z, the bubble release depth z_b, and the release layer thickness Δ_z are indicated, as well as the horizontal radius of the school, R_f.

$$\langle {s_b}^2\rangle = \frac{\rho^2\langle \dot{V}^2\rangle}{16\pi}\left(\frac{{\omega_0}^2}{\beta}+5\beta\right). \tag{6}$$

In our previous paper, it has been argued that the second case (bubble separation) occurs rarely, which increases the accuracy of this ansatz.

Equations (2), (3), and (6) constitute the general basis of our approach. The features related to the dynamics of individual bubbles are estimated from established experimental data and extended to different modes of behavior using our theoretical model of individual bubble release.[5] The large-scale features, related to size, geometry, and behavior, which in our formulation are contained in the dV_b-integration and in the Green's function $G_\omega(\boldsymbol{r},\boldsymbol{r}_b)$, are estimated from sonar and photographic images of gas-releasing herring schools.

B. Applications

We consider two situations that, according to what is known, could be considered canonical cases. First, we will look at a horizontally extended migratory school that vertically moves toward the surface. As a layer of fish reaches a critical depth, gas is released from the fish in this layer. The sound is observed from above the school, as would be the case in monitoring efforts of the arrival of spawning herring. Second, we will attend to a more compact and denser school that simultaneously releases gas either in response to a rapid vertical motion or, as has been suggested, to detract or communicate the presence of predators.[13]

It is important to note that the estimated spectra are not very sensitive to the precise shape of the active region since coherent phase effects can reasonably be ignored. This greatly enhances the confidence in the approach and allows for a significant simplification of the analysis.

1. Vertically migrating school

Like many other species, herring undergo diel vertical migrations. It has been observed acoustically and visually that migrating schools of herring release gas bubbles,[2] possibly to adapt swim bladder volume during hydrostatic pressure changes. Wahlberg and Westerberg[4] also observed this effect in a low-pressure chamber. There, pressure was constantly lowered from 1 atm down to 0.2 bar, which has about the same effect on the bladder volume as a vertical ascent of the fish from 100 to 40 m depth. This movement was observed to trigger gas release in Pacific herring.[2]

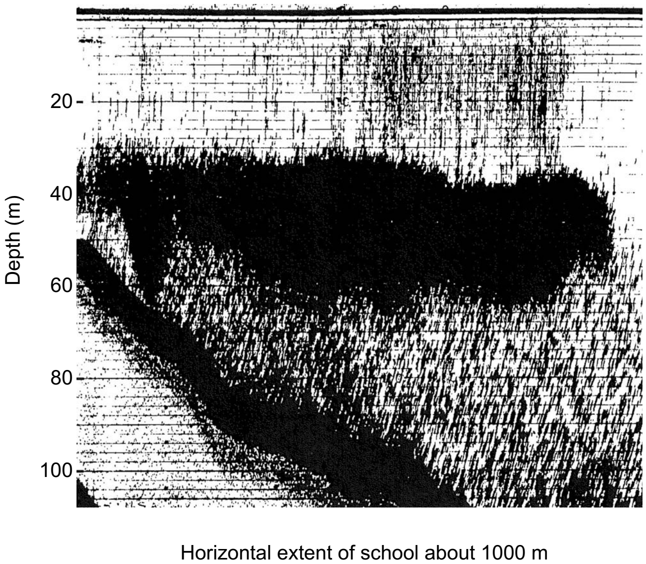


FIG. 2. Echogram of a layer of herring during upward vertical migration. The release of bubbles can clearly be seen over the dense part of the school. Reproduced from Thorne and Thomas (Ref. 2). (As indicated by these authors, the markings below the school are mixed fish and multiple scatter.)

To mathematically model this situation, we adopt the model geometry shown in Fig. 1. A vertical school of linear horizontal extent $2R_f$ moves upward such that fish in a horizontal layer at a depth of z_b and thickness Δ_z release acoustically active gas bubbles. The acoustically active bubbles are shown as gray dots in the figure. The sound is observed at a depth z directly above the school. The particular choice of this geometry is based on acoustic observations made on gas-releasing, vertically migrating herring by Thorne and Thomas.[2] Figures 2 and 3 show sample echograms from their measurements. The released bubbles are nicely picked up by the sonar in the water column above the dense part of the school. The observed parameters are given in Table I. The total duration of gas release has been estimated from visual observations in the field (the gas bubbles are visible in calm

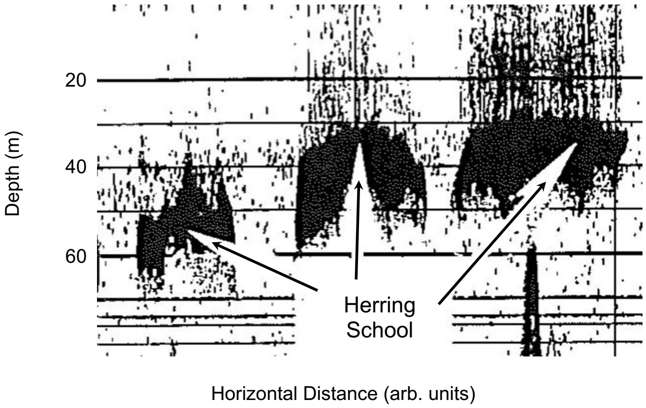


FIG. 3. Echogram of a vertically migrating school of herring over a period of 45 min. At depths of about 40 m the herring start releasing gas, which can be seen in the echogram above the gas-releasing school. Reproduced from Thorne and Thomas (Ref. 2).

TABLE I. Fish school parameters for a gas-releasing, vertically migrating herring school. (Estimated from echograms in Thorne and Thomas[2] and personal communications with these authors.)

Parameter	Symbol	Numerical value
Horizontal school size (active region)	$2R_f$	800–1200 m
Vertical extent of fish school	h_s	20–30 m
Depth of acoustically active region	z_b	35–45 m
Total duration of gas release	T_b	10–20 min
Fish school density	ρ_f	1.8–2.4 fish/m^3
Length of individual fish	l_f	21.5 cm

seas when they break the surface). The school dimension and the fish densities have been assessed from the echograms of Thorne and Thomas,[2] and the fish lengths have been measured from samples during their surveys.

Taking into account the reflections from the surface, the Green's function simply contains both the direct and the reflected propagation paths:

$$G_\omega(\boldsymbol{r},\boldsymbol{r}_b) = \frac{1}{4\pi}\left\{\frac{e^{ikr}}{r} - \frac{e^{ikr_s}}{r_s}\right\}. \quad (7)$$

In this expression, r denotes the distance from the observer to a particular point in the active volume and r_s the distance to a virtual image source that incorporates surface reflections. Because the signals add incoherently, the effect of the surface is essentially an enhancement of the received signal levels. The same is true for the effects of the sea floor on the received signal levels. However, we ignore bottom effects for simplicity and with the confidence that this leads to a more conservative estimate.

Let us assume that the event density ρ_t within the active layer is independent of position $\boldsymbol{r}_b$. In this case, the volume integral in Eq. (2) extends over the Green's function factor of the integrand only. Furthermore, we assume that $z_s \gg z$ and that the wave number $k=\omega/c$ satisfies $kz \ll 1$. Both assumptions hold sufficiently well if the point of observation is placed several meters below the surface while the active region is at a depth of some tens of meters. As is outlined in Appendix B, we then find

$$\int |G_\omega(\boldsymbol{r},\boldsymbol{r}_b)|^2 dV_b \simeq \frac{\Delta_z}{16\pi^2}\int_0^{R_f} \frac{4\pi x}{x^2+z_b{}^2}(1-\cos[k(r-r_s)])dx$$
$$\approx \frac{\Delta_z}{8\pi}\left(\log\left(1+\frac{R_f{}^2}{z_b{}^2}\right) \pm \frac{1}{kz}\frac{R_f}{z_b}\right). \quad (8)$$

The dx-integral in the first line of Eq. (8) is due to surface interference of oscillatory nature when viewed as a function of the receiver depth z. For our purpose, it is best to consider only the mean and the envelope of this function, which contribute to the upper and lower bounds on the received levels. This is what is shown in the second line of the equation and indicated by the ±-sign. The only functional dependence of the sound level on the depth is contained in the envelope terms, at least within the limits of our approximation. The mean value is independent of depth as long as the receiver is placed in the upper part of the water column.

TABLE II. Individual fish bubble dynamics parameters for gas-releasing, vertically migrating herring as well as compact "bait ball" scenarios. (Estimates from analysis in Hahn and Thomas and data from Wahlberg and Westerberg. Note that the rms volume flux and the individual herring pulse rate as given here are not independent. The large flux has been predicted from observed source levels for the small pulse rates and vice versa. This is taken into account in the given ranges. The gas-release time for the high-intensity scenario has been chosen such that 10% of the swim bladder volume is released.)

			Compact	
Parameter	Symbol	Migrating	Low intensity	High intensity
rms volume flux	$\langle \dot{V}^2\rangle^{1/2}$	0.14–4.5 ml/min	2.5 ml/min	5 ml/min
Pulse rate of individual herring (pulsed) chirps	$\dot{n}$	200–50 s^{-1}	~200 s^{-1}	~200 s^{-1}
Individual gas-release time	T_f	4.2 s	0.083 s	5.3 s

The mean spectral density for this geometry follows directly from Eqs. (2) and (8). Noting that $\dot{n}_A \equiv \rho_t \Delta_z$ is the number of pulses per time and unit area, we obtain

$$\langle S_\omega\rangle = \dot{n}_A \overline{\langle s_b{}^2\rangle |\tilde{f}_\omega|^2}\pi \log\left(1+\frac{R_f^2}{z_b^2}\right), \quad (9)$$

which forms the basis of our estimate for the vertical migration scenario.

Based on the low-pressure chamber data provided by Wahlberg and Westerberg,[4] the source strength $\langle s_b^2\rangle$ and the spectral shape $|\tilde{f}_\omega|^2$ [Eqs. (6) and (3)], describing the level of excitation and the dynamics of the individual "ringing" bubbles, respectively, have been thoroughly investigated by Hahn and Thomas.[5] The critical parameters to be addressed are the average volume flux into the bubble $\dot{V}$, the bubble size distribution affecting the resonance frequency ω_0, the damping constant β, as well as the pulse rate $\dot{n}$ for individual herring. Together with the fish density, the latter determines $\dot{n}_A$ via

$$\dot{n}_A = \dot{n}\rho_f h_f \frac{T_f}{T_b}. \quad (10)$$

For herring adapting to pressure changes, the values given in the "Migrating" column of Table II have been argued. These numbers are all based on the pressure-tank observations of individual herring source levels by Wahlberg and Westerberg.[4] The volume flux values have been extracted from these data, as laid out in Hahn and Thomas.[15] For these parameters, the prediction of the spectral density is shown in Fig. 4, which also displays expected wind-driven ambient noise background levels for various wind speeds to help interpret the results. To support these data, the cumulative wind-speed PDF for the exemplary case of winter nights in Prince William Sound, Alaska is shown in Fig. 5. Additionally, wind speed probabilities for northerly and southerly winds are summarized in Table III.

As bubbles rise toward the surface, they alter the acoustic characteristics of the medium. This, among other effects, leads to increased attenuation of the sound waves. Attenua-

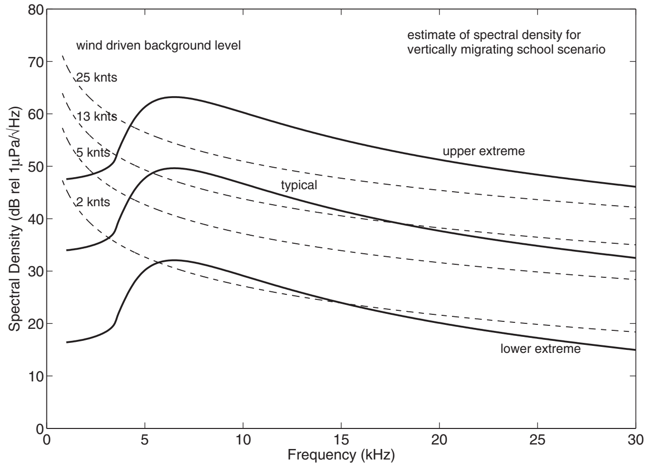


FIG. 4. Estimate of the spectral density for a vertically migrating school. The three solid lines show the prediction for the spectrum. The outer solid lines give upper and lower bounds based on the parameter range argued in the text. The middle solid line represents the prediction based on mean parameter values. The dotted lines show the expected wind-driven ambient background noise levels for wind speeds of 2, 5, 13, and 25 knots.

tion is particularly large if bubbles of resonant size are present in the water. This is not the case here for most of the propagation path: the rising bubbles grow and the ambient pressure decreases, all leading to a reduction of the resonance frequency. For water containing a bubble population of number density n_b and size-PDF $p_b(a)$, the attenuation coefficient α is given by[31,32]

$$k_e^2 = k^2 + 4\pi n_b \int_0^\infty da \frac{p_b(a)a}{\left(\frac{\omega_0^2}{\omega^2} - 1\right) + i\frac{2\beta}{\omega}},$$

$$\alpha = \operatorname{Im} k_e. \tag{11}$$

The bubble density in the water column above the fish follows directly from the total rate of released gas, which for a fish number of N_f is given by $(T_f/T_b)\langle \dot{V}^2\rangle^{1/2}N_f$, the average volume of released bubbles $\bar{V}_{a_f}$, and the terminal rise speed u_t of the bubbles:

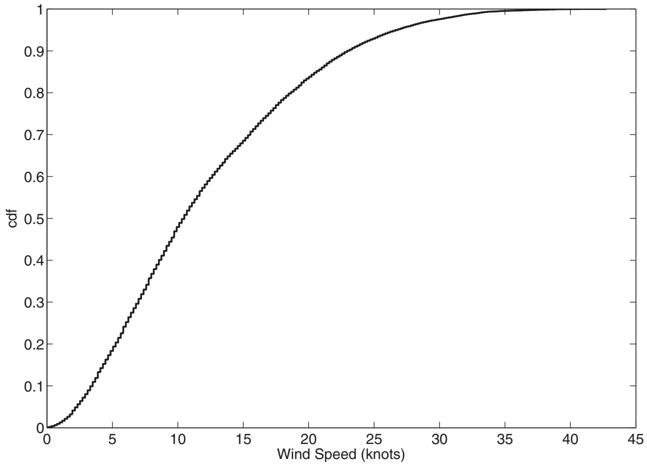


FIG. 5. Cumulative distribution function (cdf) of nighttime (18–6 h), winter (October–April) wind speeds in Prince William Sound, Alaska, obtained from the NDBC, West Orca Bay weather buoy during 2003–2006.

TABLE III. Probabilities of certain wind speeds and wind directions in Prince William Sound (Alaska) in the month of October to April during night hours (18–6 h) based on raw data from the NDBC West Orca Bay weather buoy for the years 2003–2006.

Wind-speed (knots) probabilities (%)	0–2	2–5	5–13	13–25	>25
Northerly winds (270°–90°)	2.5	10.0	32.2	12.1	1.1
Southerly winds (90°–270°)	1.7	4.2	10.6	19.6	6.0

$$n_b = \rho_f h_f \frac{T_f}{T_b}\frac{\langle \dot{V}^2\rangle^{1/2}}{u_t \bar{V}_{a_f}}. \tag{12}$$

A bubble of 1 mm radius released at 40 m has an average size of about 1.2 mm as it rises toward the surface. Such a bubble has an average terminal rise speed[33] of 28 cm/s. For this estimate, we ignore the effect of the a-dependence of the rise speed on the bubble size distribution. Using this value, our migration scenario, Table II, leads to a bubble density of 2.5 m^{-3} and a void fraction of about 1.3×10^{-8}. At any given time there is about one-half a cubic meter of gas in the water above the fish.

In our case, the attenuation coefficient α is a function of depth because it depends on the resonance frequency ω_0, Eq. (4), and on the bubble size distribution $p_b(a)$. To compute the effective attenuation, we performed a numerical analysis of the quantity

$$\alpha_{\text{eff}} = -\frac{1}{2R}\log \int d\omega \overline{|\tilde{f}_\omega|^2} e^{-2R/z_b \int dz\, \alpha(z,\omega)}. \tag{13}$$

The dz-integral extends from the surface down to the bubble release depth z_b. It takes into consideration that the pressure and bubble size distributions vary with depth. The $d\omega$-integral considers the average over the source spectrum $|\tilde{f}_\omega|^2$ that, in turn, is computed based on the distribution of acoustically *active* bubbles. For propagation distances R characteristic of our problem, we find $\alpha_{\text{eff}} \simeq 4.5\times 10^{-4}\ \mathrm{m}^{-1}$. For this purpose, all bubble size distributions have been assumed to be normal with the values for mean and standard deviation of the active, a, and *released*, a_f, bubble radii as given in the "Migrating school" column of Table IV. (The size of the released bubbles has been measured by Wahlberg and Westerberg[4] and seems to be independent of depths. From these, the statistics of the acoustically active stages is estimated according to the statistical model explained in Hahn and Thomas.[5]) With an attenuation of this magnitude, $R_f\alpha_{\text{eff}} < 1$ and the effect on the received levels is only about 2 dB. Following the analysis given in Appendix B that leads to Eq. (B7), the log-term in Eq. (8) needs to be reduced by $4\alpha_{\text{eff}}R_f$. This effect has been included in the computation of the spectra presented above.

2. Compact "bait ball"

During predator-prey interactions, many common forage fishes such as herring, anchovy, sardines, sand lance, and more can form compact schools of high densities.[34] These

TABLE IV. Bubble size distribution parameters for active and final stage (released) gas bubbles released by individual herring during sound production. (Estimates from analysis in Hahn and Thomas and data from Wahlberg and Westerberg.)

		Numerical value	
Parameter	Symbol	Migrating school (mm)	Compact schools (mm)
Mean final-stage (released) bubble radius	$\overline{a}_f$	0.96	1.2
Standard deviation of final-stage (released) bubble radii	σ_f	0.36	0.36
Mean active-stage bubble radius	$\bar{a}$	0.96	1.1
Standard deviation of active-stage bubble radii	σ_a	0.35	0.38

schools are known to exhibit complex predator-avoidance behavior, which has been the focus of attention, particularly for herring. It has been visually and acoustically observed[1] that compact schools of herring under strong predation pressure synchronously release large amounts of gas from their bladders. This gas release seems to occur over a much smaller period of time (only some seconds) than what has been observed for bubble release during vertical migration. It seems plausible[3] that such gas release plays a crucial part in these predator-avoidance behavioral modes of herring.

Albeit not directly related to the work presented here, we want to point out a physical mechanism that could play a role in this extraordinary behavior of herring. Gas release in herring affects the detection capability of their predators in at least three ways: First, because of the quick and synchronized release of gas, a bubble cloud of the dimension and shape of the school is formed. This "bubble image" of the school is visually and acoustically highly reflective and could, for a short window of time, create a false target that briefly deviates the attention of the predators away from the herring. This hypothesis is strengthened by video observations that indicate that herring seem to move *from* the bubble cloud immediately after it has been released. A snapshot of video footage showing such a situation is reproduced in Fig. 6. Judging from these pictures, this movement is not related to the diminished buoyancy due to the gas release but rather is an active horizontal escape motion. Second, gas in the bladder contributes significantly to the acoustic target strength of the herring. Common acoustic predators of herring, such as humpback whales, use frequencies mostly below 10 kHz to echolocate the schools (the ability of humpbacks to echolocate is questioned[35–37] but, based on our recordings, cannot be excluded). In this domain, the swim bladder contribution to the scattering return is dominant. The release of gas lowers this contribution and, hence, diminishes the acoustic visibility and the maximum range of detection. Third, the released bubbles are acoustically active at the moment of production which, as pointed out by Wilson *et al.*,[3] could serve to communicate danger. The trade-off will be, however, that predators might also be able to detect this noise. The following analysis will help to address such issues.

In this section, we are aiming to predict the source levels and spectra of herring bubble release with the goal of being able to identify these signals in the field and to assess the maximum range from which they can be detected against ambient noise background. Additionally, it will be necessary to understand the underlying acoustic signatures of gas release to gain insight into potential acoustic predator-prey interactions. We follow our framework synthesized in Eqs. (2)–(6). To this purpose, we need to address the three physical ingredients of this model. What is the underlying excitation of the bubbles released by the individual fish? What is the size distribution of the acoustically active bubbles? And what is the relevant Green's function capturing the acoustic environment including the approximate shape, density, and volume of the acoustically active region? To answer the first two questions, we must rely on the observations of Wahlberg and Westerberg[4] and on our previously presented sound production model,[5] which allows us to extrapolate the data beyond that of the pressure tank of Wahlberg and Westerberg,[4] where it originated. We will turn to these questions below.

To decide on the Green's function and on the geometry, we first want to mention that attenuation could play a more significant part because of the, in this case, much denser

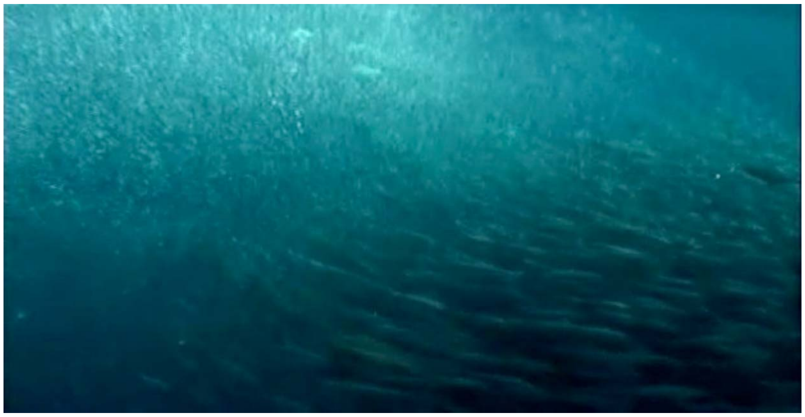


FIG. 6. (Color online) Photographic image of a compact school of herring, synchronously releasing gas during an attack of killer whales. Reproduced from the BBC documentary "The Blue Planet" with permission of the BBC Motion Gallery.

schools and bubble clouds. In essence, sound originating in the parts of the school that are most remote from the observer will be shielded by the rest of the active region. Second, the overall school *shape* aspect is the least important feature as it does not alter the *incoherently* added overall intensities much. We will explicitly see this in the computed Green's function integrals. However, the total school volume and the fish density are crucial since they, together with the dynamics of individual fish gas release, determine the overall acoustic event rates.

Once the source levels are determined, the received spectral levels can be estimated from sound propagation models that incorporate boundaries at the sea surface and at the sea floor of the acoustic environments in which this herring behavior is observed to occur. Here, we will simply project the source levels, once they have been obtained, to points of observation by applying simple scaling laws accounting for geometric spreading.

Before we go into details, let us begin with a back-of-the-envelope estimate of the spectrum, Eq. (2), which will capture the essence of this scenario. If we assume that the density ρ_t of acoustic events is constant across the school, the integral in Eq. (2) based on the free field Green's function will, at large distances r to the observer, simply be

$$\int dV_b |G_0(\boldsymbol{r},\boldsymbol{r}_b;\alpha)|^2 = \frac{e^{-2\alpha\langle l\rangle}}{16\pi^2}\frac{V_b}{r^2}. \tag{14}$$

In this expression, attenuation is included over an average propagation distance $\langle l\rangle$ within the school. For a sphere, $\langle l\rangle = 2\pi\int_0^R r^2 dr \int_{-1}^1 dx(\sqrt{(R^2-r^2)+r^2x^2}-rx) = 3/4R$ if rays origin randomly *within*. This is in slight contrast to the average chord length *through* a sphere,[38] which is $4/3R$. The attenuation coefficient α depends on the bubble *and* bladder size distribution within the herring school, which we will discuss below. For completeness, we want to mention that if distances to the point of observation are *not* large compared to size of the school, the right hand side of Eq. (14) needs to be multiplied with a geometrical factor, which for a sphere is readily shown to evaluate to $3/2\varepsilon^{-3}(\varepsilon-(1-\varepsilon^2)\tanh^{-1}\varepsilon) \to 1$ for small ε, where $\varepsilon = R/r$.

This coarse approach does not include effects that are due to the impedance contrast between inside and outside the school. This has the effect of a filter, altering the spectral shape of the signal. Although this effective filter depends modestly on the precise school shape, it is only important at very low frequencies for schools sizes of order of some meters or more.

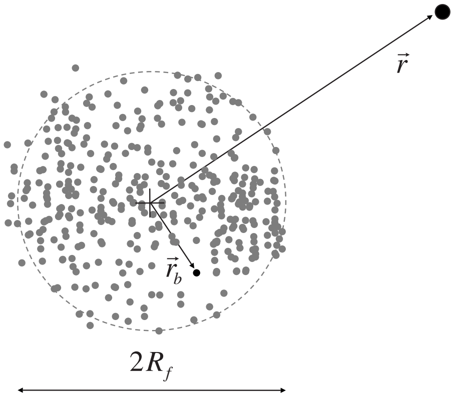


FIG. 7. Schematic diagram of the *compact bait-ball* geometry.

To see this, we consider a spherical shape, with the geometry shown in Fig. 7 and parameters chosen according to Table V, for a more detailed analysis. As outlined in Appendix A, the full Green's function for a soft sphere is

$$G_\omega(\boldsymbol{r},\boldsymbol{r}_b) = k_e \sum_{nm} Y_n{}^m(\vartheta,\varphi) Y_n{}^{m*}(\vartheta_b,\varphi_b) h_n{}^{(1)}(kr) j_n(k_e r_b)\alpha_n, \tag{15}$$

from which the necessary integral easily follows in terms of an infinite sum:

$$\begin{aligned} I &= \int |G_\omega(\boldsymbol{r},\boldsymbol{r}_b)|^2 dV_b \\ &= \frac{R_f}{8\pi}(k_e R_f)^2 \sum_n (2n+1)|h_n{}^{(1)}(kr)|^2|\alpha_n|^2 \\ &\quad \times \{j_n{}^2(k_e R_f) - j_{n-1}(k_e R_f) j_{n+1}(k_e R_f)\} \\ &\underset{r\to\infty}{\simeq} \frac{1}{16\pi^2}\frac{4\pi R^3}{3r^2}\left(\frac{k_e}{k}\right)^2 \frac{3}{2}\sum_n (2n+1)|\alpha_n|^2 \\ &\quad \times \{j_n{}^2(k_e R_f) - j_{n-1}(k_e R_f) j_{n+1}(k_e R_f)\}. \end{aligned} \tag{16}$$

The second line follows from the asymptotic expansion of $h_n^{(1)}$ for large arguments.[39] The sum in Eq. (16) can be computed numerically without complication, the number of terms that need to be carried is of order of kR. Attenuation is naturally included if k_e is extended to the complex domain. Based on physical grounds, the limit of Eq. (16) for Re $k_e = k$ is Eq. (14) with the given expression of $\langle l\rangle$ for the sphere.

Figure 8 shows the results of the computation of the integral I for a dense school of 6 m diameter corresponding to high-intensity gas release. The parameter values are given in the "High intensity" column of Table II. To demonstrate the effect of attenuation and spectral shaping, this scenario is chosen with a significantly more intense gas release than what has been observed experimentally. An upper-bound value for the volume flux is taken, and the total released gas volume amounts to 10% of the swim bladder volume. From the figure, we first note that the impedance contrast—the only influence of the spherical shape—does not critically influence the integral over the considered frequency domain.

TABLE V. Fish school parameters used for the computation of the source levels of compact gas-releasing herring schools.

		Compact	
Parameter	Symbol	Low intensity	High intensity
Size of fish school	$2R_f$	30 m	6 m
Number of fish	N_f	~212 000	~1700
Total duration of gas release	T_b	10 s	10 s
Depth of fish school	z_b	3 m	3 m
Fish school density	ρ_f	15 fish/m^3	15 fish/m^3
Length of individual fish	l_f	21.5 cm	21.5 cm

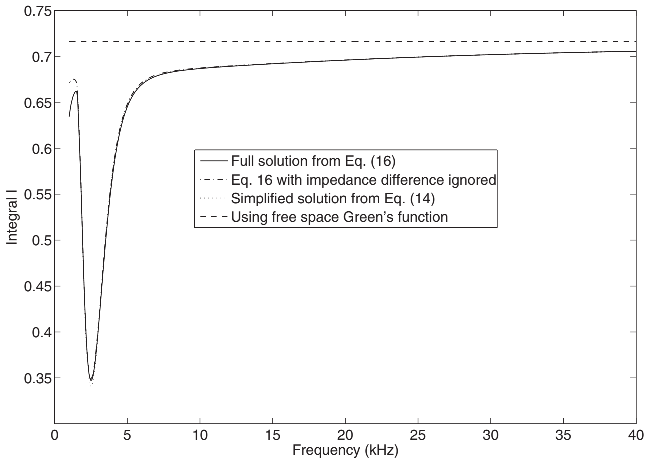


FIG. 8. Green's function integral for compact fish schools. The solid line represents the evaluation according to Eq. (16) with fully evaluated k_e, the dash-dotted line shows the same result, however, with impedance differences ignored, and the dotted line (graphically mostly indistinguishable from the dash-dotted line) gives the result from approximation Eq. (14). The integral is evaluated for the high-intensity release scenario. The dashed line shows, for reference, the value resulting from using the free-space Green's function G_0.

Second, for dense schools and significant gas release, attenuation of sound in the school is an important effect.

To capture the attenuation and sound speed in the fish school correctly, two bubble populations were taken into consideration in this computation: the released gas bubbles and the swim bladders. For the released bubbles, the size statistics[5] given in Table IV has been used and a number density of

$$n_b = \frac{T_f \langle \dot{V}^2 \rangle^{1/2} V_b \rho_f}{(V_b + R_f^2 \pi u_t T_b) \bar{V}_{a_f}}, \tag{17}$$

which takes into account the gradual release of bubbles. The bladder sizes have been computed according to the predominant length of herring and, derived from that, the expected bladder eccentricities.[40] From these, the small amplitude bladder dynamics, ω_0 and β, can be computed in close analogy to gas bubbles.[41] The density of bladders is, of course, equal to the fish density ρ_f. The result of this is shown in Fig. 9 for the high-intensity scenario as well as for a large school of about 200 thousand fish that releases gas according to Wahlberg and Westerberg's[4] field data, which are indicated as "Low intensity" in Table V. The velocity u_t indicates the relative velocity between the released bubbles and the herring school and is taken to be 2 m/s in the computations. This dilutes the bubble densities in the fish school somewhat but takes into account the observed agitated behavior of herring schools under attack.

Putting everything together, the source levels follow again from Eq. (2), formally evaluated at $r=1$, supplied with the acoustic event density

$$\rho_t = \frac{\rho_f \dot{n} T_f}{T_b}. \tag{18}$$

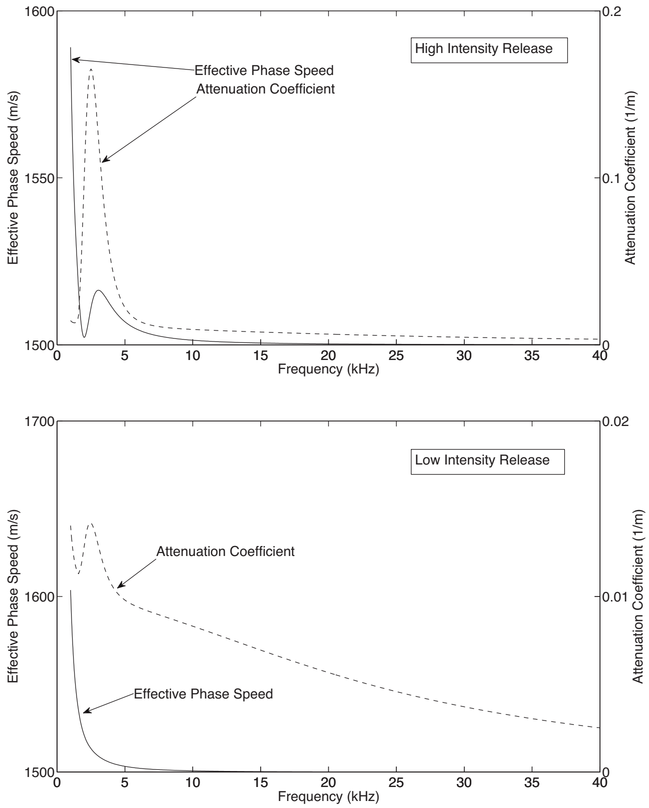


FIG. 9. Effective phase speeds and attenuation coefficients for the high-intensity (top panel) and low-intensity (bottom panel) compact fish school scenario.

Figure 10 shows the final result. For reference, the ambient background noise levels are also indicated as well as the expected transmission loss due to geometrical spreading for 20 log r and 15 log r average transmission loss assumptions. Both scenarios result in about the same source level but based on different premises. With the slight exception of visible attenuation signatures close to the spectral peak frequencies, it is not possible to differentiate a dense school releasing a significant fraction of the bladder gas from a mas-

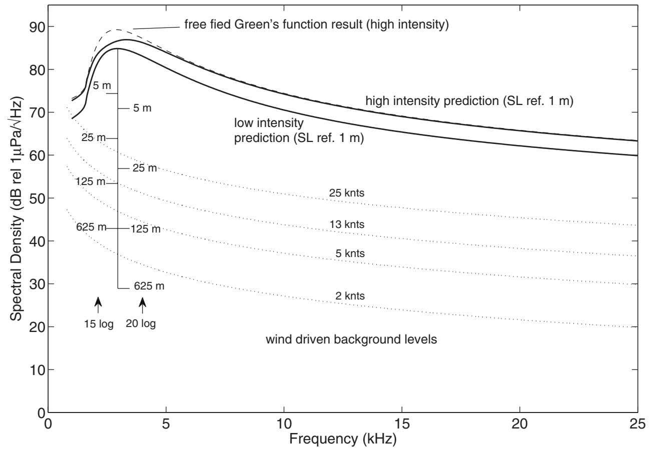


FIG. 10. Estimate of the source level for the high-intensity and low-intensity compact fish school scenarios. To estimate the received spectral densities, the intensity loss due to geometric spreading is also indicated for 1, 5, 25, 125, and 625 m, both for 20 log and 15 log spreading laws. The dashed line shows the high-intensity result based on the free-field Green's function. The dotted lines indicate the expected wind-driven ambient background noise levels for wind speeds of 2, 5, 13, and 25 knots.

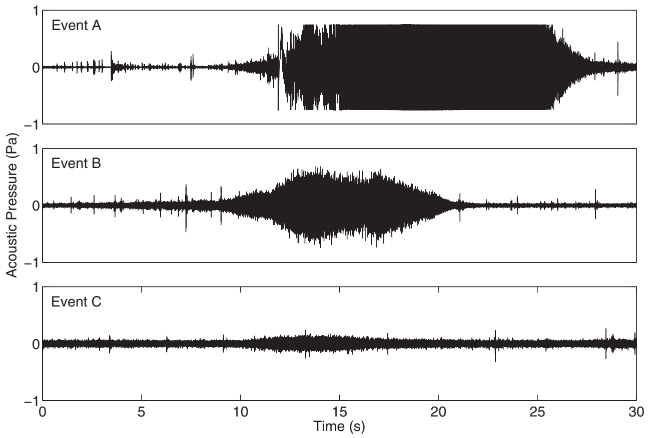


FIG. 11. Three potential gas-release events of compact herring schools.

sive school releasing only some few ml/min over a fraction of a second.

3. Field recordings and comparison with theory

Recordings of underwater sound in the presence of large aggregations of herring were made in Prince William Sound, Alaska in April 2007. Based on a preliminary analysis of acoustic survey data, several thousand metric tons of herring were present in the sound as well as 15–20 humpback whales, several hundred sea lions, and thousands of sea birds, all feeding on the herring.

We used an ITC 6050C hydrophone (with a flat receive response at the frequency range of interest) deployed about 10 m below a 12 ft skiff, together with a Reson VP2000 voltage preamplifier to record a total of about 2 h of broadband data on a Sony TCD-D7 portable digital tape recorder for subsequent analysis. No underwater visual observations were made during the recording. However, the feeding activity of three large humpback whales was evident in the immediate neighborhood of the skiff, and a prior acoustic survey indicated the presence of about 1000 metric tons of herring in the vicinity of the recording location. The sea was calm during the entire duration of the recording.

Based on the expected signatures, primarily the sound duration and the shape of the spectrum, we were able to identify eight events that most likely correspond to the "compact school" scenario. The time-series of three of these events (A, B, and C) are shown in Fig. 11. The event depicted in the upper panel (event A) was the strongest observed. Interestingly, it was immediately preceded by the noise of the humpbacks (communication or possibly echolocation) and followed by herring jumping out of the water around the skiff. As shown below (for events A and B), the spectral and temporal signatures for these events are very similar. The overall received levels are different, however, because of varying source distances (visual cues of this were the location of the feeding humpbacks), which were not measured. Event C is given as an example of a gas-release event in the distance, such that its signature appears just above the background. The strong distortion of the signal at about 12 s might indicate that the school of herring physically touched the hydrophone as it was moving upward. Gas bubbles were visible on the sea surface surrounding the skiff shortly after the bubble noise was audible on the headset. The bubbles observed on the surface were mainly millimeter-sized, and no larger bubbles were visible in the direct vicinity of the skiff. Visually assessed, the footprint of the bubbles on the surface spanned several tens of meters.

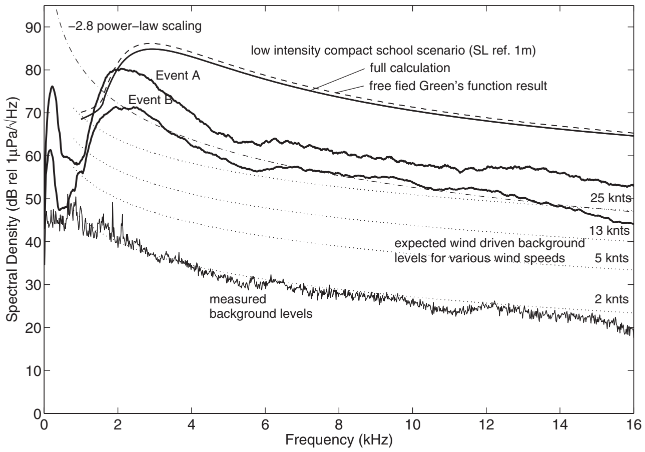


FIG. 12. Spectral densities of two potential gas-release events of compact herring schools corresponding to the two upper panels (events A and B) of Fig. 11, as well as the spectral density of the ambient noise background. For comparison, the expected spectral source level for the low-intensity release compact school from Fig. 10 is also shown (heavy line). The dashed line shows the model computation without the effects of attenuation and impedance contrast. Overlaid on the measured spectrum of the lower-level event is the predicted −2.8 power-law prediction for the high-frequency tail of the spectrum (dash-dotted line). The spectrum indicated for the event in the top row of Fig. 11 is calculated for the later part of the event, where clipping was minor.

Let us for now assume that these events are indeed due to bubble release of a compact school of herring under strong predation pressure. Figure 12 depicts the spectra of events A and B and allows a comparison with the predictions given above. The measured spectra show a strong similarity to our prediction. The spectrum peaks in the neighborhood of 2 kHz and gently rolls off toward higher frequencies, in close resemblance to the predicted power-law decay.[5] At low frequencies, there is a sharper roll-off toward the smaller frequencies. The theoretical curve, chosen to be the low-intensity scenario, also features most of these characteristics. There are, however, some notable deviations. First, the main peak occurs at a slightly higher frequency. Our model parameters have been largely chosen according to the tank data of Wahlberg and Westerberg.[4] A slight increase in the mean radii of the active bubble distribution could account for this shortcoming. A release depth shallower than assumed would help too. In addition, the peak is slightly more pronounced as compared to the model computations.

We also note that there seems to be no indication of significant attenuation effects, which would alter the shape of the peak. This could indicate that the low-intensity gas-release scenario is more likely appropriate. This, independently, is also hinted by the size of the bubble footprint on the sea surface.

We finally want to point out that the overall levels of our prediction are quite reasonable. From visual observations of surfacing herring and whales, we know that the recorded

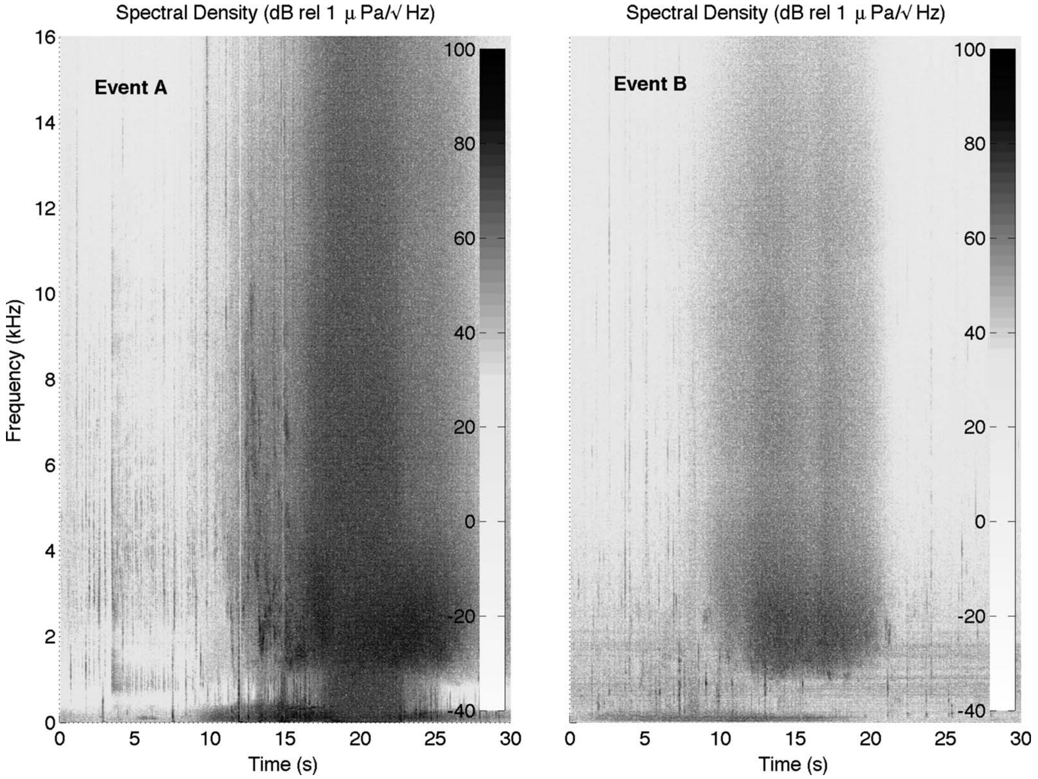


FIG. 13. Periodogram of the top two events (A and B) from Fig. 11. The left panel corresponds to the large signal in the top row. The time axis runs horizontally and displays half a minute of data for each event. The shading is adapted to the spectral intensity in dB rel. 1 μPa/ $\sqrt{}$Hz according to the displayed color bars.

herring were in direct vicinity of the recording location. Hence, the recorded sound levels should be close to the predicted spectral levels. The computation for the low-intensity scenario is based on about 200 thousand herring, more or less simultaneously releasing bubbles. This corresponds to about 40 metric tons of herring, or 4% of all herring in the vicinity of the skiff.

It has been observed that humpback whales can release bubbles while hunting for herring.[42–44] Despite the convincing argument that can be made, care has to be taken to be sure that the recorded signal is indeed due to the herring and not caused, for example, by the hunting activities of the present humpback whales. From an acoustic point of view, not much is known about humpback "bubble curtains." While the detailed bubble size distribution within humpback bubble curtains is unknown, photographic evidence and observations of raising whale bubbles suggest that these bubbles are, on the upper side of the distribution, at least centimeter-sized.[43] Large rising spherical caps, several tens of centimeters in diameter, have been observed at field surveys and also by fishermen when whales are active. Bubbles of this size, even when released at a depth of 60 m (which was the depth of the seafloor at the site of our observations) would ring below 1 kHz. Clearly, small active bubbles could be produced through breakup. In this case, there should be a power-law-like spectrum extending down to frequencies below 1 kHz. We did not observe this. It is noteworthy that the footprint of humpback bubble curtains at the sea surface[43–45]—that often resemble a net surrounding their prey—looks markedly different compared to our visual observations of large compact areas covered with small rising bubbles. Finally, we want to point out again that the particular high-frequency roll-off of the recorded spectrum is consistent with the proposed sound production mechanism.

The periodograms of events A and B are displayed in Fig. 13. Again, we note the rather sharp decline in spectral power below about 2 kHz that was predicted by our model. Especially important for future monitoring applications is the fact that both events (and several others not displayed) have very similar time- and frequency-domain signatures, which also becomes evident when exploring both events featured in Fig. 12.

Also noticeable are low-frequency bands at a few hundred hertz that are detached from the main events and that appear several seconds before the hypothesized herring bubble release occurs. We can only speculate on the origin of these bands. One possibility would be to attribute this noise to large bubbles released by whales. In this light, it might be useful to point out the coda of whale sounds that occurs immediately before the low-frequency bands appear (broadband clicks from 0 to 10 s in the high-intensity event). Another possibility to explain these low-frequency bands could be collective oscillations of the fish school. This, however, seems unlikely because of improper timing with respect to the main gas-release event.

Additionally, we note a broadband signal, from about 17 to 23 s, embedded in the first event that appears to be of different origin than the rest. During this time interval, we note a broadband component that stretches from very low frequencies into the domain of the gas-release event where it lifts the spectral levels. This could be the effect of strong signal clipping that occurred during early recordings, as the source was closer than expected.

To summarize, the spectral features of the main signatures agree well with all our predictions for compact herring school bubble release. However, we cannot be absolutely certain that the observed "candidate" events are indeed herring. Evidence that they are, nonetheless, is as strong as it can be without visual confirmation. Additional features hint that interesting predator-prey dynamics could be illuminated with simple acoustic techniques, but more comprehensive simultaneous measurements including video and high-frequency surveying acoustics are necessary as a preparatory step. Furthermore, an initial analysis hints that scattering of the strong noise generated by the bubble release from herring in the vicinity could reveal herring sizes from passive measurements! This will be the subject of further research.

III. DISCUSSION AND OUTLOOK

The main purpose of this study is to explore the possibility of detecting herring with simple passive broadband acoustic means to complement traditional active higher-frequency surveying techniques. Since passive monitoring platforms could be permanently deployed in locations that are known to be critical habitats for this species, remote monitoring and management of herring are now possible.

Based on the presented theoretical examples, we conclude that passive detection is feasible. This is true both for the vertically migrating and for the predator response scenario. First measurements in Prince William Sound, Alaska confirm this assessment. Within a small period of 2 h, several potential events could easily be identified and often be visually matched to predatory activities of humpback whales at ranges of up to a mile, basically covering an entire embayment where herring was also measured and assessed as part of a traditional herring survey. The observed signatures have characteristic features that indicate the presence of herring without the need for visual confirmation.

To reach this goal, an extension of this work is necessary, mainly on the observational side. The observed events need to be unambiguously matched with herring gas release by visual or independent simultaneous high-frequency acoustics measurements. This is our next objective as we progress from this starting point toward the goal of implementing passive herring monitoring platforms.

Finally, we want to point out again that, in principle, much more than the simple presence of herring can be learned from observations such as the ones presented. Many predators of herring communicate and target acoustically. Since the gas-release response of herring also has a large acoustic signature, deep insight into the dynamics of this interaction can be gained acoustically. The insights will be even bigger when passive acoustics is combined with underwater and surface visual observations.

ACKNOWLEDGMENTS

The authors would like to thank Dick Thorne for many helpful discussions and financial support through a grant from the NMFS, Rob Cermak from the Alaska Ocean Observing System providing support in obtaining and understanding PWS wind data, and Otis Brown, the dean of the Rosenstiel School, for his assistance and support in developing this multidisciplinary line of research. This work was supported by Dr. Ellen Livingston, the Office of Naval Research.

APPENDIX A: SOFT SPHERE GREEN'S FUNCTION

The computation of the Green's function $G_\omega(\boldsymbol{r},\boldsymbol{r}_b)$ for a soft sphere is a standard problem that can be found, for example, in Morse and Ingard.[46] For ease of reference, we give the result. For a (minus) unit strength point source at a location $\boldsymbol{r}_b$ *inside* the sphere, one finds for the acoustic field at a point of observation $\boldsymbol{r}$ outside the sphere

$$G_\omega^{>}(\boldsymbol{r},\boldsymbol{r}_b)=\sum_{nm}A_{nm}Y_n{}^m(\vartheta,\varphi)h_n^{(1)}(kr). \tag{A1}$$

In this and the following expressions, $h_n^{(1)}$ are the spherical Hankel functions of the first kind, Y_n^m are the spherical harmonics, and (r,ϑ,φ) and $(r_b,\vartheta_b,\varphi_b)$ are the spherical coordinate representations for $\boldsymbol{r}$ and $\boldsymbol{r}_b$, respectively. For the coefficients A_{nm}, one finds

$$A_{nm}=k_e\alpha_nY_n{}^{m*}(\vartheta_b,\varphi_b)j_n(k_er_b) \tag{A2}$$

together with the following definitions:

$$\alpha_n=i\frac{h_n{}^{(1)}(k_er)}{h_n{}^{(1)}(kr)}\frac{\beta_n{}^h-\beta_n{}^j}{\bar{\beta}_n{}^h-\beta_n{}^j},$$

$$\beta_n{}^b=i\frac{\rho c}{\rho_ec_e}\frac{h_n{}^{(1)\prime}(k_er)}{h_n{}^{(1)}(k_er)},$$

$$\beta_n{}^j=i\frac{\rho c}{\rho_ec_e}\frac{j_n{}'(k_er)}{j_n(k_er)},$$

$$\bar{\beta}_n{}^h=i\frac{h_n{}^{(1)\prime}(kr)}{h_n{}^{(1)}(kr)}. \tag{A3}$$

In the limit $c=c_e$ and $\rho=\rho_{el}$, we have $\alpha_n=i$, which leads to $G_\omega^{>}(\boldsymbol{r},\boldsymbol{r}_b)\rightarrow G_0(\boldsymbol{r},\boldsymbol{r}_b;k_e)$, as expected. The subscript e identifies the corresponding effective quantities inside the sphere.

APPENDIX B: COMPUTATION OF TWO DIMENSIONAL GREEN'S FUNCTION INTEGRAL

The integral in Eq. (8), which at the frequency range of interest is the product of a rapidly and a slowly varying function, allows for a simple estimate of its mean value and of the envelope of the fluctuations for vales of $z<z_b$. Noting that (the bar denotes the average over a full cycle of the oscillatory factor of the integrand)

$$\overline{(1-\cos[k(r-r_s)])}=1, \tag{B1}$$

we find that the mean value of the integral $\bar{I}$ is

$$I = \int_0^{R_f} 2\pi x \left| \frac{e^{ikr}}{r} - \frac{e^{ikr_s}}{r_s} \right|^2 dx$$

$$\simeq \int_0^{R_f} \frac{4\pi x}{x^2 + {z_b}^2}(1 - \cos[k(r - r_s)])dx = \bar{I} + \Delta I,$$

$$\bar{I} = \int_0^{R_f} \frac{4\pi x}{x^2 + {z_b}^2} dx = 2\pi \log\left(1 + \frac{R_f^2}{z_b^2}\right). \qquad \text{(B2)}$$

The fluctuations of the integral, ΔI, are determined by the behavior of the integrand at the endpoints of integration. To see this, we change variables to the phase $\varphi = k(r - r_s)$; both r and r_s are functions of x and z:

$$\Delta I(z) = 4\pi \int_{\varphi_{\min}(z)}^{\varphi_{\max}(z)} F(\varphi, z)\cos \varphi d\varphi,$$

$$F(\varphi, z) = \frac{x(\varphi, z)}{x(\varphi, z)^2 + z_b^2} \frac{1}{\varphi'(z)}. \qquad \text{(B3)}$$

The fluctuation integral ΔI contains the receiver depth z as a parameter in the slowly varying function F as well as in the endpoints of integration. The prime in the second line denotes differentiation with respect to x. Explicitly expressing $r - r_s$ (in the phase) as $\sqrt{x^2 + (z - z_b)^2} - \sqrt{x^2 + (z + z_b)^2}$, we obtain

$$F(\varphi, z) = \frac{1}{\varphi} \frac{\varphi^4 - 16k^4 z^2 z_b^2}{\varphi(\varphi^4 + 16k^4 z^2 z_b^2 - 4k^2 z^2 \varphi^2)} \simeq \frac{1}{\varphi},$$

$$\varphi_{\min}(z) = -2kz,$$

$$\varphi_{\max}(z) = -k\{\sqrt{R_f^2 + (z - z_b)^2} - \sqrt{R_f^2 + (z + z_h)^2}\} \simeq -2kz \frac{z_b}{R}. \qquad \text{(B4)}$$

In this calculation $z < z_b$ has been assumed to approximate F and $z_b < R_f$ to simplify $\varphi_{\max}(z)$. From Eq. (B4) we finally arrive at a compact expression for the fluctuations in terms of the cosine integral $\mathrm{Ci}(z) = -\int_z^\infty (\cos t/t)dt$:

$$\Delta I(z) = 4\pi \left[\mathrm{Ci}\left(2kz \frac{z_b}{R}\right) - \mathrm{Ci}(2kz) \right]. \qquad \text{(B5)}$$

Under our assumptions, the first part dominates in magnitude and modulates with a wavelength of $(R/2z_s)$ times the acoustic wavelength. The second term is of smaller magnitude and oscillates with half the acoustic wavelength. For our purpose it suffices to state, using Eq. (B5), that

$$|\Delta I| < \frac{2\pi}{kz} \frac{R_f}{z_b}. \qquad \text{(B6)}$$

To consider the effect of bubbles in the water column above the gas-releasing fish, absorption can be incorporated into this framework in a straightforward manner. The simplest, and for our purposes sufficient, approach is to include an imaginary part α in the wave number k of Eq. (7). Within the domain of our approximations, Eq. (B2) becomes

$$I = \int_0^{R_f} \frac{4\pi x}{x^2 + {z_b}^2} e^{-2\alpha\sqrt{x^2 + {z_b}^2}}(1 - \cos[k(r - r_s)])dx = \bar{I} + \Delta I,$$

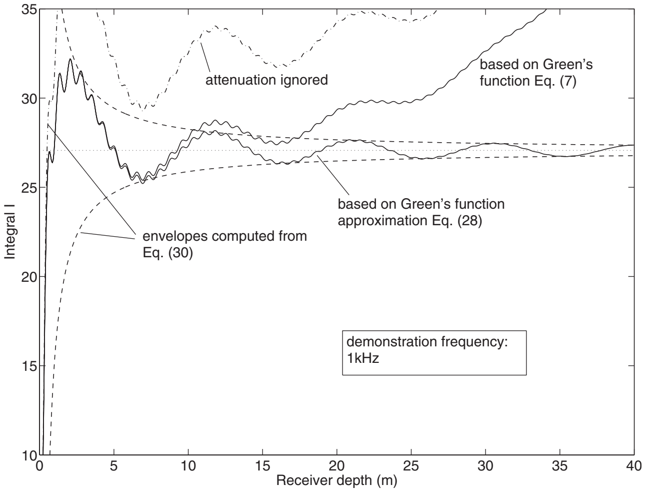


FIG. 14. Green's function integral I including the effect of (average) attenuation for the mean migrating fish school scenario. The solid lines show the full solution based on Green's function Eq. (7) (increasing toward larger receiver depth), as well as the approximate solution, Eq. (B7). The dotted and dashed lines, respectively, indicate the mean value and the envelopes, Eq. (B9). The dash-dotted line shows the integral without attenuation. The computation has been performed for demonstration at a frequency of 1 kHz to make the oscillations more visible.

$$\bar{I} = \int_0^{R_f} \frac{4\pi x}{x^2 + {z_b}^2} e^{-2\alpha\sqrt{x^2 + {z_b}^2}} dx$$

$$= -4\pi \mathrm{Ei}(-2\alpha z_b) + 4\pi \mathrm{Ei}(-2\alpha\sqrt{R_f^2 + {z_b}^2})$$

$$\underset{\alpha R_f \gg 1}{\simeq} -4\pi \mathrm{Ei}(-2\alpha z_b) \underset{\alpha R_f \ll 1}{\simeq} 2\pi \log\left(1 + \frac{R_f^2}{z_b^2}\right) - 8\pi\alpha R_f. \qquad \text{(B7)}$$

As expected, if attenuation is present the logarithmic divergence as $R_f \to \infty$ is removed and the limiting value of the integral depends only on αz_b. The exponential integral function used above is defined in the standard way as[47] $\mathrm{Ei}(z) \equiv -\int_{-z}^\infty e^{-t}/t dt$. For small negative values of the argument,[47] $\mathrm{Ei}(x) = \gamma + \ln(-x) + \Sigma_{k=1}^\infty (x^k / k \cdot k!)$, from which the limiting form for small values of αR_f follows.

The magnitude of the fluctuations of the integral with varying depth, $|\Delta I|$, can be estimated using the same ideas as above once we note that in the presence of attenuation ΔI has the following approximate integral representation:

$$\Delta I = \int_{2kzz_b/R_f}^{2kz} e^{-4\alpha(kzz_b/\varphi)} \frac{\cos \varphi}{\varphi} d\varphi. \qquad \text{(B8)}$$

From this we infer

$$|\Delta I| < \frac{2\pi}{kz} \frac{e^{-2\alpha R_f} R_f}{z_b}. \qquad \text{(B9)}$$

The quality of various approximations involved is shown in Fig. 14.

[1] L. Nottestad, "Extensive gas bubble release in Norwegian spring-spawning herring (*Clupea harengus*) during predator avoidance," ICES J. Mar. Sci. **55**, 1133–1140 (1998).

[2] R. E. Thorne and G. L. Thomas, "Acoustic observations of gas bubble release by Pacific herring (*Clupea-harengus-pallasi*)," Can. J. Fish. Aquat. Sci. **47**, 1920–1928 (1990).

[3] B. Wilson, R. S. Batty, and L. M. Dill, "Pacific and Atlantic herring

produce burst pulse sounds," Proc. R. Soc. London, Ser. B **271**, S95–S97 (2004).

[4] M. Wahlberg and H. Westerberg, "Sounds produced by herring (*Clupea harengus*) bubble release," Aquat. Living Resour. **16**, 271–275 (2003).

[5] T. R. Hahn and G. L. Thomas, "Modeling the sound levels produced by bubble release of individual herring," J. Acoust. Soc. Am. **124**, 1849–1857 (2008).

[6] R. A. Rountree, R. G. Gilmore, C. A. Goudey, A. D. Hawkins, J. J. Luczkovich, and D. A. Mann, "Listening to fish: Applications of passive acoustics to fisheries science," Fisheries **31**, 433–446 (2006).

[7] R. Rountree, C. Goudey, and T. Hawkins, in *Listening to Fish: An International Workshop on the Applications of Passive Acoustics in Fisheries*, edited by R. Rountree, C. Goudey, T. Hawkins, J. J. Luczkovich, and D. Mann (Sea Grant, Cambridge, MA, 2002), pp. 4–10, http://web.mit.edu/seagrant/digitalocean/listening.pdf.

[8] M. A. Connaughton and M. H. Taylor, "Seasonal and daily cycles in sound production associated with spawning in the weakfish, cynoscion regalis," Environ. Biol. Fish. **42**, 233–240 (1995).

[9] J. J. Luczkovich and M. W. Sprague, in *Listening to Fish: An International Workshop on the Applications of Passive Acoustics in Fisheries*, edited by R. Rountree, C. Goudey, T. Hawkins, J. J. Luczkovich, and D. Mann (Sea Grant, Cambridge, MA, 2002), pp. 59–63, http://web.mit.edu/seagrant/digitalocean/listening.pdf.

[10] J. Ramcharttar, D. P. Gannon, and A. N. Popper, "Bioacoustics of fishes of the family sciaenidae (croakers and drums)," IEEE Trans. Power Syst. **135**, 1409–1431 (2006).

[11] A. D. Hawkins and M. C. P. Amorim, "Spawning sounds of the male haddock, *Melanogrammus aeglefinus*," Environ. Biol. Fish. **59**, 29–41 (2000).

[12] A. D. Hawkins, L. Casaretto, M. Picciulin, and K. Olsen, "Locating spawning haddock by means of sound," Bioacoustics **12**, 284–286 (2002).

[13] I. Svellingen, B. Totland, and J. T. Oevredal, "A remote-controlled instrument platform for fish behaviour studies and sound monitoring," Bioacoustics **12**, 335–336 (2002).

[14] S. Allen and D. A. Demer, "Detection and characterization of yellowfin and bluefin tuna using passive-acoustical techniques," Fish. Res. **63**, 393–403 (2003).

[15] R. A. Charif, P. J. Clapham, and C. W. Clark, "Acoustic detections of singing humpback whales in deep waters off the British Isles," Marine Mammal Sci. **17**, 751–768 (2001).

[16] C. W. Clark and P. J. Clapham, "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring," Proc. R. Soc. London, Ser. B **271**, 1051–1057 (2004).

[17] T. F. Norris, M. Mc Donald, and J. Barlow, "Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the Eastern North Pacific during their northbound migration," J. Acoust. Soc. Am. **106**, 506–514 (1999).

[18] A. S. Burgess and D. J. Kewley, "Wind-generated surface noise source levels in deep-water east of Australia," J. Acoust. Soc. Am. **73**, 201–210 (1983).

[19] B. R. Kerman, "Underwater sound generation by breaking wind-waves," J. Acoust. Soc. Am. **75**, 149–165 (1984).

[20] A. R. Kolaini, "Sound radiation by various types of laboratory breaking waves in fresh and salt water," J. Acoust. Soc. Am. **103**, 300–308 (1998).

[21] W. M. Carey and M. P. Bradley, "Low-frequency ocean surface noise sources," J. Acoust. Soc. Am. Suppl. **78**, S1 (1985).

[22] W. M. Carey and J. W. Fitzgerald, "Low-frequency noise and bubble plume oscillations," J. Acoust. Soc. Am. Suppl. **82**, S62 (1987).

[23] T. R. Hahn, T. K. Berger, and M. J. Buckingham, "Acoustic resonances in the bubble plume formed by a plunging water jet," Proc. R. Soc. London, Ser. A **459**, 1751–1782 (2003).

[24] A. Prosperetti, in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, 1988), pp. 151–171.

[25] P. Tkalich and E. S. Chan, "Breaking wind waves as a source of ambient noise," J. Acoust. Soc. Am. **112**, 456–463 (2002).

[26] M. S. Longuet-Higgins, "An analytical model of sound production by raindrops," J. Fluid Mech. **214**, 395–410 (1990).

[27] H. C. Pumphrey and P. A. Elmore, "The entrainment of bubbles by drop impacts," J. Fluid Mech. **220**, 539–567 (1990).

[28] T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).

[29] A. Prosperetti, "Bubble phenomena in sound fields: Part one," Ultrasonics **22**, 69–77 (1984).

[30] A. Prosperetti, L. A. Crum, and K. W. Commander, "Nonlinear bubble dynamics," J. Acoust. Soc. Am. **83**, 502–514 (1988).

[31] K. W. Commander and A. Prosperetti, "Linear pressure waves in bubbly liquids—Comparison between theory and experiments," J. Acoust. Soc. Am. **85**, 732–746 (1989).

[32] L. L. Foldy, "The multiple scattering of waves," Phys. Rev. **67**, 107–109 (1945).

[33] T. Maxworthy, C. Gnann, M. Kurten, and F. Durst, "Experiments on the rise of air bubbles in clean viscous liquids," J. Fluid Mech. **321**, 421–441 (1996).

[34] T. R. Hahn, "Low-frequency sound scattering from spherical assemblages of bubbles using effective medium theory," J. Acoust. Soc. Am. **122**, 3252–3267 (2007).

[35] W. W. L. Au, A. Frankel, D. A. Helweg, and D. H. Cato, "Against the humpback whale sonar hypothesis," IEEE J. Ocean. Eng. **26**, 295–300 (2001).

[36] L. N. Frazer and E. Mercado, "A sonar model for humpback whale song," IEEE J. Ocean. Eng. **25**, 160–182 (2000).

[37] E. Mercado and L. N. Frazer, "Humpback whale song or humpback whale sonar? A reply to Au *et al.*," IEEE J. Ocean. Eng. **26**, 406–415 (2001).

[38] T. B. Borak, "A method for computing random chord length distributions in geometrical objects," Radiat. Res. **137**, 346–351 (1994).

[39] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1965), pp. xiv and 1046.

[40] O. Diachok, "Absorption spectroscopy: A new approach to estimation of biomass," Fish. Res. **47**, 231–244 (2000).

[41] R. H. Love, "Resonant acoustic scattering by swimbladder-bearing fish," J. Acoust. Soc. Am. **64**, 571–580 (1978).

[42] F. A. Sharpe and L. M. Dill, "The behavior of Pacific herring schools in response to artificial humpback whale bubbles," Can. J. Zool. **75**, 725–730 (1997).

[43] T. G. Leighton, D. Finfer, E. Grover, and P. R. White, "An acoustical hypothesis for the spiral bubble nets of humpback whales and the implications for whale feeding," Acoust. Bull. **32**, 17–21 (2007).

[44] T. G. Leighton, S. D. Richards, and P. R. White, "Trapped within a 'wall of sound': A possible mechanism for the bubble nets of humpback whales," Acoust. Bull. **29**, 24–29 (2004).

[45] D. Gendron and J. Urban, "Evidence of feeding by humpback whales (*Megaptera-novaeangliae*) in the Baja-California breeding ground, Mexico," Marine Mammal Sci. **9**, 76–81 (1993).

[46] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[47] I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, Yu. V. Geronimus, and M. Yu. Tseytlin, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1965), pp. xlv and 1160.

Three-dimensional source tracking in an uncertain environment

Dag Tollefsen
Norwegian Defence Research Establishment (FFI), Box 115, 3191 Horten, Norway

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

(Received 5 January 2009; revised 17 February 2009; accepted 17 February 2009)

This paper develops an approach to three-dimensional source tracking in an uncertain ocean environment using a horizontal line array (HLA). The tracking algorithm combines matched-field focalization for environmental (seabed and water column) and source-bearing model parameters with the Viterbi algorithm for range-depth estimation and includes physical constraints on source velocity. The ability to track a source despite environmental uncertainty is examined using synthetic test cases for various track geometries and with varying degrees of prior information for environmental parameters. Performance is evaluated for a range of signal-to-noise ratios in terms of the probability of estimating a track within acceptable position/depth errors. The algorithm substantially outperforms tracking with poor environmental estimates and generally obtains results close to those obtained with exact environmental knowledge. The approach is also applied to measured narrowband data recorded on a bottom-moored HLA in shallow water (the Barents Sea) and shown to successfully track both a towed submerged source and a surface ship in cases where simpler tracking algorithms failed.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097495]

PACS number(s): 43.30.Wi, 43.30.Pc [AIT] Pages: 2909–2917

I. INTRODUCTION

Localizing and/or tracking an unknown acoustic source in the ocean is an important and challenging problem that has received much attention.[1–13] Matched-field processing (MFP)[1–3] is a widely-used source localization technique based on matching the acoustic field measured at an array of sensors with replica fields computed by a numerical propagation model over a grid of trial source positions. The grid of matches (or mismatches) constitutes an ambiguity surface, and the source position estimate is taken to be the position of maximum match (or minimum mismatch) on this surface. For a moving source, additional properties of source speed and course can be derived; several matched-field tracking methods[4–6] have been developed to estimate source-track parameters.

MFP requires knowledge of the acoustic environment (including water-column sound-speed profile (SSP) and seabed geoacoustic parameters), and environmental model mismatch[7,8] can pose a severe limitation for localization. One approach to reducing environmental model mismatch is to carry out a geoacoustic inversion survey using controlled sources at known positions to estimate seabed parameters, e.g., by matched-field inversion,[9] and then use the estimated (fixed) seabed model in subsequent matched-field localization of unknown sources.[10] An alternative approach that does not require a preliminary survey is to treat environmental parameters and source positions as joint unknowns and seek the minimum-mismatch solution over all unknowns using a numerical optimization algorithm. In its original formulation, the method of focalization[11] searched for environmental model parameters and a single source position through a series of parameter and coordinate perturbations driven by the global search method of simulated annealing. This approach was designed for improved localization without necessarily obtaining the correct environmental parameters due to the non-uniqueness of the acoustic inverse problem.[11] (Subsequent formulations[12] used environmental parameter perturbations with source coordinates searched exhaustively over ambiguity surfaces.) Focalization has recently been extended to two-dimensional tracking of a moving source,[13] with an efficient optimization-tracking approach developed and applied to synthetic data on a vertical line array.

This paper develops an approach for three-dimensional (3D) tracking in an uncertain shallow-water environment with application to horizontal line array (HLA) data (Sec. II). The method makes use of data from multiple observations (source positions) along the track, an efficient hybrid optimization algorithm, and the Viterbi algorithm[14] to determine the optimal source track within applied limits on source horizontal and vertical velocities. The approach is applied to noisy synthetic data in a series of test cases that include different track geometries, varying signal-to-noise ratios (SNRs), and varying levels of prior information on seabed and water-column parameters (Sec. III). The results are evaluated in terms of the probability of estimating a track that is acceptably close to the true track. The tracking-optimization algorithm is found to substantially outperform tracking with poor environmental estimates and in general obtains results close to those obtained with exact environmental knowledge. The method is also applied to data collected with a bottom-moored HLA in shallow waters of the Barents Sea, including data due to a continuous-wave towed source and ship noise at several ranges (Sec. IV). Finally, Sec. V summarizes and discusses this work.

II. TRACKING ALGORITHM

In this section, the 3D optimization-tracking algorithm is described. Several modifications to the original focalization algorithm[11] have been implemented. First, focalization is extended to use of multiple data observations for a moving source. This increases the data information applied but also leads to a more challenging optimization problem since more unknowns (i.e., multiple source positions) are introduced. Second (taking on the approach of Ref. 12), source range and depth coordinates are computed from the mismatch ambiguity surfaces generated for the environmental and bearing parameters of each model in the optimization. Third, an efficient tracking algorithm is applied to find the optimal source track through the set of ambiguity surfaces. This approach is similar to the two-dimensional optimization-tracking algorithm recently developed by Dosso and Wilmut,[13] but is extended here to 3D tracking with a HLA.

Consider acoustic data $\mathbf{d}=\{\mathbf{d}_{fj}, f=1,F; j=1,J\}$ at N sensors, F frequencies, and J observations (data segments) for a moving source. The data mismatch (often referred to as energy) is taken to be the (negative) log-likelihood function under the assumption of uncorrelated, complex Gaussian-distributed errors with unknown source amplitude and phase and unknown error variance at each frequency. This leads to energy E_j for the jth data segment given by[13]

$$E_j(\mathbf{m}) = N\sum_{f=1}^{F} \log_e B_{fj}(\mathbf{m}), \qquad (1)$$

and a total track energy of

$$E(\mathbf{m}) = \sum_{j=1}^{J} E_j(\mathbf{m}), \qquad (2)$$

where $B_{fj}(\mathbf{m})$ is the Bartlett mismatch defined by

$$B_{fj}(\mathbf{m}) = \mathrm{Tr}\{\mathbf{C}_{fj}\} - \frac{\mathbf{d}_{fj}^{\dagger}(\mathbf{m})\mathbf{C}_{fj}\mathbf{d}_{fj}(\mathbf{m})}{|\mathbf{d}_{fj}(\mathbf{m})|^2}. \qquad (3)$$

Here, $\mathrm{Tr}\{\bullet\}$ represents the matrix trace, † represents conjugate transpose, $\mathbf{d}_{fj}(\mathbf{m})$ is the replica acoustic field computed for environmental and track model $\mathbf{m}$, and $\mathbf{C}_{fj}$ is the data cross-spectral density matrix (CSDM) at the fth frequency and the jth data segment.

The search for an optimal (lowest-energy) model is driven by the adaptive simplex simulated annealing (ASSA) hybrid search algorithm,[15] which combines the global search method of very fast simulated annealing with the local downhill simplex (DHS) method. The DHS method[16] operates on a simplex of models and repeatedly applies geometric operations (reflection, expansion, and contraction) to the highest-energy model of the simplex. In ASSA, DHS operations are followed by a random perturbation of all parameters, and the resulting model is conditionally accepted based on the Metropolis criterion, i.e., if a random number ξ drawn from a uniform distribution on $[0,1]$ satisfies

$$\xi \leq e^{-\Delta E/T}, \qquad (4)$$

where ΔE is the energy difference from the original model and T is a control parameter (temperature). After a required number of accepted perturbations, the temperature is reduced according to $T_{k+1}=\beta T_k$ with $\beta<1$ to decrease the probability of accepting a higher-energy model. This procedure is repeated until convergence, defined to be when the difference between the highest and lowest energies in the simplex (relative to their average) is less than a pre-defined threshold. The ASSA algorithm employs several techniques for increased efficiency, including adaptive adjustment of the perturbation size for each parameter (based on a running average of recently-accepted perturbations) and drawing the parameter perturbations from Cauchy distributions.

The variation in source bearing (θ) with time is included in the inversion and modeled by a second-order polynomial

$$\theta_j = b_0 + b_1 t_j + b_2 t_j^2, \qquad (5)$$

where $j=1, J$ is the data segment index, $t_j=(j-1)\Delta t$ is the observation time (Δt is a constant time increment), and (b_0, b_1, b_2) are unknown polynomial coefficients. The coefficients are included as parameters in the model $\mathbf{m}$, with b_2 constrained to have the same sign as b_1 to avoid looping tracks. The use of higher-order polynomials could allow for closer fit to a larger number of track types; however, each polynomial coefficient introduces an additional parameter in the optimization and may complicate the search for simple-shaped (e.g., linear) tracks.

A key aspect of the tracking algorithm is that source range and depth along the track are not included as explicit parameters in the ASSA optimization, but are treated implicitly. For each model $\mathbf{m}$ considered in the optimization, the energy function [Eq. (1)] is evaluated for each data segment at the range-depth grid of source positions. The Viterbi algorithm is then applied to determine the lowest-energy (range-depth) track through this series of ambiguity surfaces, subject to constraints on maximum horizontal and vertical source velocities (of v_H and v_z, respectively). The Viterbi algorithm progresses as follows: for each grid point on the second surface, the sum of the energy at that point and the point of minimum energy on the first surface within a region limited to $\pm\sqrt{(v_H\Delta t)^2-(r\Delta\theta)^2}$ in range and $\pm v_z\Delta t$ in depth of the current position is computed and stored (where $\Delta\theta$ is the difference in bearing and r is the grid-point range). The point on the first surface that produces the smallest energy sum is taken to be the antecedent to the grid point on the second surface; this is also stored. For each grid point of the third surface, the sum of the energy at that point and the point of minimum energy on the stored energy-sum surface within source velocity constraints is computed and stored. The point on the second surface that produces the smallest energy sum is taken to be the antecedent to the grid point on the third surface. This procedure is continued until the Jth surface has been examined. The minimum of the final energy-sum surface then defines the end point of the optimal track, and the optimal track through all surfaces (satisfying source velocity constraints) is determined since antecedents on all previous surfaces have been stored.

TABLE I. Model parameters and search bounds used in simulation study.

Parameter and units	True value	Wide bounds	Narrow bounds
Geoacoustic			
h (m)	12	[0, 40]	[11, 13]
c_{1T} (m/s)	1503	[1450, 1600]	[1494, 1512]
c_{1B} (m/s)	1560	[1500, 1650]	[1547, 1573]
c_2 (m/s)	1750	[1600, 1900]	[1670, 1830]
ρ_1 (g/cm^3)	1.50	[1.20, 2.00]	[1.36, 1.64]
ρ_2 (g/cm^3)	1.85	[1.40, 2.20]	[1.48, 1.82]
α_1 (dB/λ)	0.22	[0.01, 1.00]	[0.14, 0.30]
α_2 (dB/λ)	0.12	[0.01, 1.00]	[0.01, 1.00]
D (m)	115	[113, 117]	[114.5, 115.5]
SSP			
c_{w1} (m/s) at 0 m	1472	[1465, 1480]	[1469, 1475]
c_{w2} (m/s) at D	1468	[1465, 1480]	[1465, 1471]
Bearing			
b_0 (deg)		[−180, 180]	
b_1 (deg/min)		[−9, 9]	
b_2 (deg/min^2)		[0, 0.5]	

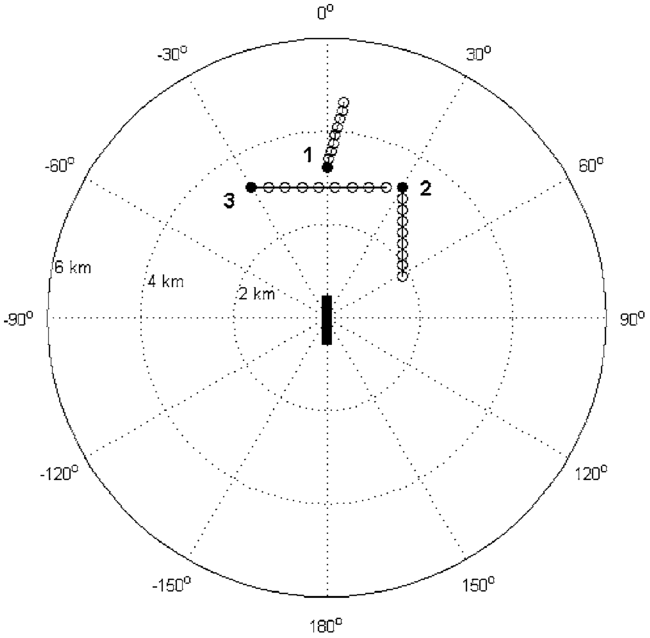


FIG. 1. Tracks (1–3) used in simulation study. Start positions indicated with filled circles. Array (not to scale) is centered at the origin of the coordinate system.

III. NUMERICAL SIMULATIONS

A. Test cases

An environmental model representative of the shallow continental shelf[17] was chosen for the simulation study. The model consists of a water column of depth D and a seabed with a sediment layer of thickness h over a semi-infinite basement. Seabed geoacoustic parameters include sound speed at the top and bottom of the sediment layer (c_{1T} and c_{1B}), density (ρ_1) and attenuation (α_1) in the sediment, and sound speed (c_2), density (ρ_2), and attenuation (α_2) in the basement. The water-column SSP is described by two parameters (c_{w1} and c_{w2}) at depths of 0 and D m. Table I lists the true values for all environmental parameters together with both wide and narrow search bounds applied in this study. (The narrow bounds represent relatively small uncertainties such as would typically result from carrying out a geoacoustic inversion survey and water-column SSP measurements before tracking. The wide bounds represent an unknown environment with no such information available.) In the first part of this study, wide search bounds are applied for the eight seabed parameters and water depth, and narrow search bounds for the SSP parameters.

The test cases involve acoustic data at frequencies of 200 and 300 Hz received at a 256-m long HLA on the seafloor (oriented north-south). The array is comprised of 33 sensors equidistant spaced at 8-m intervals. This relatively sparse array (element spacing greater than one acoustic wavelength) was chosen to keep computational efforts reasonable given the large number of inversions considered in this study. Simulated data were generated for a source moving at 24-m depth at constant speed and course for three tracks, each consisting of nine data segments separated by 60 s in time. The tracks are plotted in Fig. 1 and include a source moving outbound at a speed of 3 m/s near array endfire at range 3.23–4.63 km (track 1), a source moving at 4 m/s parallel to and toward the array broadside at range 3.23–1.84 km (track 2), and a source moving across array endfire at 6 m/s at range 3.23–2.80 km (track 3).

The range-depth search grid defining the two-dimensional ambiguity surfaces cover 0.3–6.0 km in range (grid spacing 50 m) and 4–112 m in depth (grid spacing 2 m). (Note that the grid points do not coincide exactly with track points in range.) Table I lists search limits for the bearing model parameters. Track constraints are imposed on the source horizontal and vertical velocities of 9 and 0.067 m/s, respectively.

To simulate noisy data, the CSDM is computed using synthetic acoustic fields (for the true environment) with noise added to the acoustic pressure vectors, i.e.,

$$\mathbf{C}_{fj} = (\mathbf{d}_{fj} + \sigma_f \mathbf{n}_{fj})(\mathbf{d}_{fj} + \sigma_f \mathbf{n}_{fj})^\dagger, \tag{6}$$

where $\mathbf{n}_{fj}$ is a Gaussian-distributed complex noise vector and σ_f^2 is the noise variance (a single data snapshot is used). Under the assumption of independent Gaussian errors, the noise variance corresponding to a specific SNR can be computed as

$$\sigma_f^2 = 10^{-\mathrm{SNR}_f/10}/N. \tag{7}$$

The SNR was set to a fixed value at the start of the track and adjusted with range to yield constant noise variance. The resulting variations in SNR over the tracks are displayed in Fig. 2.

The normal-mode numerical propagation model ORCA (Ref. 18) was used to compute synthetic data and replica acoustic pressure fields. The number of acoustic field computations required for each model $\mathbf{m}$ is $M \cdot N \cdot J$, where M is the number of range-depth grid points and the factor J (number of data segments) is required due to variation in grid to array element ranges with bearing (computed using the law

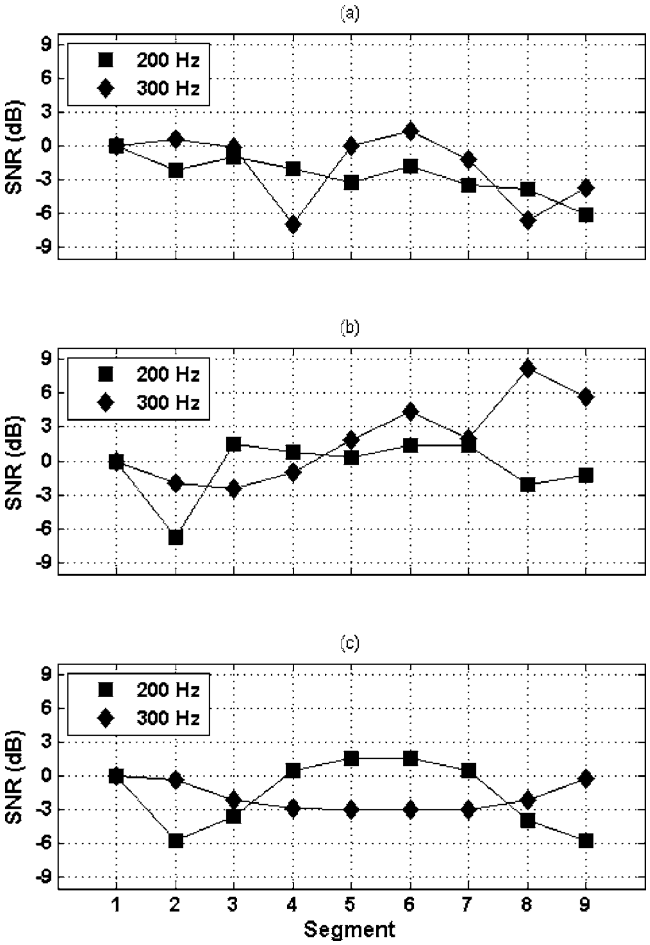


FIG. 2. SNR vs track segment, relative to 0 dB at start of track, at 200 Hz (boxes) and 300 Hz (diamonds) for (a) track 1, (b) track 2, and (c) track 3.

of cosines). For the 33-element HLA, given ambiguity surface grid size, and nine data segments, 1.87×10^6 field computations are required per frequency.

A preliminary study indicated that suitable algorithm control parameters for efficient track estimation consist of a temperature reduction factor $\beta=0.95$ with nine accepted perturbations required per temperature step and a convergence threshold of 0.001. Furthermore, the algorithm is terminated after six consecutive reductions in energy with unchanged track range and depth parameters.

A track estimate is considered to be acceptable if the mean position error is less than 600 m and the mean absolute depth error is less than 6 m. An error in track orientation with respect to the array length axis is not considered unacceptable since this can result from the inherent inability of a HLA for left-right discrimination, and not a failure due to the algorithm (i.e., if the track estimate reflected about the array axis satisfies the mean position and depth-error criteria, the result is considered acceptable).

B. Tracking results

To examine the ability to track sources in an uncertain environment with different levels of noise, the tracking-optimization algorithm was applied to 20 realizations of noisy data at SNR values from −6 to +6 dB (at the start of track). The results are compared with those obtained using the same 20 realizations of noisy data and either the *exact* model or a *random* environment. For the exact case, the environmental parameters were fixed to their exact values and the bearing model parameters to the least-squares fit to the exact bearings (this represents the closest approximation to the true bearing for the given bearing parametrization.) For the random case, the environmental parameters were fixed to random values drawn uniformly from within their search bounds. Results for the three tracks are summarized in Figs. 3–5 in terms of the probability of an acceptable track (PAT) vs average SNR (averaged in decibels over all track segments and frequencies), with one standard deviation error

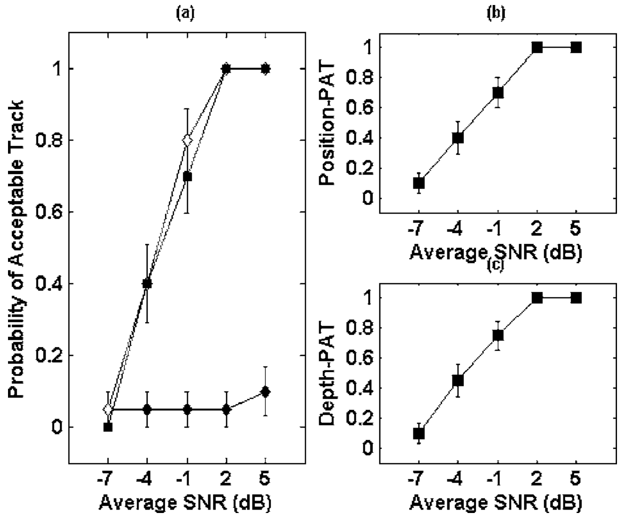


FIG. 3. (a) PAT vs average SNR for track 1 for focalization in uncertain environment (squares), random environment (closed diamonds), and exact environment (open diamonds). Vertical lines indicate one standard deviation error bars. (b) PAT for position-error and (c) PAT for depth-error only for focalization in uncertain environment.

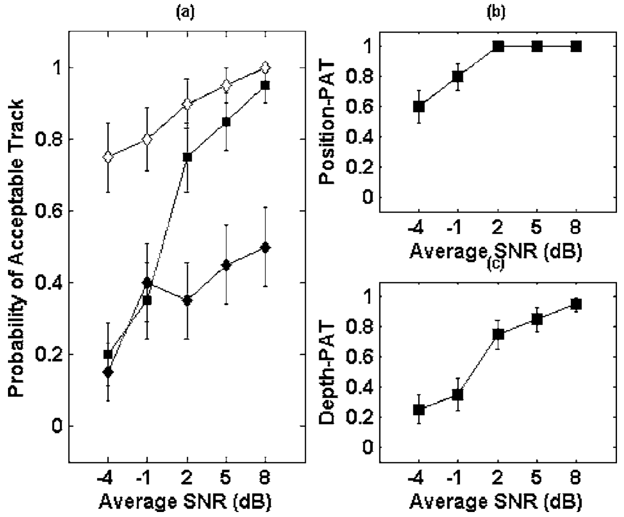


FIG. 4. (a) PAT vs average SNR for track 2 for focalization in uncertain environment (squares), random environment (closed diamonds), and exact environment (open diamonds). Vertical lines indicate one standard deviation error bars. (b) PAT for position-error and (c) PAT for depth-error only for focalization in uncertain environment.

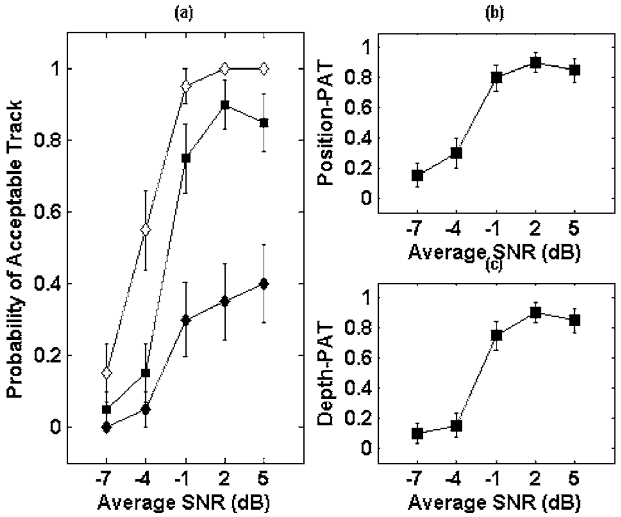


FIG. 5. (a) PAT vs average SNR for track 3 for focalization in uncertain environment (squares), random environment (closed diamonds), and exact environment (open diamonds). Vertical lines indicate one standard deviation error bars. (b) PAT for position-error and (c) PAT for depth-error only for focalization in uncertain environment.

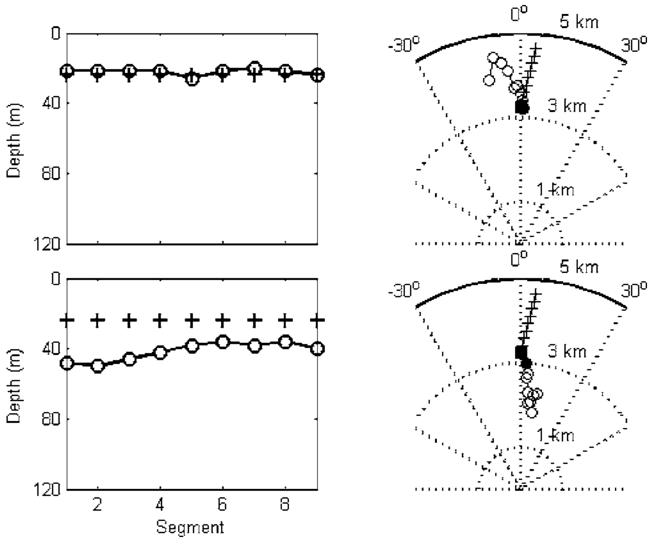


FIG. 6. Examples of acceptable (upper panels) and unacceptable (lower panels) track estimates for track 1 at −1-dB average SNR. Symbols indicate true (+) and estimated (○) coordinates. Start points indicated with filled symbols.

estimates computed for a binomial distribution according to[19]

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{8}$$

where $\hat{p}$ is the number of acceptable track estimates in n trials.

Figures 3–5 show that PAT generally increases with increasing SNR for all tracking approaches. The PAT values obtained by optimization-tracking are generally much higher than those obtained for a random environment (except at the lowest SNRs) and in some cases approach the PAT for the exact model.

Figure 3 shows that for track 1, the PAT for focalization is close to the exact-model PAT at all SNRs and reaches 1.0 at +2-dB SNR. The good performance for this track can be understood since the ability of a HLA for estimation of environmental parameters and range-depth localization is highest for a source at endfire, where the effective vertical aperture of the array is maximized.[3] (Bearing resolution for a HLA is poorest at endfire; this can contribute to an increase in position error, but appears to be a less important effect in this case.) To further investigate the results, Figs. 3(b) and 3(c) show the PAT for position error only and for depth error only, which are quite similar for this track. Note that a track can be acceptable in position but not depth (or vice versa); for example, at −7-dB SNR the position-error and depth-error PAT values are both 0.1 while the position-and-depth PAT value is 0.

Figure 4 shows that for track 2, the PAT for focalization at +2-dB SNR and above is close to that obtained for the exact model. At −4- and −1-dB SNRs, the PAT for focalization is significantly lower than for the exact model. As the source moves toward array broadside, the effective vertical aperture of the array diminishes, and matched-field range and depth localization degrades. This loss of localization in range can in part be offset by the ability for ranging by sampling wavefront curvature; this effect is most prominent for a source close to array broadside and at close source-receiver ranges. In addition, the bearing resolution for a HLA in general improves as the source moves toward array broadside. Figures 4(b) and 4(c) show that the position-error PAT is higher than the depth-error PAT at all SNRs, indicating that depth error is a more important cause of degraded track estimation for this track. Figure 5 shows that for track 3, the PAT for focalization increases with SNR but does not reach that obtained for the exact model. Overall, Figs. 3–5 show that the best focalization results are obtained for the two tracks near array endfire, with better results for track 1 than for track 3. Results are overall slightly poorer for the track that approaches the array broadside (track 2).

To further illustrate results of the tracking algorithm, typical examples of acceptable and unacceptable tracks obtained via focalization are presented in Figs. 6–8. In Fig. 6 (track 1 at −1-dB average SNR), the acceptable track has excellent depth estimates and is roughly linear except for the track end point, but is headed in a north-west (rather than north-east) direction; this is an example of the left-right ambiguity about the HLA. The deviation at the end of the track is related to the fact that end points are subject to one-sided track constraints only. The unacceptable track is at wrong depth and does not predict outward motion; however, the bearings are approximately correct. In Fig. 7 (track 2 at +2-dB SNR), the acceptable track closely matches the true track except for an end point deviation. The unacceptable track has large depth errors and is unacceptable in position, although range and bearing errors are small. In Fig. 8 (track 3 at +2-dB SNR), the acceptable track closely matches the true track. The unacceptable track is approximately 20 m deeper and at approximately 600-m shorter range than the true track (both position and depth errors above the threshold), but the bearing estimates are good and the track direction is correct.

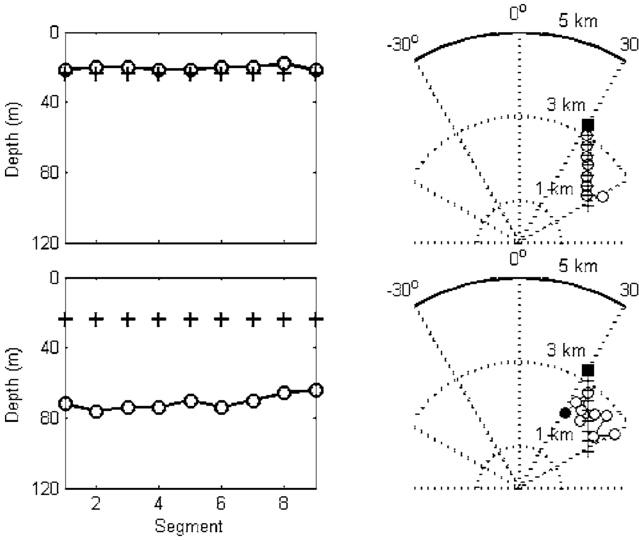


FIG. 7. Examples of acceptable (upper panels) and unacceptable (lower panels) track estimates for track 2 at +2-dB average SNR. Symbols indicate true (+) and estimated (○) coordinates. Start points indicated with filled symbols.

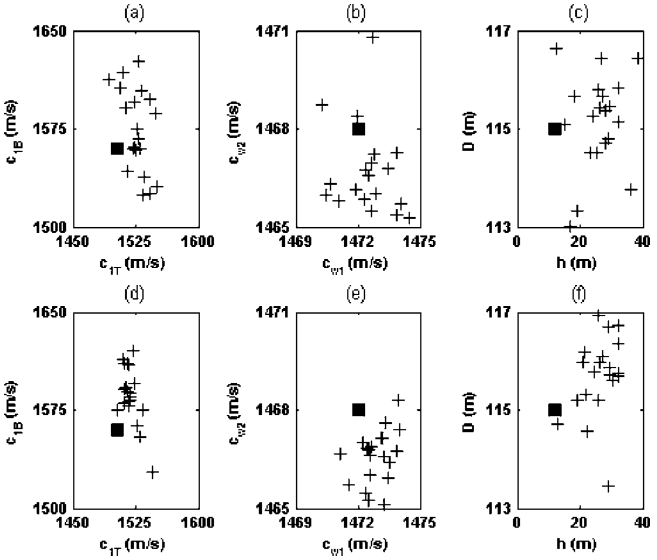


FIG. 9. Scatter plots of optimal model values for selected geoacoustic parameters. Track 1 at (a)–(c) −1-dB and (d)–(f) +5-dB average SNRs. True model values indicated with boxes.

The primary goal of focalization is localization/tracking, not recovery of environmental parameters. However, it is interesting to investigate to what extent the true environmental parameters can be determined given data information from multiple source positions along a track. (For example, Ref. 20 considered to what extent geoacoustic parameters could be recovered using data from an unknown moving source, with applications to geoacoustic inversion of noise from a ship of unknown position.) Figure 9 shows scatter plots of the optimal parameter values from each of the 20 runs of the optimization-tracking algorithm for combinations of six of the most sensitive environmental parameters for track 1 at −1- and +5-dB average SNRs (results for both acceptable and unacceptable tracks are included). In general, parameter estimates appear widely scattered within their search bounds with values that differ significantly from the true values, although some observations can be made. The spread of estimated values is larger at −1 dB than +5 dB. Estimates of sediment sound speed (c_{1T}, c_{1B}) tend to be aligned with negative slope; this correlation suggests that the data are sensitive to the average sound speed in sediment, but have limited ability to resolve these parameters individually. At +5 dB, the estimates of sound speed in water (c_{w1}, c_{w2}) tend to be aligned with positive slope; this suggests that the data are more sensitive to the shape of the SSP than to the average sound speed. Similar inter-parameter correlations have been observed previously in geoacoustic inversion.[9,15,20] Estimates of sediment thickness (h) and water depth (D) tend to be higher than their true values. Overall, it appears that the environmental parameters are not well determined individually but provide effective environments that allow for correct track estimation.

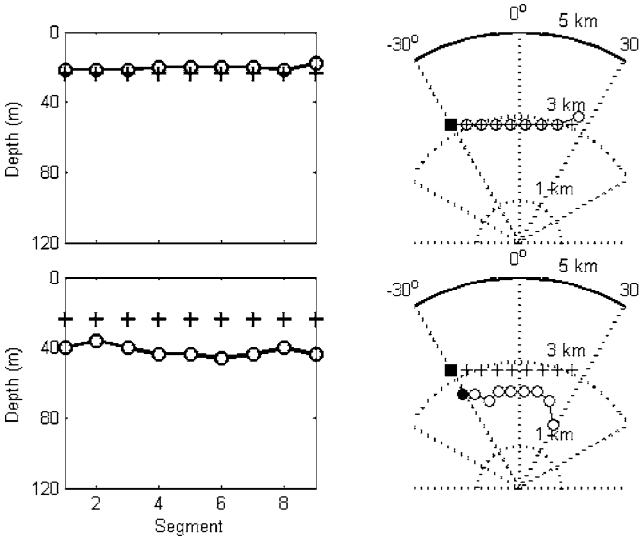


FIG. 8. Examples of acceptable (upper panels) and unacceptable (lower panels) track estimates for track 3 at −1-dB average SNR. Symbols indicate true (+) and estimated (○) coordinates. Start points indicated with filled symbols.

C. Effects of prior environmental uncertainty

This section examines and quantifies the dependence of track estimation on prior information of environmental parameters (e.g., whether the parameters are essentially completely unknown, or whether previous measurements, with uncertainties, exist). The tracking examples so far have used wide search bounds on geoacoustic parameters and water depth, and narrow search bounds on water sound speed as given in Table I. This section also considers narrow or wide bounds on geoacoustic parameters and water depth in combination with narrow or wide bounds on water sound speed (Table I). The narrow geoacoustic bounds correspond to the 95%-highest probability density credibility intervals obtained in a simulated matched-field geoacoustic inversion experiment in this environment.[17] Only tracks 1 (endfire) and 2 (toward array broadside) are considered here.

Figure 10(a) shows that for track 1 the results obtained for wide geoacoustic/wide SSP bounds are consistently the poorest (except at the lowest SNR). Generally similar results

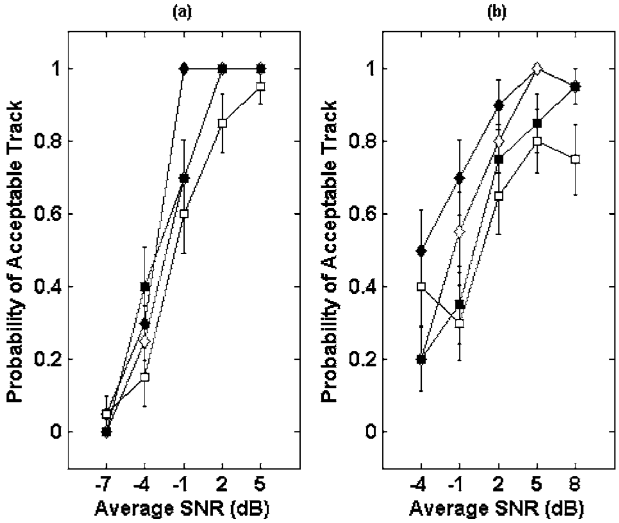


FIG. 10. PAT vs average SNR for (a) track 1 and (b) track 2 for varying search bounds on geoacoustic/SSP parameters: narrow/narrow (closed diamonds), wide/narrow (closed boxes), narrow/wide (open diamonds), and wide/wide (open boxes). Vertical lines indicate one standard deviation error bars.

are obtained for the other three cases (except at −1-dB SNR, where the narrow geoacoustic/narrow SSP result is significantly better). Note that narrow geoacoustic bounds allow for wide SSP bounds without significant degradation of results obtained for narrow geoacoustic/narrow SSP bounds. Figure 10(b) shows that for track 2, the results obtained for narrow geoacoustic/narrow SSP bounds are significantly better than those obtained for wide geoacoustic/narrow SSP bounds. Results for narrow geoacoustic/wide SSP bounds are generally better than results for wide geoacoustic/narrow SSP bounds. Results for wide geoacoustic/wide SSP bounds are poorest.

These results indicate that improved prior information on environmental parameters can have an important effect on tracking performance, particularly for tracks extending toward HLA broadside, where the data have reduced ability to resolve environmental parameters. The results also indicate that (for this environment) use of narrow bounds for geoacoustic parameters allows for wide bounds on SSP parameters without significant degradation of tracking performance.

IV. EXPERIMENTAL RESULTS

A. Shallow-water experiment

An acoustic experiment to study geoacoustic inversion and acoustic localization/tracking was conducted by the Norwegian Defense Research Establishment (FFI) in the shallow waters of the Barents Sea in 2003.[21,22] Acoustic data were collected with an 18-element HLA of length 900 m deployed on the relatively flat seabed at a depth of approximately 282 m. The array consisted of 7 sensors spaced at 10-m intervals followed by 11 sensors at increasing spacing to a maximum of 160 m. An acoustic source was towed at depth of 54 m and speed of 5.2 kn by the R/V H U SVERDRUP II along an outward radial track oriented at 30° angle with respect to the array endfire. The acoustic data considered here were due to a continuous-wave tone emitted by the acoustic source at a frequency of 80 Hz for source-to-array (closest end) ranges of 4.0–4.7 and 7.8–8.5 km (referred to as short-range and long-range towed-source data, respectively). In addition, ship-noise at a frequency component of 144 Hz from ship-to-array ranges of 1.6–2.3 and 5.1–5.8 km (referred to as short-range and long-range ship noise, respectively) is also considered.

TABLE II. Geoacoustic model parameters, estimated values from matched-field inversion in Ref. 21, and search bounds used with experimental data.

Parameter and units	Estimated value	Search bounds
h (m)	11.2	[1, 40]
c_1 (m/s)	1510	[1450, 1900]
c_2 (m/s)	1753	[c_1, c_1+30h]
ρ_1 (g/cm^3)	2.03	[1.4, 3.0]
ρ_2 (g/cm^3)	2.06	[1.4, 3.0]
α_1 (dB/m kHz)	0.32	[0.01, 1.0]
α_2 (dB/m kHz)	0.21	[0.01, 1.0]
D (m)	282	[278, 288]

The processing sequences for the towed-source (ship-noise) data consisted of forming CSDM estimates from five (ten) consecutive 50% overlapping data snapshots of length 6.6 s (3.3 s), converted to the frequency domain using a fast Fourier transform with a frequency bin width of 0.15 Hz (0.3 Hz), with a Hamming time windowing function applied. The total averaging time for each data segment was 19.8 s (18.2 s). Nine data segments, separated by 33 s between the start of each segment, are used in tracking; these extended over a time span of 4 min 42 s over which the ship moved approximately 755 m. The average SNR for the towed-source data was estimated to be 3.3 dB (short-range) and 0.2 dB (long-range), and for the ship noise 4.5 dB (short-range) and −0.5 dB (long-range). (The estimates are based on an average of signal and noise power spectral densities, after snapshot-averaging, over the elements of the array[21] and over the data segments.)

The model environment consists of a water column with a known SSP over a two-layer seabed. The seabed model[21] consists of an upper layer with depth-dependent properties over a homogeneous basement half-space. The parameters that describe the upper model layer are the layer thickness (h), sound speed at the top and bottom of the layer (c_1 and c_2), attenuation coefficient at the top and bottom of the layer (α_1 and α_2), and a depth-independent density ρ_1. The lower layer is described by constant sound-speed and attenuation values that are identical to those at the base of the upper layer (c_2 and α_2) and by an independent density ρ_2. The seven geoacoustic model parameters used to describe the seabed, their estimated values (mean values from Bayesian matched-field inversion of multi-tone towed-source data at a range of 1.58 km),[21] and the search bounds employed for tracking are listed in Table II. The range-depth ambiguity surfaces covered 8 km in range (1–9 km for short-range and 4–12 km for long-range data, grid spacing 50 m) and almost the entire water column in depth (6–270 m, grid spacing 2 m). A second-order polynomial was used to model track bearing, with search bounds for bearing parameters set as

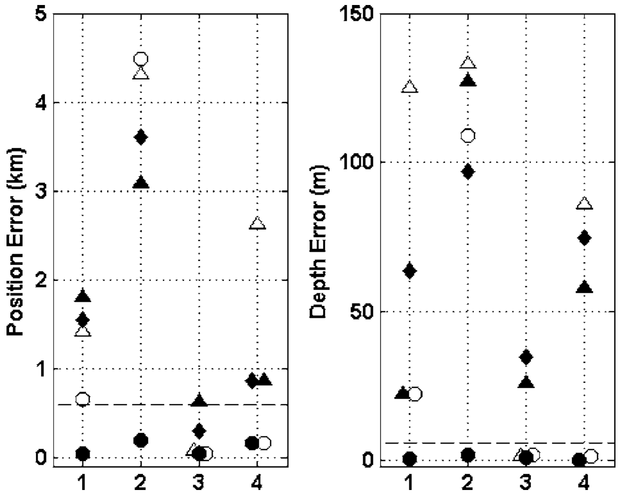


FIG. 11. 3D tracking results for Barents Sea data: (1) towed-source data at range 4.0–4.7 km, (2) towed-source data at range 7.8–8.5 km, (3) ship noise at range 1.6–2.3 km, and (4) ship noise at range 5.1–5.8 km. Algorithms applied (described in the text) are as follows: 1pp—open triangles, 9pp—closed triangles, 9pv—open circles, 9vv—closed circles, and 9rv—diamonds. Dashed lines indicate 600-m mean position error and 6-m mean depth error, defining an acceptable track.

listed in Table I. Track constraints were imposed on source horizontal and vertical velocities of 6 and 0.067 m/s, respectively.

B. Tracking results

To assess the 3D optimization-tracking algorithm developed in Sec. II for source tracking in an uncertain environment, four alternative (simpler) approaches to track estimation are also applied here to the measured data. In *independent focalization* (referred to as case 1pp) focalization is applied independently for each data segment, with no model applied for track bearing. The track is constructed as the sequence formed by the optimal source range, depth, and bearing coordinates from each focalization, with no constraints applied on source velocity. In *simultaneous focalization*, data from all segments are inverted simultaneously, with the polynomial model [Eq. (5)] applied for track bearing. For each model evaluated in the optimization, the minimum of each ambiguity surface is taken as the optimal range-depth coordinates. The final track is constructed as the sequence of optimal range-depth minima, with no constraints applied on source velocity (case 9pp), or by application of the Viterbi algorithm to the final set of ambiguity surfaces with constraints applied on source velocity (case 9pv). These approaches are compared to results for the 3D optimization-tracking algorithm described in Sec. II (case 9vv) and for a *random environment* (case 9rv) in which the 3D optimization-tracking algorithm is applied with environmental parameters fixed to values drawn randomly from the uniform search bounds (results represent averages over 100 runs with different environment realizations).

Tracking results are summarized (Fig. 11) in terms of mean position error and mean depth error for the five approaches. Errors are measured with respect to the source position for the towed-source data or ship stern position for ship noise and with respect to the source tow depth or ship propeller depth. Figure 11 shows that the optimization-tracking algorithm provides acceptable track estimates for all cases. For the towed-source data, only the 3D optimization-tracking algorithm (9vv) estimates an acceptable source track, with a mean position error of 36 m at short range and 200 m at long range (mean depth error <2 m in both cases); all other algorithms provide large range and depth errors. For ship noise, the 3D optimization-tracking algorithm estimates an acceptable track with mean position errors less than 100 m at short range and 200 m at long range (mean depth error <2 m in both cases). At short range, 1pp and 9pv also estimate tracks with small position and depth errors. At long range, 9pv also estimates an acceptable track, but the other algorithms do not. These results indicate that simultaneous environmental optimization and tracking with source velocity constraints are required for reliable source-track estimation for this experiment.

V. SUMMARY

This paper developed an approach to 3D source tracking with a HLA in an unknown ocean environment. The approach is based on ASSA optimization for unknown environmental and bearing model parameters and applies the Viterbi algorithm (with source velocity constraints) for range-depth track estimation. The algorithm was applied to synthetic data in a simulated environment for several linear source-track test cases. In general, the algorithm performance increased with SNR and in many cases approached the results obtained with exact environmental knowledge. The algorithm consistently outperformed tracking with an environmental model selected at random from the prior uncertainty bounds. Best tracking results were obtained for tracks near array endfire, with some degradation for a track that approached the array broadside. Improved prior information on the environmental parameters, in the form of narrowed parameter search bounds, was shown to improve tracking performance, with the largest impact for a track extending toward HLA broadside, and best results achieved with narrow bounds on both seabed geoacoustic and water sound-speed parameters. Results also indicated that, for the environment studied, use of narrow bounds for geoacoustic parameters allows for wide bounds on SSP parameters without significant degradation of tracking performance. Finally, the algorithm was applied to relatively low-SNR narrowband data from a towed submerged source and noise from a surface ship recorded on a bottom-moored HLA in a shallow-water environment. The 3D source tracking algorithm successfully estimated the tracks of the source and ship and substantially outperformed alternative simpler algorithms that either did not make use of multiple data segments simultaneously, did not optimize over environmental parameters, or did not apply source velocity constraints via the Viterbi algorithm.

[1] J. M. Ozard, "Matched field processing in shallow water for range, depth, and bearing determination: Results of experiment and simulation," J. Acoust. Soc. Am. **86**, 744–753 (1989).

[2] 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).

[3] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World

Scientific, Singapore, 1993).

[4] H. Bucker, "Matched-field tracking in shallow water," J. Acoust. Soc. Am. **96**, 3809–3811 (1994).

[5] M. J. Wilmut, J. M. Ozard, and P. Brouwer, "Evaluation of two efficient target tracking algorithms for matched-field processing with horizontal arrays," J. Comput. Acoust. **3**, 311–326 (1995).

[6] L. T. Fialkowski, J. S. Perkins, M. D. Collins, M. Nicholas, J. A. Fawcett, and W. A. Kuperman, "Matched-field source tracking and ambiguity surface averaging," J. Acoust. Soc. Am. **110**, 739–746 (2001).

[7] 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).

[8] A. Tolstoy, "Sensitivity of matched field processing to sound-speed prone mismatch for vertical arrays in a deep water Pacific environment," J. Acoust. Soc. Am. **85**, 2394–2404 (1989).

[9] M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **112**, 1523–1535 (1992).

[10] M. Nicholas, J. S. Perkins, G. J. Orris, and L. T. Fialkowski, "Environmental inversion and matched-field tracking with a surface ship and L-shaped receiver array," J. Acoust. Soc. Am. **116**, 2891–2901 (2004).

[11] M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **90**, 1410–1422 (1991).

[12] L. T. Fialkowski, M. D. Collins, J. Perkins, and W. A. Kuperman, "Source localization in noisy and uncertain ocean environments," J. Acoust. Soc. Am. **101**, 3539–3545 (1997).

[13] S. E. Dosso and M. J. Wilmut, "Comparison of focalization and marginalization in Bayesian tracking in an uncertain environment," J. Acoust. Soc. Am. **125**, 717–722 (2009).

[14] A. J. Viterbi, "Error bounds on convolutional codes and an asymptotically optimal decoding algorithm," IEEE Trans. Inf. Theory **13**, 260–299 (1967).

[15] S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1989).

[17] D. Tollefsen and S. Dosso, "Geoacoustic information content of horizontal line array data," IEEE J. Ocean. Eng. **32**, 651–662 (2007).

[18] E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).

[19] G. K. Bhattacharyya and R. A. Johnson, *Statistical Concepts and Methods* (Wiley, New York, 1977).

[20] S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2008).

[21] 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).

[22] D. Tollefsen and S. E. Dosso, "Bayesian geoacoustic inversion of ship noise on a horizontal array," J. Acoust. Soc. Am. **124**, 788–795 (2008).

Absolute calibration of hydrophones immersed in sandy sediment

Gary B. N. Robb
National Oceanography Centre, University of Southampton, European Way, Southampton SO14 3ZH, United Kingdom

Stephen P. Robinson, Pete D. Theobald, and Gary Hayman
National Physical Laboratory, Teddington, Middlesex TW11 0LW, United Kingdom

Victor F. Humphrey, Timothy G. Leighton, and Lian Sheng Wang
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

Justin K. Dix and Angus I. Best
National Oceanography Centre, University of Southampton, European Way, Southampton SO14 3ZH, United Kingdom

(Received 27 June 2008; revised 27 February 2009; accepted 3 March 2009)

An absolute calibration method has been developed based on the method of three-transducer spherical-wave reciprocity for the calibration of hydrophones when immersed in sandy sediment. The method enables the determination of the magnitude of the free-field voltage receive sensitivity of the hydrophone. Adoption of a co-linear configuration allows the acoustic attenuation within the sediment to be eliminated from the sensitivity calculation. Example calibrations have been performed on two hydrophones inserted into sandy sediment over the frequency range from 10 to 200 kHz. In general, a reduction in sensitivity was observed, with average reductions over the frequency range tested of 3.2 and 3.6 dB with respect to the equivalent water-based calibrations for the two hydrophones tested. Repeated measurements were undertaken to assess the robustness of the method to both the influence of the sediment disturbance associated with the hydrophone insertion and the presence of the central hydrophone. A simple finite element model, developed for one of the hydrophone designs, shows good qualitative agreement with the observed differences from water-based calibrations. The method described in this paper will be of interest to all those undertaking acoustic measurements with hydrophones immersed in sediment where the absolute sensitivity is important. [DOI: 10.1121/1.3106530]

PACS number(s): 43.30.Xm, 43.30.Ma, 43.30.Yj, 43.58.Vb [KGF] Pages: 2918–2927

I. INTRODUCTION

Marine engineers require a detailed knowledge of the physical properties of seafloor sediments to site foundations for oil and gas drilling rigs and pipelines, offshore wind farms, and telecommunication cables. At present these properties are predominantly obtained using techniques for analyzing acoustic reflection/refraction data that require ground-truthing cores and samples.[1–3] A thorough understanding of the relationships between the acoustical and physical properties of seafloor sediments would negate the need for ground-truthing. Unfortunately, the nature of these relationships is still under debate, with a variety of equally valid geoacoustic theories in circulation for both saturated sediment[4–6] and sediment containing free-gas bubbles.[7–9]

In situ experiments allow the acoustical properties of well-defined sediment volumes to be measured, the physical properties of which can be determined through the laboratory analysis of sediment samples.[10–15] The examination of acoustical and physical data obtained using *in situ* techniques therefore offers a valuable way of refining our knowledge of the acoustical-physical relationships required for more effective inversion of acoustic data. While a variety of more complex *in-situ* experimental techniques exist, which require the insertion of three or more acoustic probes into the sediment, the use of a single source-receiver pair is still frequently adopted owing to its relatively simple deployment and a reduced sediment disturbance.[11,12,14,16] The processing techniques necessary for the analysis of the resulting transmit-receive data require knowledge of the transmitting current response of the source and sensitivity of the receiver, both of which are commonly determined through water-based calibrations. It has, however, been noted that these transducer-based properties may vary with the acoustic impedance of the medium into which the transducer is inserted.[8,11,17]This is particularly true for frequencies where the radiation impedance is a significant contribution to the overall device impedance such that the impedance of the medium can affect the hydrophone performance (typically, this would be at frequencies close to the resonance frequency). The characteristic acoustic impedance of sediment (for example, 1.7×10^6 kg m^{-2} s^{-1} for a mud to 3.8×10^6 kg m^{-2} s^{-1} for a sand) may considerably exceed that of water (i.e., $(1.4-1.5) \times 10^6$ kg m^{-2} s^{-1} for water with temperatures from 0 to 20 °C and salinities from 0‰ to 35‰). Therefore, cali-

0001-4966/2009/125(5)/2918/10/$25.00

bration data derived from water-based measurements may not be valid when the hydrophone is used in sediment. This motivates the work described here to develop a method which may be applied to sediment-based calibrations. The method will be beneficial to sediment acousticians relying on the absolute sensitivity of sediment-immersed hydrophones.

This paper describes and illustrates the calibration method with results from two types of hydrophone which were calibrated while immersed in sandy sediment. Section II presents the calibration technique which is based on a modified version of the three-transducer spherical-wave reciprocity method originally described by Luker and Van Buren for use in hydrophone phase calibration in water.[18] The method allows sediment-based absolute calibrations of receive sensitivity and transmit current response to be determined without knowledge of the acoustic properties of the sediment. Section III presents the results of reciprocity calibrations performed in saturated fine sands under laboratory conditions. Section IV shows a comparison of the results with a simple finite element (FE) model of one of the hydrophones, while conclusions are drawn in Sec. V.

II. SEDIMENT-BASED RECIPROCITY CALIBRATION TECHNIQUE

At present, the primary method in which hydrophones are calibrated is that of three-transducer spherical-wave reciprocity.[19] The method provides free-field values for both the receive voltage sensitivity and transmitting current response of the hydrophones under test. Throughout the work described here, the definitions used are those that are common in the calibration of electroacoustic transducers and are defined by the International Electrotechnical Commission.[19] Specifically, the free-field receive voltage sensitivity in a given direction and for a given frequency is the ratio of the open-circuit voltage developed by the hydrophone to the sound pressure that would exist in the undisturbed free-field at the position of the hydrophone if the hydrophone were absent. The SI units are V Pa^{-1}, but the sensitivity is commonly expressed in decibels as dB re 1 V μPa^{-1}. The transmitting current response in a given direction and for a given frequency is the ratio of the sound pressure at a reference distance from the transducer to the electrical current flowing through the terminals. The reference distance is defined as 1 m. The SI units are Pa m A^{-1}, but the sensitivity is commonly expressed in decibels as dB re 1 μPa A^{-1} at 1 m. Note that this is a far-field quantity, and a spherical-wave field is implied with the sound pressure being inversely proportional to the distance from the source. In the work described here, only the magnitude of the hydrophone response is considered.

The method of three-transducer spherical-wave reciprocity involves the transmission of signals between three pairs of transducers, commonly designated P, T, and H. The typical measurement configurations used require successive acoustic transmission from P to T, P to H, and T to H (see Fig. 1). The sensitivity of hydrophone H, which is denoted by M_{HW}, can then be derived from the measured transfer impedance (i.e., the quotient of the received voltage divided by the driving current) for each the transmitter pair using

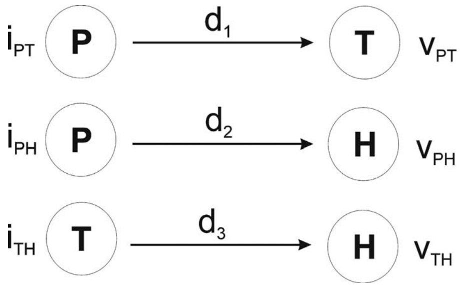


FIG. 1. Measurement configurations required for a standard water-based three-transducer reciprocity calibration. To obtain the sensitivity of hydrophone H, two additional transducers are required, namely, P and T, and measurements are required between the three transducer pairs of P to T, P to H, and T to H, which are separated by the distances d_1, d_2, and d_3, respectively. The respective driving currents for these three pairs are denoted by i_{PT}, i_{PH}, and i_{TH}, while the received voltages are denoted by v_{PT}, v_{PH}, and v_{TH}.

$$M_{HW} = \sqrt{\left(\frac{2d_2d_3}{\rho f d_1}\right)\left(\frac{Z_{PH}Z_{TH}}{Z_{PT}}\right)}, \tag{1}$$

where d_1, d_2, and d_3 are, respectively, the distances between P and T, P and H, and T and H,f is the acoustic frequency, ρ is the bulk density of the surrounding medium, and Z_{PH}, Z_{TH}, and Z_{PT} are the respective measured transfer impedances for transmission from P to H, T to H, and P to T. The transmitting current response of hydrophone H, which is denoted by S_{HW}, can also be determined through the spherical-wave reciprocity parameter J using

$$J = \frac{M_{HW}}{S_{HW}} = \frac{2}{\rho f}. \tag{2}$$

This standard reciprocity method makes a number of assumptions. First, the conditions are assumed to be free-field. This limits the time-window that can be included in the analysis since measurements must be made before the arrival of boundary reflections. Second, all receivers are assumed to lie in the far-field of the corresponding projector where, for a simple source, pressure is inversely proportional to the distance from the source. Third, it is necessary that transducer T be reciprocal, i.e., it is electrically passive, linear, and reversible. Fourth, it is assumed that the signals analyzed are steady-state, and therefore any initial transducer transients must be allowed to settle before any measurement window is applied. A final assumption present in Eq. (1) is that the attenuation of sound in water, which arises from absorption losses only, is negligible. This assumption is generally valid for kilohertz frequencies (for example, absorption at 200 kHz is 8.6×10^{-3} Np m^{-1} or less[20]) while the homogeneous nature of water allows correction factors to be determined for higher frequencies.[19]

In contrast to water, the attenuation of acoustic waves in the range 16–100 kHz in saturated sediment, which consists of absorption and scattering losses, can reach values of 2.9 Np m^{-1} in muds and 9.5 Np m^{-1} in sand (predictions from the grain-shearing theory[5] for typical sediment properties[21,22]). If this attenuation could be accurately predicted, suitable correction factors could be obtained. However, this is not practical owing to the highly variable nature

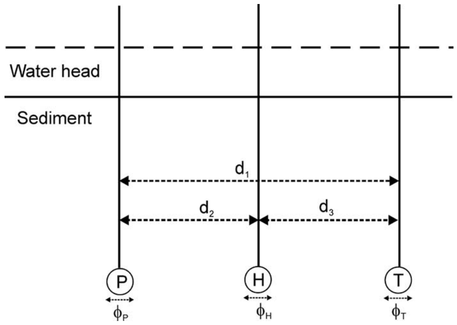


FIG. 2. Co-linear arrangement for three-transducer reciprocity required for sediment-based measurements which utilizes a projector P and reciprocal transducer T to calibrate hydrophone H.

of these attenuations. For example, a compilation of attenuation data from marine sediment displays a scatter of ±31% for a unique mean grain size,[6] while attenuation coefficients measured in sandy sediment varies by up to 2.8 Np m^{-1} for sediments with similar physical properties lying within a 100 m distance of one another.[22] This variability makes it extremely difficult to predict, and to account for, the attenuation losses in sediment. It is therefore preferable to devise a calibration method which does not critically depend on the need to correct for attenuation losses. This may be achieved through the co-linear arrangement displayed in Fig. 2, based on the configuration originally used by Luker and Van Buren for hydrophone phase calibration in water.[18] The method is described below.

Consider the general case of a pair of transducers that are embedded in sediment, comprising a projector P and hydrophone H whose reference centers are separated by a distance d. If a driving current i_p is applied to the projector, and it is assumed that spherical spreading losses apply and the sediment is homogeneous, the voltage v_H received by the hydrophone can be expressed as

$$v_H = \frac{M_H S_P i_P}{d} e^{-\alpha d}, \tag{3}$$

where S_P is the transmitting current response of the projector, M_H is the sensitivity of the hydrophone, and α is the attenuation coefficient of the sediment (in Np/m). The specific transfer impedances for transducer pairs P to $T(Z_{PT})$, P to $H(Z_{PH})$, and T to $H(Z_{TH})$ are therefore given by

$$Z_{PT} = \frac{M_{TS} S_P}{d_1} e^{-\alpha[d_1-(\phi_P+\phi_T)/2]},$$

$$Z_{PH} = \frac{M_{HS} S_P}{d_2} e^{-\alpha[d_2-(\phi_P+\phi_H)/2]},$$

$$Z_{TH} = \frac{M_{HS} S_T}{d_3} e^{-\alpha[d_3-(\phi_T+\phi_H)/2]}, \tag{4}$$

where M_{HS} and M_{TS} are the sediment-based sensitivities of hydrophone H and transducer T, respectively, and S_T and S_P are the transmitting current responses of the transducers T and P, respectively. The separations between the reference centers of P and T, P and H, and T and H are denoted by d_1, d_2, and d_3, respectively, while the diameter of transducers P, T, and H are denoted by ϕ_P, ϕ_T, and ϕ_H, respectively. Equation (4) assumes that the acoustic reference center of each transducer lies at the geometric center of the device. Using the reciprocity parameter J, as described in Eq. (2), it is possible to combine the expressions for the transfer impedances Z_{PH}, Z_{PT}, and Z_{TH} to derive an expression for the sensitivity M_{HS} of the hydrophone H:

$$M_{HS} = \sqrt{\left(\frac{2d_2d_3}{\rho f d_1}\right)\exp[\alpha(d_2+d_3-d_1-\phi_H)]\left(\frac{Z_{PH}Z_{TH}}{Z_{PT}}\right)}. \tag{5}$$

As a consequence of the co-linear arrangement adopted, the separation distances are related by the expression $d_1=d_2+d_3$ and Eq. (5) reduces to

$$M_{HS} = \sqrt{\left(\frac{2d_2d_3}{\rho f d_1}\right)\exp(-\alpha\phi_H)\left(\frac{Z_{PH}Z_{TH}}{Z_{PT}}\right)}. \tag{6}$$

For an infinitesimally thin hydrophone, the exponential term in Eq. (6) will reduce to unity and the sensitivity of the central hydrophone can be determined from the approximate form

$$M_{HS} = \sqrt{\left(\frac{2d_2d_3}{\rho f d_1}\right)\left(\frac{Z_{PH}Z_{TH}}{Z_{PT}}\right)}, \tag{7}$$

which corresponds to the water-based scenario displayed in Eq. (1). For the sands examined in the present work, the maximum attenuation coefficient for the frequency range examined (10–200 kHz) was estimated using the grain-shearing model[5] to be 8.16 Np m^{-1}. Combined with the diameters of the hydrophones used, which are 20–21 mm, respectively (see Sec. III A), the sensitivities determined from the exact equation [Eq. (6)] will vary from those determined from the approximate equation [Eq. (7)] by a maximum of 8.9%, i.e., sensitivity levels will deviate by less than 0.74 dB. As this deviation lies within the variability associated with sediment disturbance (see Sec. III B), the approximate equation is assumed valid for sediment-based reciprocity calibrations and is used throughout the remainder of this paper. Although the attenuation coefficients, frequency ranges and transducer diameters used in this work are typical of *in-situ* experiments, future users are advised to assess the validity of Eq. (7) for their own situations.

There are, however, certain alignment problems that arise from the co-linear arrangement shown in Fig. 2. First, hydrophones cannot be assumed to be omni-directional and so must be calibrated in a reference direction. The receiving device must therefore be aligned such that its reference direction is pointing toward the transmitting device. For the co-linear arrangement, this means that the central device H must be rotated between measurements of Z_{PT} and Z_{TH} so that H is facing the transmitting hydrophone in each case. Second, when measuring Z_{PT} or Z_{TP}, the central device should ideally be removed to avoid any "shadowing" effect on the acoustic field. While both the rotation and removal of H will have no effect for water-based measurements, for sediment-based measurements these adjustments may introduce some degree of disturbance into sediment and therefore alter the propagation loss between consecutive measurements used in the same calibration. The impact of these disturbance and shadowing effects is examined in Sec. III.

III. SEDIMENT-BASED RECIPROCITY MEASUREMENTS

A. Experiment

A series of reciprocity calibrations was performed on transducers inserted into saturated sands contained in two laboratory tanks. Initial measurements were made in a small test tank located at the National Physical Laboratory to validate the sediment-based reciprocity technique detailed in Sec. II. A second laboratory tank located at the University of Southampton was used for calibrations over a larger frequency range, the use of two test tanks providing a test of the robustness of the method. These tanks have been designated as Tank 1 and Tank 2, respectively.

The manner in which the sand was placed in both tanks was designed to minimize the possibility of trapping or forming air bubbles within the sediment. The inclusion of such bubbles would introduce strongly frequency-dependent compressional wave velocities and attenuations,[7] which would disrupt the waveforms received and, therefore, make it extremely difficult to identify the time-windows over which the signals are steady-state. To minimize the possibility of introducing air bubbles, the sand was sprinkled into degassed water (degassed using a vacuum pump prior to filling the tank) using a large container with 5 mm diameter holes drilled in the bottom, see Fig. 3(a), and the sediment was left to settle for a week prior to the measurement phase. The method used for inserting the transducers was also chosen to prevent the entrapment of air bubbles. This involved inserting an open-ended tube into the sediment and excavating the sand within it using a smaller tube. The transducer was then placed in the excavated hole through the outer tube, and the tube was then removed to allow the sediment to envelop the transducer. After each insertion, the sediment was given a few hours to settle.

The physical properties of the sand were measured through the collection and analysis of sediment cores. Bulk density, porosity, and compressional wave velocity were measured at 1 cm depth increments using a multi-sensor core logger,[23] while grain size distributions were measured at 5 cm intervals using a laser particle analyser. The sand was classified as a medium sand with a mean grain size of 304 ± 49 μm, a porosity of 33% ± 1.0%, a bulk density of 2177 ± 18 kg m^{-3}, and a compressional wave velocity of 1746 ± 8 m s^{-1} (all values quoted are mean values with standard deviations as the corresponding uncertainties). Here, the mean grain size is expressed as $-\log_2(d/d_0)$, where $d_0 = 1$ mm and d is the grain diameter for each class determined by sieving (i.e., the maximum grain diameter that can pass through a sieve mesh, assuming spherical grains). The mean grain size was determined from the cumulative frequency curve of percentage mass of the sample passing through each sieve against grain size fraction (sieve size) by reading off the grain sizes corresponding to the 16, 50, and 84 quartiles (i.e., 16% of the sample mass has a grain size less than D16, 50% less than D50, and 84% less than D84) and calculating (D16+D50+D84)/3. The Friedman & Sanders scale was then used to categorize the sediment according to mean grain size.[24] For the work here, the mean grain size was measured on a number of samples from an assumed homogeneous sediment, thus getting a standard deviation ± linear error.

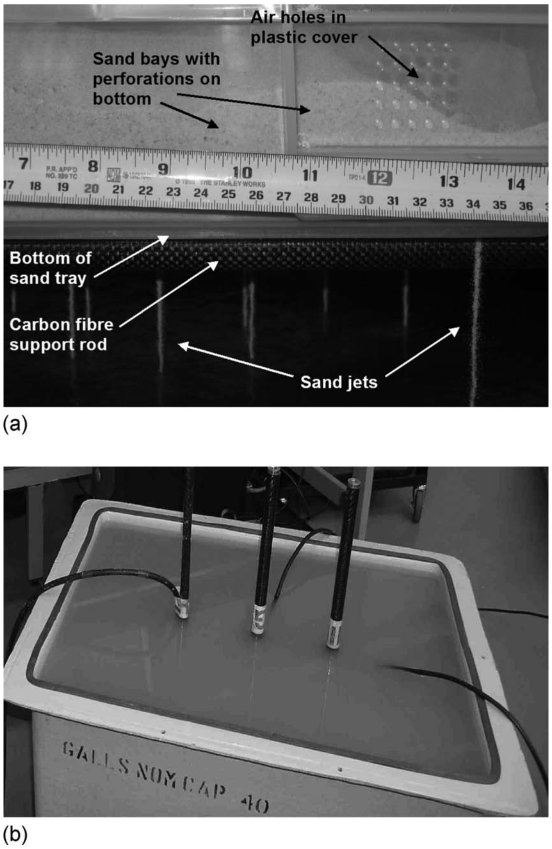


FIG. 3. (a) Manner in which sediment tank was filled (through sprinkling of sand from container at top of image into degassed water. (b) Reciprocity experiment being performed on Tank 1, with the poles attached to three transducers clearly visible.

The absence of gas bubbles was confirmed through a number of observations. First, the high fraction of the 500 kHz pulses that were transmitted though the core during the core analysis indicated a bubble-free medium. Second, all signals acquired during the reciprocity measurements displayed clean waveforms, without any additional frequency components that would be indicative of scattering from bubbles. Third, upon gentle stirring of the sediment no bubbles were observed to emerge from the sediment.

Tank 1 contained a sediment volume measuring 0.67 m by 0.49 m with a depth of 0.45 m and a 50 mm head of water, see Fig. 3(b). All transducers were inserted to a depth of 0.22 m in the sediment, with the central hydrophone H placed at the centre of the tank and, to maximize the echo-free time, the outer devices P and T placed 0.13 m on either side of H. To determine the transfer impedances required to calculate the sensitivity of hydrophone H, the steady-state driving current and received voltage were measured for each transducer pair (namely, P to T, P to H, and T to H). These measurements were acquired for tone-burst signals with frequencies that covered the range from

20 to 150 kHz in 2 kHz steps. The steady-state voltage of the received signal was measured using standard techniques,[19] which involve the application of a time-window to select the portion of the received signal that satisfies both steady-state conditions (i.e., contains no ringing affects associated with the start of the received signal) and free-field conditions (i.e., contains no reflected signals). Least squares fitting techniques were then applied to this windowed signal to measure the amplitude of the received voltage, with reliable results obtained when the duration of the time-window contained at least half a period of the received signal.[25] A type 8104 transducer (manufactured by Brüel and Kjær, Denmark) with a diameter of 21 mm was selected as the central hydrophone (H) because the minimal ringing effects associated with this relatively heavily damped device allowed a relatively long time-window (32.9 μs) to be used and therefore frequencies as low as 20 kHz to be examined. In order to allow signals with sufficient amplitude to be transmitted over the required frequency range an ITC 1042 transducer with a diameter of 35 mm (manufactured by International Transducer Corporation) was selected as the projector (P) and a Brüel and Kjær 8100 with a diameter of 21 mm was used as the reciprocal transducer (T).

While the transfer impedances Z_{PH} and Z_{TH} required to derive the sensitivity of H from Eq. (7) can only be measured with H present, Z_{PT} was measured under three sets of conditions. First, the use of Z_{PT} measured before H was inserted allowed a "reference" sensitivity to be calculated that was subject to no shadowing effect and only relatively minor disturbances associated with the insertion of P and T. Second, a "shadowed" sensitivity was calculated through the use of Z_{PT} measured with H present, which introduces both a shadowing effect and an additional disturbance associated with the insertion of H. Third, on the removal of H, a final measure of Z_{PT} was obtained; while this measurement will have no shadowing effects associated with it, it will be affected by the sediment disturbance associated with the insertion and removal of H and is therefore referred to as the "disturbed" sensitivity. In addition, the transfer impedance between the outer transducers was measured in both directions, i.e., from P and $T(Z_{PT})$ and from T and $P(Z_{TP})$. This allowed the sensitivity to be calculated with either Z_{PT} or Z_{TP} in the denominator of Eq. (7), and the validity of the reciprocal assumption to be tested (if both P and T are reciprocal then both sensitivities should be the same).[19]

Calibrations were also performed in Tank 2, with the aim of repeating the earlier work and obtaining sensitivities at a wider range of frequencies. Here, the hydrophone under test consisted of a spherical element with a diameter of 12.5 mm embedded in a cylindrical polyurethane boot with an outer diameter of 20 mm. This hydrophone was manufactured by Neptune Sonar Ltd. UK, and consisted of a modified version of their D140 design (the design used here has the same spherical element, but is encapsulated in a cylindrical boot). This hydrophone was designed for *in-situ* fieldwork over the frequency range 10–200 kHz. Tank 2 was cylindrical, with a diameter of 0.97 m, a sediment depth of 0.87 m, and water head of 0.10 m. The slightly larger dimensions of Tank 2 allowed the length of the time-window over which the received signal satisfied steady-state and free-field conditions to be extended to 50 μs, enabling measurements to be made at somewhat lower frequencies than for Tank 1. The central hydrophone was placed at a depth of 0.45 m in the center of the tank and, to maximize echo free time, with P and T placed in a co-linear arrangement 0.17 m on either side of H. Other details of the experiment, such as the transducers used for P and T and the pulse lengths, remained as described above for Tank 1. However, since degassed water was not available, a submersible pump was used to assist with degassing of water and sediment. This pump, capable of driving a 30 m head of water, had a constricting valve fitted to the inlet which was used to throttle the flow. The resulting pressure drop allowed the water/sediment mixture to degas by drawing dissolved gases out of the water. The outlet was vented to atmospheric pressure at the water surface with a glass beaker used to collect the exiting gas (allowing the gas production to be monitored). This degassing pump was used to degas the water prior to filling the tank with sand, with steadily decreasing amounts of gas harvested from the tank and captured in the glass beaker as the water reached a degassed state. Once degassed, the water tank was filled with sediment to envelop the degassing pump which was protected by a flooded cylindrical column. As the column lay below the water surface, degassing continued during sand filling of the tank and prior to measurements. A small increase in the amounts of gas harvested from the tank during sand filling provided some evidence that the filling process did indeed introduce air/gas into the water. The degassing pump was used for several hours in the saturated sediment prior to measurements until the volume collection rate of gas reached an equilibrium. While there remains the possibility that small fractions of gas were present in both sediment tanks, the evidence presented earlier indicates that these gas fractions were not sufficient to impact on the measurement technique presented here.

For comparison purposes, water-based reciprocity measurements were also performed in a water tank measuring 2.0 m long by 1.5 m wide and 1.5 m deep for both the Brüel and Kjær 8104 and the spherical test hydrophone. In order to ensure that the water and sediment calibrations are directly comparable the transducer configurations, mounting, drive voltage, and pulse lengths remained unchanged. The P to T and T to P stages were again measured both with and without H present, i.e., under reference and shadowed conditions. For the water-based reciprocity calibrations, the overall measurement uncertainties (expressed for a confidence level of 95%) were estimated to be typically 0.5 dB.[26]

It is normal for three-transducer spherical-wave reciprocity calibrations to be undertaken with hydrophone separation distances that are sufficiently large to ensure that far-field conditions are achieved and that the acoustic field is sufficiently close to a spherically- diverging wave.[19] For the sediment-based calibrations reported here, relatively short transducer separations were used because of the use of relatively small test tanks (separations of 0.13 and 0.17 m, respectively). For the higher frequencies used here, the strictest acoustic far-field criterion is violated, for example, that specified in the international standard IEC 60565.[19] This

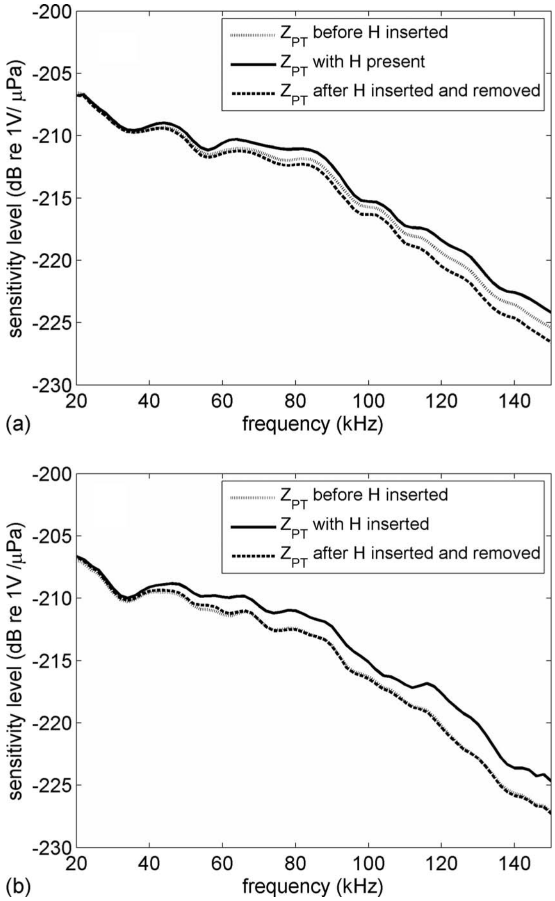


FIG. 4. An examination of the impact of shadowing and disturbance effects for the Brüel and Kjær 8104. Two sets of repeat measurements are displayed in (a) and (b), respectively. These measurements include reference sensitivity levels calculated using Z_{PT} measured before H was inserted (dotted line), shadowed sensitivity levels calculated using Z_{PT} measured with H present (solid line), and disturbed sensitivity levels calculated using Z_{PT} measured after H was inserted and then removed (dashed line). All measurement used the same P to H and T to H measurements. Note that the repeatability of the measurements with repeated insertion of the hydrophones was estimated to be between ±0.1 and ±0.4 dB (expressed as a standard deviation).

means that the acoustic field will not approximate to a spherical-wave to within a maximum tolerance of ±2% over the full frequency range, as required by the standard. With the relatively small transducer elements used here, this was not considered to be a significant source of error. However, to ensure that any observed differences in the calibration results were not due to this factor, the water-based calibrations were undertaken at the same separation distances as were used in the sediment-based calibrations. In addition, extra calibrations were undertaken in water at large separation distances (minimum of 1.1 m), so that comparisons were also possible with full far-field water-based results.

B. Results

Sensitivity levels measured on the Brüel and Kjær 8104 in Tank 1, which are used to validate the technique described in Sec. II, are displayed in Figs. 4 and 5. The sensitivities are expressed in decibels relative to a reference value of 1 V/μPa. It should be noted that the agreement between the sensitivity levels calculated using Z_{PT} and Z_{TP} was excellent, with discrepancies of less than 1% seen over much of the frequency range for both water-based and sediment-based measurements. This strongly suggests that both P and T were behaving reciprocally. The impact of shadowing and disturbance effects on the sensitivity levels are displayed in Fig. 4 through the use of the reference, shadowed, and disturbed sensitivity levels discussed in Sec. III A. These three sensitivity levels were derived for two sets of measurements obtained from two deployments of H, with a period of 48 h allowed between the first and second sets to allow the sediment to re-settle. The combined shadowing and disturbance effects associated with the insertion and presence of H causes sensitivity levels to increase from the reference sensitivity level by a maximum of 1.3 dB for the first set of measurements and 2.8 dB for the second set, with the effect more pronounced at higher frequencies. This increase can be primarily explained by the presence of H reducing the amplitude of the signal that is transmitted from P to T and, therefore, reducing the value of Z_{PT} in Eq. (7). The disturbed sensitivity levels lie a maximum of 1.2 dB below the reference sensitivity level for the first set of measurements and deviate from the corresponding reference values by less than 0.3 dB for the second set, which again is more pronounced at higher frequencies. This reduction can be explained by considering that the removal of H reduced the compaction of sediment in the region around the hydrophone position. The resulting increase in porosity causes the attenuation of the sediment to be reduced and the amplitude of the signal transmitted from P to T to be increased. The subsequent increase in the value of Z_{PT} in Eq. (7) leads to lower sensitivity levels.

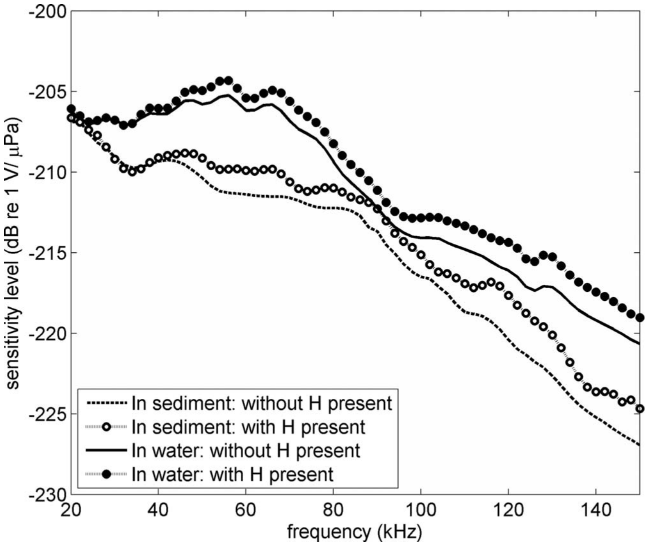


FIG. 5. Sensitivity levels measured on the Brüel and Kjær 8104 hydrophone in sediment and water. Sediment-based sensitivity levels measured in Tank 1 are displayed using P and T measurements both with H present (open circles) and without H present (dashed line). Water-based measurements performed in a larger water filled tank (measuring $2 \times 1.5 \times 1.5$ m^3) are also displayed for P and T measurements both with H present (closed circles) and without H present (solid line). Note that the overall uncertainty for the calibrations were estimated to be ±0.5 dB for the water-based calibrations, and between ±0.9 and ±1.5 dB for the non-shadowed sediment-based calibrations (expressed for confidence levels of 95%).

The water-based and sediment-based sensitivities of the Brüel and Kjær 8104 are compared in Fig. 5. While reference

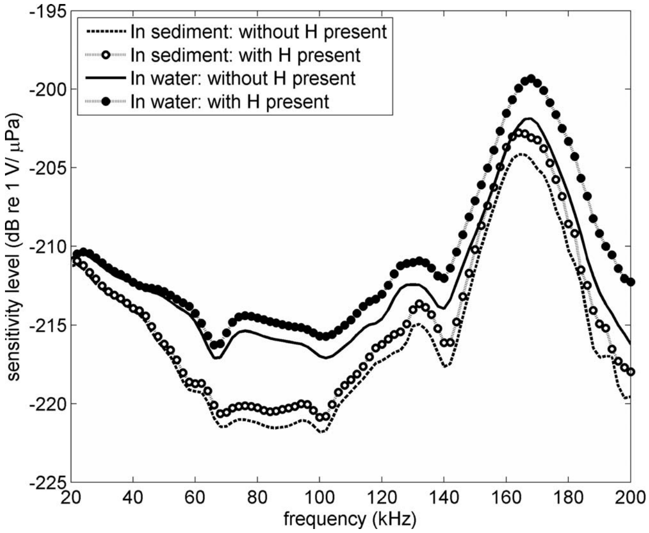


FIG. 6. Sensitivity levels measured for the spherical test hydrophone in sediment and water. Sediment-based sensitivity levels measured in Tank 2 are displayed using *P* and *T* measurements with *H* present (open circles) and without *H* present (dashed line). Water-based measurements performed in a larger water filled tank (measuring 2 m by 1.5 m by 1.5 m) are also displayed for *P* and *T* measurements with *H* present (closed line) and without *H* present (solid line). Note that the overall uncertainty for the calibrations were estimated to be ±0.5 dB for the water-based calibrations, and between ±0.9 and ±1.5 dB for the non-shadowed sediment-based calibrations (expressed for confidence levels of 95%).

sensitivity levels (i.e., those that use *P* to *T* measurements without *H* inserted) are traditionally used for water-based calibrations, shadowed measurements (i.e. those that use *P* to *T* measurements with *H* present) may be a more practical scenario for *in-situ* calibrations, as these would allow the three transducers to be attached to a single rig with well-defined separations for a single insertion into the sediment. For this reason both shadowed and non-shadowed sensitivities have been included in the following discussion. Comparison of the reference and shadowed sensitivity levels for each medium display a maximum increase from shadowing effects of 2.7 dB in sediment and 1.9 dB in water, both of which are more pronounced as frequency increases. The sensitivity levels in sediment are generally significantly lower than that in water, with a mean reduction across the frequency range of 3.8 dB for the shadowed measurements and 3.6 dB for the reference measurements. This reduction in sensitivity is greater than the degree of variability associated with disturbance effects, which, as discussed above, is less than 1.2 dB. The difference between sediment-based and water-based sensitivity levels varies with frequency, with this difference less than 2 dB for the range 20–28 kHz and 84–98 kHz, and greatest between these regions in the vicinity of the water-based resonance frequency of *H* (65 kHz).

The measured sensitivities of the spherical test hydrophone are displayed in Fig. 6, which again includes both reference and shadowed measurements. As for the Brüel and Kjær 8104, sediment-based sensitivity levels are lower than water-based values, with a mean reduction of 3.6 dB for shadowed measurements and 3.2 dB for non-shadowed measurements. This reduction is most pronounced for frequencies between 70 and 100 kHz, which, in contrast to the Brüel and Kjær 8104, are below the water-based resonance frequency of the hydrophone (i.e., 170 kHz). The increase in sensitivity caused by shadowing effects is observed to be considerably greater for water-based measurements (maximum increase of 4.0 dB at 200 kHz) than sediment-based measurements (maximum increase of 2.0 dB at 200 kHz).

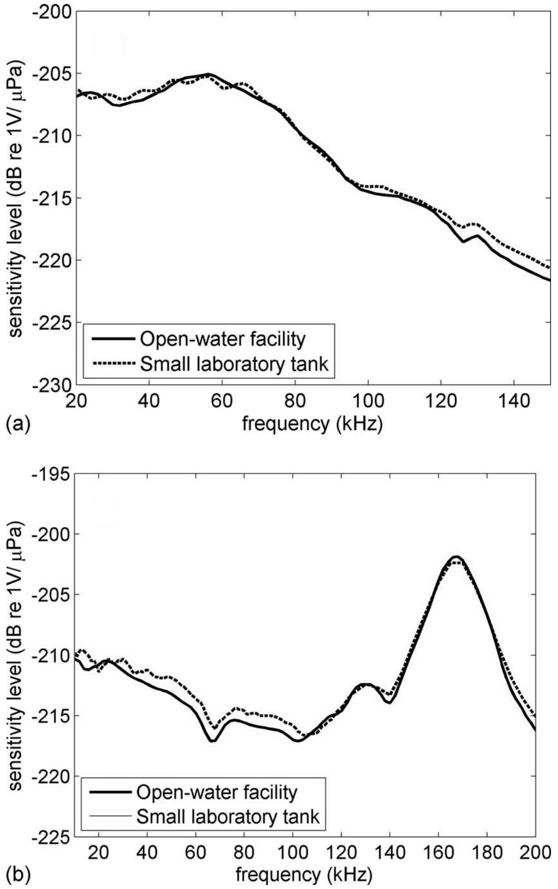


FIG. 7. Comparison of free-field sensitivity levels measured using large separations in an open-water facility (solid line) with water-based sensitivity levels measured in a small laboratory tank using the shorter transducer separation required for sediment-based work (dashed line) for (a) Brüel and Kjær 8104 and (b) spherical test hydrophone. Note that the overall uncertainty for the water-based calibrations were estimated to be ±0.5 dB (expressed for confidence levels of 95%).

In addition, to examine the effect of the restricted acoustic far-field conditions, water-based sensitivity levels measured at the short separation distances were compared to those obtained at larger separations, the latter measurements being undertaken in an open-water facility. The use of larger facilities allowed a minimum transducer separation of 1.1 m to be used and the resulting sensitivities to represent true far-field conditions. For the Brüel and Kjær 8104 transducer arrangement [see Fig. 7(a)] sensitivity lie within 0.6 dB of the free-field values from 20 to 122 kHz, and deviated from these free-field values by less than 1.1 dB as frequency increases to 150 kHz. The spherical test hydrophone displayed a similar deviation [see Fig. 7(b)] from true free-field values, with deviations less than 1 dB from 10 to 200 kHz.

The above results indicate that, for the two transducers examined, the change in the medium loading arising from the insertion of transducers into sediment will reduce sensitivity levels, averaged over the frequency range tested, by

3.2 and 3.6 dB respectively. These reductions exceed variations introduced by sediment disturbance effects (which are less than 1.2 dB) and are more pronounced for shadowed sensitivity levels. The reduction in sensitivity is probably a consequence of the increased acoustic impedance of the fine sand (3.79×10^6 kg m^{-2} s^{-1}) with respect to the acoustic impedance of the water (1.45×10^6 kg m^{-2} s^{-1}) and the resulting increased mismatch with the boot material used to protect the hydrophone element.

C. Uncertainties

For free-field reciprocity calibrations performed in water, the overall uncertainty is typically ±0.5 dB (expressed for a 95% confidence level). A comprehensive description of the sources of uncertainty is provided elsewhere in the literature.[19,26] Many of these sources of uncertainty are common to the calibrations performed in sediment, examples being lack of steady-state conditions, lack of spherical-wave propagation, instrument calibration uncertainty, and lack of far-field conditions. The values of these uncertainties can vary with frequency, for example, the contribution from lack of steady-state conditions is greater at the lower limit of the frequency range where there are fewer cycles available within the echo-free time-window for analysis. Conversely, the lack of far-field conditions contributes greater uncertainty at higher frequencies where the near-field region extends for a greater distance.

In addition, there are additional uncertainties specific to the calibrations in sediment, a number of which have already been the subject of discussion in the previous sections. These include the residual effect of the finite size of the hydrophones, which was estimated to be a maximum value of 0.74 dB in Sec. II. The influence of shadowing caused by the central hydrophone has already been discussed in Sec. III B. The reduced separations used for the sediment calibrations introduced the potential for increased contributions from lack of far-field conditions. This was assessed by undertaking two sets of water-based calibrations, one set at the same marginal separations used for the sediment work, and one set at increased separations where far-field conditions were comfortably satisfied. This contribution was estimated in Sec. III B at between 0.6 and 1.1 dB.

Any calibration method will also suffer from a "random" uncertainty due to the lack of perfect repeatability in the measurements. This will be exacerbated for the sediment calibrations due to disturbance of the sediment on repeated immersion of the hydrophones. This was examined experimentally by measuring the transfer impedance between hydrophones under repeated immersion, extraction, and re-immersion of the Brüel and Kjær 8104 hydrophone. The repeatability obtained varied relatively little with frequency and was typically in the range 0.1–0.4 dB (expressed as a standard deviation).

An additional potential contribution specific to the sediment-based calibrations is that of lateral variations in attenuation (causing the value of α to be different for paths d_2 and d_3). Such a variation can be incorporated in to the analysis by using different attenuations in Eq. (4). As the total attenuation over the sediment part of the path d_1 is equal to the total attenuation over the paths d_2 and d_3 then the resulting equations will reduce to Eq. (6) indicating that such lateral variations should not be an issue. The experimental protocol (including the method of lying down a uniform degassed sediment described in Sec. III A) makes lateral variation in α far less likely than would occur in real shallow marine sediment, with its potential for the presence of flora and fauna (including shells), gas, and mineral inclusions. Assessment of any lateral variation in α would require removal and re-insertion of hydrophones into the sediment, which introduces another error that is potentially at least as large as the one being measured. To investigate the potential for any lateral variation in α to affect the results of this work, during the initial measurements using the Brüel and Kjær 8104, the roles of P and T (and the identity of the hydrophones) were interchanged. This provided a test of the robustness of the method to any lateral variations since the transducer pairs associated with the separations d_2 and d_3 (and the corresponding transfer impedances) were now interchanged. Although these measurements were only undertaken at a subset of frequencies within the overall frequency range, the results showed that the differences obtained were within the repeatability obtained from simply removing and re-inserting the hydrophones (typically between 0.1–0.4 dB as indicated above). This provided confidence that lateral variations in attenuation were not a significant source of error for the sediment tanks used here.

Although the work described here is mainly intended as a statement of the methodology rather than a definitive study of uncertainties, the additional sources of uncertainty have been combined with those common sources from the water-based calibrations in an attempt to make a provisional estimate of the overall uncertainty for sediment-based calibrations at a subset of frequencies within the overall frequency range. Combining the uncertainties according to the ISO Guide to Uncertainties in Measurement,[27] overall values of between ±0.9 and ±1.5 dB (depending on frequency) are obtained when using the reference "non-shadowed" method (expressed for a confidence level of 95%).

When considering the differences observed between the water-based and sediment-based results, it should be remembered that as far as possible the same instrumentation and experimental procedure were used for both media. This means that any systematic bias introduced into the results due to uncertainty contributions which are common to both experiments will be the same for both media and so will not influence the *differences* between the results. As can be seen from the results presented in Sec. III B, sensitivity differences are observed which exceed the estimated uncertainties, and therefore it is believed that these differences are real for the hydrophones used here and are not experimental artifacts.

IV. COMPARISON WITH HYDROPHONE MODEL

In order to investigate the effect of immersion in sand on a hydrophone performance and confirm that the observed changes in sensitivity were realistic, preliminary FE calculations were performed using a simplified model of the spheri-

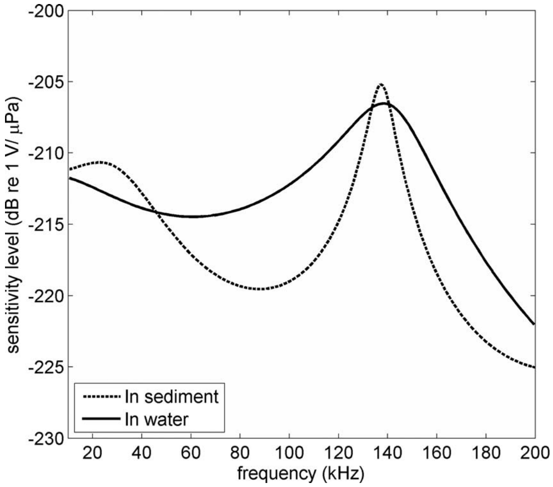


FIG. 8. Absolute sensitivity of spherical hydrophone with spherical boot in water and sand calculated using FE model.

cal hydrophone. These used a commercial FE code (PAFEC, PACSYS Ltd., Nottingham, UK). For reasons of simplicity, the model was only developed for the spherical hydrophone. The Brüel and Kjær 8104 hydrophone was not modeled since the complex element design, consisting of four coaxial rings, was not readily amenable to a simplified model.

The model consisted of a radially-poled spherical shell of PZT 5 piezoelectric ceramic with inner and outer radii of 5.0 and 6.35 mm, respectively, covered by a uniform coating layer of outer radius 10 mm (the hydrophone boot), based on information provided by the hydrophone manufacturer. The hydrophone was modeled when immersed in water, with a speed of sound 1500 m s^{-1} and density 1000 kg m^{-3}, and in sand, with a speed of sound 1746 m s^{-1} and density 2176 kg m^{-3}. The parameters for the piezoelectric material were taken from the database within PAFEC and the coating layer was assumed to have a compressional wave speed of 1626 m s^{-1} and a density of 1040 kg m^{-3}. In the model, the coating and the sediment were assumed to act as fluids (and so would not support shear waves). However, realistic values of the compressional wave speed have been used which correspond to those for the sediment. The model utilized boundary elements on the outer surface of the coating to simulate the effect of the infinite fluid medium (water or sand). These properties are considered to be reasonable estimates of the properties of the tested device.

In order to evaluate the performance of the hydrophone as a receiver, the transmit voltage sensitivity of the device was initially calculated. This was then converted into a current sensitivity using the predicted impedance of the device. Finally, the receive sensitivity was calculated by use of the principle of reciprocity [see Eq. (2)]. It should be noted that no attempt has been made to correct the resulting element receive sensitivities to end of cable sensitivities.

The resulting receive sensitivities are shown in Fig. 8 as a function of frequency from 10 to 200 kHz. These results show that the sensitivities in sediment and water are very similar at the lowest frequency considered, but that the sensitivity in sand is much lower in the region of 90 kHz, being some 6 dB lower than in water. However, the sensitivities in the region of the resonance at 140 kHz are very similar, with that in sand being about 1.5 dB higher. The peak in response has a higher *Q* for the in-sand measurements than the in-water measurements. It should be noted that changes to the *Q* factor of the resonance have been predicted elsewhere for hydrophones immersed in sediment.[28]

The qualitative features displayed by the model are in very good agreement with those displayed by the experimental measurements of sensitivity (Fig. 6). In particular, the similar low frequency sensitivities, the significantly reduced sensitivity in the midfrequency region in sand, and similar sensitivity in the resonance region are all well replicated. This gives added confidence that these features are real and not the result of systematic uncertainties.

It should be noted that this FE model was not intended to model exactly the tested hydrophone with its complex construction of a spherical element in a truncated cylindrical boot. It is intended to develop a more extensive model in the future in order to understand in more detail how sediment affects the performance of buried hydrophones.

V. CONCLUSIONS

An absolute calibration method has been developed to investigate hydrophone performance in sediment. The method is based on the method of three-transducer spherical-wave reciprocity with a co-linear transducer arrangement which enables the sensitivity of the central hydrophone to be determined without *a priori* knowledge of the sediment properties. A series of reciprocity calibrations has been performed in sediment and in water for two types of hydrophone. Through comparison with equivalent water-based measurements, the immersion of the hydrophones into the sediment was observed to reduce their sensitivity levels by varying amounts depending on the frequency. The observed reductions varied between a minimum of less than 1 dB and a maximum of just over 7 dB, the average reduction in sensitivity for the two hydrophones being 3.2 and 3.6 dB, respectively (averaged over the all measurement frequencies). The effect of the sediment disturbance associated with the necessary insertion, rotation, and removal of the central hydrophone caused measured sensitivities to deviate from reference sensitivity levels by less than 1.2 dB. Shadowing effects associated with the use of *P* to *T* measurements with the central hydrophone present increased sensitivity levels by between 1.3 and 4.0 dB. Despite the relatively short transducer separations that were required for the sediment-based measurements, the lack of perfect far-field conditions for part of the frequency range did not cause significant error in the measurements. The reduction in sensitivity levels associated with insertion into sediment can be explained through the higher impedance of the sediment and increased mismatch with the hydrophone boot material. A simple FE model has been developed for one of the hydrophones, the results of which show good qualitative agreement with the measured data and indicate that the observed changes are realistic.

ACKNOWLEDGMENTS

This work was funded by the Engineering and Physical Sciences Research Council (Grant No. EP/D000580/1 based at the University of Southampton; Principal Investigator: Tim Leighton) and the Acoustics and Ionising Radiation Programme part of the National Measurement System Programmes Unit of the UK Department of Innovation, Universities and Skills. Thanks are extended to Richard Hazelwood for invaluable assistance in the use of vacuum pump techniques to degas the sediment, Jeremy Sothcott for core analysis, and Veerle Huvenne for grain size analysis.

[1] I. R. Stevenson, C. McCann, and P. B. Runciman, "An attenuation-based sediment classification technique using chirp sub-bottom profiler data and laboratory acoustic analysis," Mar. Geophys. Res. **23**, 277–298 (2002).

[2] S. P. R. Greenstreet, I. D. Tuck, G. N. Grewar, E. Armstrong, D. G. Reid, and P. J. Wright, "An assessment of the acoustic survey technique, Roxann, as a means of mapping seabed habitat," ICES J. Mar. Sci. **21**, 939–959 (1997).

[3] T. G. Leighton and G. B. N. Robb, "Preliminary mapping of void fractions and sound speeds in gassy marine sediments from subbottom profiles," J. Acoust. Soc. Am. **124**, EL313–EL320 (2008).

[4] R. D. Stoll, *Sediment Acoustics*, Lecture Notes in Earth Science 26 (Springer-Verlag, Berlin, 1974).

[5] M. J. Buckingham, "Compressional and shear wave properties of marine sediments: Comparisons between theory and data," J. Acoust. Soc. Am. **117**, 137–152 (2005).

[6] E. L. Hamilton, "Compressional-wave attenuation in marine sediments," Geophys. J. **36**, 620–646 (1972).

[7] A. L. Anderson and L. D. Hampton, "Acoustics of gas bearing sediments II. Measurements and models," J. Acoust. Soc. Am. **67**, 1890–1903 (1980).

[8] F. A. Boyle and N. P. Chotiros, "Nonlinear acoustic scattering from a gassy poroelastic seabed," J. Acoust. Soc. Am. **103**, 1328–1336 (1998).

[9] D. Gei and J. M. Carcione, "Acoustic properties of sediments saturated with gas hydrate, free gas and water," Geophys. Prospect. **51**, 141–157 (2003).

[10] M. J. Buckingham and M. D. Richardson, "On tone-burst measurements of sound speed and attenuation in sandy marine sediments," Inf. Sci. (N.Y.) **27**, 429–453 (2002).

[11] G. B. N. Robb, A. I. Best, J. K. Dix, P. R. White, T. G. Leighton, J. M. Bull, and A. Harris, "The measurement of the in situ compressional wave properties of marine sediments," Inf. Sci. (N.Y.) **32**, 484–496 (2007).

[12] T. J. Gorgas, R. H. Wilkens, S. S. Fu, L. N. Frazer, M. D. Richardson, K. B. Briggs, and H. Lee, "In situ acoustic and laboratory ultrasonic sound speed and attenuation measured in heterogeneous soft seabed sediments: Eel River shelf, California," Mar. Geol. **182**, 103–119 (2002).

[13] E. L. Hamilton, G. Shumway, H. W. Menard, and C. J. Shipek, "Acoustic and other physical properties of shallow-water sediments off San Diego," J. Acoust. Soc. Am. **28**, 1–15 (1956).

[14] E. I. Thorsos, K. L. Williams, N. P. Chotiros, J. T. Christoff, K. W. Commander, C. F. Greenlaw, D. V. Holliday, D. R. Jackson, J. I. Lopes, D. E. McGehee, J. E. Piper, M. D. Richardson, and D. Tang, "An overview of SAX99: Acoustic measurements," IEEE J. Ocean. Eng. **26**, 4–25 (2001).

[15] M. A. Zimmer, L. D. Bibee, and M. D. Richardson, "Acoustic sound speed and attenuation measurements in seafloor sands at frequencies from 1 to 400 kHz," in Proceedings to the International Conference on Underwater Acoustic Measurements: Technologies and Results, Heraklion, Crete (2005), pp. 327–334.

[16] J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Somerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).

[17] W. S. Burdic, *Underwater Acoustic Systems Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1991), Chap. 3.

[18] L. D. Luker and A. L. Van Buren, "Phase calibration of hydrophones," J. Acoust. Soc. Am. **70**, 516–519 (1981).

[19] IEC 60565:2006, *Underwater Acoustics—Hydrophones—Calibration in the Frequency Range* 0.01 Hz to 1 MHz (International Electrotechnical Commission, Geneva, Switzerland, 2006).

[20] M. Schulkin and H. W. Marsh, "Sound absorption in seawater," J. Acoust. Soc. Am. **34**, 864–865 (1962).

[21] E. L. Hamilton, "Geoacoustic modelling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

[22] G. B. N. Robb, A. I. Best, J. K. Dix, J. M. Bull, T. G. Leighton, and P. R. White, "The frequency dependence of compressional wave velocity and attenuation coefficient of intertidal marine sediments," J. Acoust. Soc. Am. **120**, 2526–2537 (2006).

[23] A. I. Best and D. G. Gunn, "Calibration of multi-sensor core logger measurements for marine sediment acoustic impedance studies," Mar. Geol. **160**, 137–146 (1999).

[24] G. M. Friedman and J. E. Sanders, *Principles of Sedimentology*, (Wiley, New York, 1978).

[25] R. Micheletti, "Phase angle measurement between two sinusoidal signals," IEEE Trans. Instrum. Meas. **40**, 6–9 (1991).

[26] S. P. Robinson, P. M. Harris, J. Ablitt, G. Hayman, A. Thompson, A. L. Van Buren, J. F. Zalesak, R. M. Drake, A. E. Isaev, A. M. Enyakov, C. Purcell, H. Zhu, Y. Wang, Y. Zhang, P. Botha, and D. Krüger, "An international key comparison of free-field hydrophone calibrations in the frequency range 1 kHz to 500 kHz," J. Acoust. Soc. Am. **120**, 1366–1373 (2006).

[27] *ISO/IEC Guide 98-3:2008 Uncertainty of Measurement—Part 3: Guide to the Expression of Uncertainty in Measurement* (GUM:1995) (International Organization for Standardization, Geneva, Switzerland, 2008).

[28] S. G. Kargl, "Mechanical loading of a spherical hydrophone embedded in a sediment," J. Acoust. Soc. Am. **113**, 2300 (2003).

Material properties from acoustic radiation force step response

Marko Orescanin
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Kathleen S. Toohey
Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Michael F. Insana
Department of Bioengineering, Department of Electrical and Computer Engineering, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 6 October 2008; revised 2 February 2009; accepted 24 February 2009)

An ultrasonic technique for estimating viscoelastic properties of hydrogels, including engineered biological tissues, is being developed. An acoustic radiation force is applied to deform the gel locally while Doppler pulses track the induced movement. The system efficiently couples radiation force to the medium through an embedded scattering sphere. A single-element, spherically-focused, circular piston element transmits a continuous-wave burst to suddenly apply and remove a radiation force to the sphere. Simultaneously, a linear array and spectral Doppler technique are applied to track the position of the sphere over time. The complex shear modulus of the gel was estimated by applying a harmonic oscillator model to measurements of time-varying sphere displacement. Assuming that the stress-strain response of the surrounding gel is linear, this model yields an impulse response function for the gel system that may be used to estimate material properties for other load functions. The method is designed to explore the force-frequency landscape of cell-matrix viscoelasticity. Reported measurements of the shear modulus of gelatin gels at two concentrations are in close agreement with independent rheometer measurements of the same gels. Accurate modulus measurements require that the rate of Doppler-pulse transmission be matched to a priori estimates of gel properties. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106129]

PACS number(s): 43.35.Mr, 43.25.Qp, 43.80.Ev [PEB] Pages: 2928–2936

I. INTRODUCTION

Elasticity imaging is a promising diagnostic technique for discriminating between benign and malignant breast lesions.[1–7] Its diagnostic value stems from the important role of the cellular mechanoenvironment in regulating tumor growth[8] and from the large tumor contrast observed for various mechanical properties.[9] Biological sources of elastic strain contrast in mammary tissues include edema, hyperplasia, acidosis, fibrosis, desmoplasia, and inflammatory responses characteristic of the reaction of breast stroma to cancer cells. Physical sources of elasticity contrast are related to the spatial variations in flow velocity of fluids through the extracellular matrix (poroelasticity) and the rate at which the matrix itself mechanically relaxes (viscoelastic) in response to applied forces.[10] Classification of nonpalpable, isoelastic lesions is possible by imaging time-varying strain features.[6,7]

Despite early clinical successes, the visibility of lesions in elasticity imaging can vary widely. We hypothesize that some of the clinical variability may be reduced by improving our understanding of elasticity imaging contrast mechanisms and adapting the imaging techniques accordingly. Our approach to mechanism discovery is to establish relationships between the physical and biological sources of contrast listed above across the spectrum of force frequencies used by the various approaches to elasticity imaging.

Quasi-static elasticity imaging methods apply a ramp force suddenly and hold it constant while strain is imaged over time.[11] The methods is "quasi-static" for patient imaging because modest forces (1–5 N) are manually applied slowly (~1-s ramp on) to the breast surface through the ultrasound transducer. Quasi-static methods interrogate tissues at a very low applied-load frequency bandwidth that is bounded from above at approximately 1 Hz and from below at 0.01 Hz depending on the total acquisition time for the strain image recording sequence.[12] At the other load bandwidth extreme are acoustic radiation force imaging methods.[5] A focused push-pulse applies a weak impulse force deep in tissue for about 1 ms after which displacements are imaged in time as the tissue relaxes. This load bandwidth is nominally 100–1000 Hz depending on experimental details. Other acoustic-based approaches, including ultrasound-stimulated vibro-acoustic spectography,[2] shear wave elasticity imaging,[3] and harmonic motion imaging,[4] probe load bandwidths somewhere between these two extremes.

It is difficult to design studies to discover disease-specific sources of elasticity contrast for any of these imaging techniques. *In vivo* breast tissue properties are spatially heterogeneous, frequently anisotropic, and have poorly defined boundaries. Hence complex internal stress fields are common, making it difficult to even rigorously define a modulus. Excised tissue samples are nonrepresentative be-

0001-4966/2009/125(5)/2928/9/$25.00 © 2009 Acoustical Society of America

cause of changes caused by the lack of perfusion, decomposition, or use of fixatives. Gelatin hydrogels are structurally simpler, homogeneous, and able to mimic some properties of breast stroma as required for imaging system development.[12] However, hydrogels do not mimic cell-driven dynamic properties normally associated with malignant progression or responses to treatment; many of these features are assumed associated with tumor contrast. Rodent models of mammary cancer can accurately represent genomic, biochemical, metabolic, and some perfusion aspects of tumor physiology,[13] but are less representative of the macrostructures of human breast tumors that strongly influence mechanical behavior.

We are exploring the use of three dimensional (3D) cell cultures.[14,15] While they suffer many of the same problems experienced in excised tissue measurements, they have the advantage of containing living mammary cells embedded in hydrogel volumes. The cells can be biochemically or mechanically stimulated and then observed under sterile conditions. Cell cultures do not simulate the tumor macroanatomy but they can mimic the responses of tumor-cell clusters to their microenvironment. Gels combine geometric simplicity for ease of mechanical measurements with dynamic cellular processes that can be independently verified via optical microscopy.

Many biological tissues and all of the gels we considered are biphasic polymers, which means their mechanical properties are determined by a polymeric matrix (solid phase) embedded in a liquid (fluid phase). The mechanical responses of multiphasic polymers depend significantly on the rate at which force is applied. For example, the complex shear modulus is known to vary widely with force frequency in lightly-cross-linked amorphous polymers,[16] breast tissues,[6] and even within individual cells of the body.[17]

Our research goal is to develop a radiation force technique for estimating shear modulus and shear viscosity of gel types often used in 3D cell cultures and engineered tissues. These measurements will eventually be made over the bandwidth of force frequencies used in various elasticity imaging techniques. This report focuses on the application of Doppler measurements to describe transient dynamic responses of gelatin gels to a step change in radiation force. Particle velocity estimates are related to modulus and viscosity through a second-order rheological model. The results provide an estimate of the impulse response function of shear wave imaging.

II. METHODS

The goal of the proposed method is to remotely and quantitatively estimate material properties using acoustic radiation force. Acoustic pressure fields exert localized forces with a magnitude that depends on the energy density of the field and the scattering and absorption properties of target media. Gelatin gels are used in this study that describes the measurement system and rheological models applied for material property estimation.

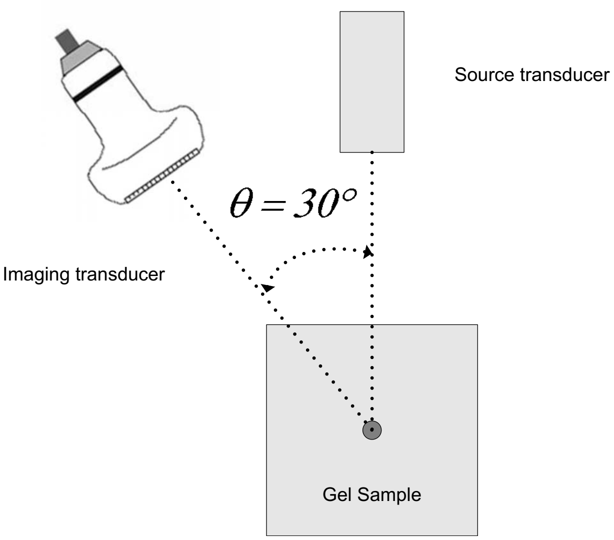


FIG. 1. Diagram of the experiment to measure viscoelastic properties of gel samples. Acoustic force applied by a source transducer displaces a sphere embedded in the gel. An imaging transducer tracks the induced motion of the sphere.

A. Acoustic radiation force

Acoustic radiation force is generated when momentum of the acoustic wave is transferred to the propagation medium via attenuation and scattering interactions. We study low-attenuation gels to which a strongly scattering sphere is embedded. Scattering from the sphere efficiently couples the acoustic field to the gel to induce forces that measurably deform gels at relatively low acoustic intensity.

For sphere diameters small compared with the beam width (1.5 and 6 mm, respectively), we can assume local plane waves and the time-averaged force on the scattering sphere is approximately[18]

$$F = \pi a^2 Y \bar{E}. \tag{1}$$

The quantity a is the sphere radius and Y is the radiation force function as determined by the mechanical properties and geometry of the sphere and the surrounding gel. $\bar{E}$ is the average energy density of the incident field. The time average is over several cycles of the carrier frequency (microseconds) but typically varies over the period of the amplitude modulation (milliseconds). We measured the acoustic radiation force on a steel sphere suspended in water and found that it agreed with the prediction of Eq. (1) within experimental error.[19]

B. Source transducer

Figure 1 illustrates the experiment depicting a gel sample containing a stainless steel sphere. Force is applied by the acoustic field of a circular, 19-mm-diameter, $f/4$, lead zirconate titanate (PZT) element that is transmitting sine-wave bursts at the resonant frequency of 1 MHz. Bursts 200 ms in duration were transmitted every 2 s to induce a maximum sphere displacement >20 μm for gels containing 3% w/w gelatin. The pressure field from the source transducer was measured in water using a recently calibrated polyvanylidene fluoride (PVDF) membrane hydrophone (GEC-Research Ltd., Marconi Research Center, Chelmsford, UK). The results were used to estimate a primary radiation force at 60 μN.[19] The error on the force estimate was approximately 16% of the mean value and was determined primarily by the uncertainty in pressure estimates. The sphere was positioned on the beam axis at the 76-mm radius of curvature of the source. The location of the sphere was tracked in time by measuring and integrating the instantaneous sphere velocity.

C. Sphere velocity and displacement estimation

A Siemens Sono-line Antares system was used to estimate sphere velocity via pulsed-Doppler methods. A VF5-10 linear array transducer was driven by 1-cycle, 7.27-MHz voltage pulses to transmit nominally 2.5-cycle, 7-MHz acoustic pulses. Doppler-pulse transmission was repeated for a fixed beam-axis position on the time interval $T_s = 76.8\ \mu$s. rf echo waveforms were sampled at 40 Msamples/s using the ultrasound research interface of the Antares system[20] and stored for offline processing. The axes of the source transducer and linear array intersected at the 1.5-mm-diameter steel sphere, and the beam axes were separated by $\theta = 30°$, as illustrated in Fig. 1.

The demodulated complex envelope $V[n,m']$ was computed for each Doppler echo waveform. The sample index $1 \le n \le N$ counts echo samples within an echo waveform in what is commonly referred to as "fast-time." The index $1 \le m' \le M'$ counts the waveforms in "slow-time."

We compute the lag-one correlation function estimate between adjacent pairs of echo waveforms using

$$\hat{\phi}[n,m] = V^*[n,2m-1]V[n,2m], \quad m' = 2m-1. \tag{2}$$

The change in index from m' to $m(1 \le m \le M)$ avoids counting by 2. The estimate of instantaneous sphere velocity $\hat{v}_s$ from complex correlation estimates is[21,22]

$$\hat{v}_s[m] = \left(\frac{-c}{4\pi f_c T_s \cos\theta} \right) \frac{1}{N_0} \sum_{n=n_0}^{n_0+N_0-1} \arg(\hat{\phi}[n,m]). \tag{3}$$

c is the compressional-wave speed of sound in the gel medium (1.5 mm/μs), n_0 marks the first fast-time sample in the region of interest near the sphere-echo peak, N_0 is the number of fast-time samples in the region of interest, and arg(·) indicates the phase angle obtained from the arctangent of the ratio of imaginary to real parts of the argument. High-pass filtering in slow-time, which is frequently used in blood flow measurements (wall filter), was unnecessary because scattering from the gel was negligible compared to the sphere.

Finally, sphere displacement is estimated by integrating velocity estimates $\hat{x}(t) \equiv \int_0^t \hat{v}_s(t')dt'$, where $t' = 2mT_s$. Integration was performed numerically using a cumulative trapezoidal scheme.[23]

D. Hydrogel sample construction

Gelatin gel samples (250 bloom strength, type B, Rousselot, Buenos Aires, Argentina) were constructed to test acoustic radiation force measurements of shear modulus and viscosity. Gelatin powder and distilled water are heated in a water bath at a temperature between 65 and 68 °C for 1 h and periodically stirred. When the sample is cooled to 50 °C, 0.1% by weight formaldehyde is added and thoroughly mixed. Molten gelatin is poured into a cylindrical sample mold (diameter 7.5 cm, height 5.5 cm). Two or three stainless steel spheres 1.5 mm in diameter are widely dispersed within the cooling gel just prior to gelation. Samples with 3% or 4% w/w gelatin concentrations are homogeneous except for the isolated spheres that are separated by at least 1.5 cm.

Narrowband through-transmission measurements of compression-wave speed and attenuation coefficient[24] were made on samples without steel spheres and with 4% gelatin concentration. Measurements made at 21 °C in degassed water were first calibrated using a castor oil sample. Two phantoms were measured every 0.5 MHz between 7 and 12 MHz. The slope of the attenuation coefficient as a function of frequency was estimated to be 0.027 ± 0.003 dB mm^{-1} MHz^{-1}. Using no alcohol in the sample, the average speed of compressional waves was $c = 1506 \pm 0.34$ m s^{-1} over the frequency range of the measurement.

The material properties of the gelatin gels were verified independently through oscillatory rheometer experiments. Parallel plate shear experiments were conducted on an AR-G2 rheometer (TA Instruments, New Castle, DE). Circular specimens, 25 mm diameter and 2–4 mm high, were molded from the same gelatin used to make the large samples containing spheres. After 1 day of gelation, the specimens were removed from the molds and bonded to parallel plate fixtures using cyanoacrylate (Rawn America, Spooner, WI). 5% strain was applied over a frequency range from 0.1 to 10 Hz with ten sample points per decade of frequency. For both concentrations of gelatin, the measured storage modulus was averaged over the test range giving 321 ± 14 and 640 ± 17 Pa for 3% and 4% gelatin concentrations.

E. Modeling

The rheological behavior of hydrogels on a scale larger than the ultrasonic wavelength may be described as that of a continuum.[16] We propose to model the displacement $x(t)$ of a sphere embedded in gelatin as a simple harmonic oscillator,[25]

$$M_t \frac{d^2x(t)}{dt^2} + R\frac{dx(t)}{dt} + kx(t) = F(t). \tag{4}$$

$F(t)$ is the driving force, M_t is the total mass on which the force acts, R is a damping constant related to the mechanical impedance of the gel (see Appendix), and k is an elastic constant. Because the uniaxial load is applied along the source transducer beam axis and movement of the sphere is in the same direction, x and F are the axial components of the corresponding vectors. For a step change in force over time from a constant value to zero, $F(t) = F_0(1 - \text{step}(t))$, the homogeneous solution for displacement obtained from Eq. (4) has the form

$$x(t) = \begin{cases} x_0, & t \leq 0 \\ Ae^{-\alpha t}\cos(\omega_d t + \varphi), & t > 0. \end{cases} \quad (5)$$

A is the displacement amplitude, $\alpha = R/2M_t$, $\omega_d = \sqrt{\omega_0^2 - \alpha^2}$ is the resonant frequency with damping, and $\omega_0 = \sqrt{k/M_t}$ is the resonant frequency without damping. From the initial conditions, $A = x_0/\cos\varphi$ and $\tan\varphi = -\alpha/\omega_d$.

It is important to include the surrounding gel in estimating the dynamic inertia of the system.[26] The total mass that reacts to the radiation force is $M_t = M_s + M_a$, where M_s is the mass of the sphere and $M_a = \frac{2}{3}\pi a^3 \rho_g$ is the added mass of surrounding gel, where a is the sphere radius and ρ_g is the density of the gel. The next step is to relate the constants k and R to rheological parameters μ and η.

The viscous drag force F_d experienced by a 1.5-mm sphere as it moves through incompressible and viscous gel at velocities <10 mm/s has a Reynolds number on the order of 0.02. Consequently Eq. (4) gives the linear approximation $F_d(t) = -Rv_s(t)$, and the classic Stokes equation for R is[27]

$$R = 6\pi a \eta, \quad (6)$$

where the parameter η has the SI units Pa s. In the Appendix, we show that R is the mechanical resistance or the real part of the impedance. This implies that η may be interpreted as the shear damping parameter, which within the frequency range of the experiments is defined as $\mu_2 + \mu_1 a/c_s$, where μ_2 is shear viscosity or the imaginary part of the complex shear modulus μ', μ_1 is the real part of the complex shear modulus μ', and c_s is the shear wave speed.

Ilinskii *et al.*[28] applied an analysis parallel to Stokes derivation to show that the elastic constant in the restoring force equation, $F_r(t) = -kx(t)$, is

$$k = 6\pi a \mu, \quad (7)$$

where μ, with the SI units Pa, approximates the shear elasticity, μ_1 (see Appendix for details).

Combining Eqs. (5)–(7) sphere displacement is modeled in terms of shear elasticity and shear damping parameter. The approach is to measure M_t and a independently and then numerically fit normalized displacement estimates $\hat{x}'(t) = \hat{x}(t)/\hat{x}_0$ to model values $x'(t) = x(t)/x_0$ obtained from Eqs. (5)–(7) with μ and η as free parameters. Normalization scales and shifts the response so that displacements have values between 0 and 1. Thus μ and η are estimated without knowledge of the applied force magnitude F_0.

III. RESULTS

We verified the proposed model and assumptions by conducting radiation force experiments. The 1-MHz source transducer transmitted 200-ms voltage bursts with the same amplitude in each experiment. Originally at rest, the sphere was suddenly displaced away from the transducer by the pulse a maximum distance x_0 (see Fig. 2) before being released to return to its original location. The imaging probe measuring the sphere velocity was transmitting and receiving Doppler pulses during the entire process.

The rf echo waveform in Fig. 3 shows that each Doppler pulse causes the steel sphere to ring. Because the echo signal-to-noise ratio for tracking sphere velocity was very high, Doppler-pulse durations were set to 2.5 cycles to temporally resolve the first echo from subsequent ringing echoes. Echo phase is estimated near the peak of the first echo in Fig. 3.

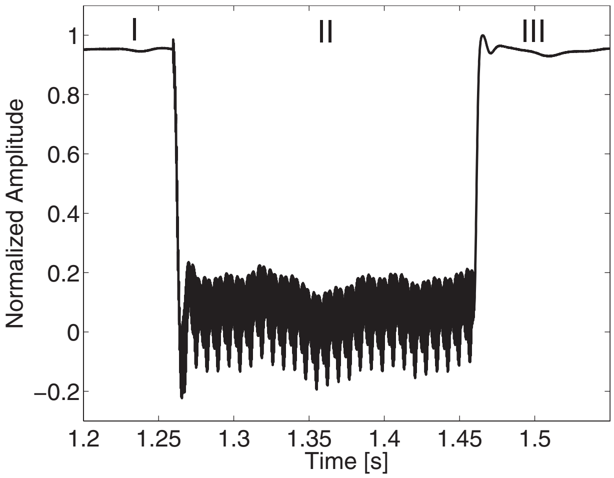


FIG. 2. Measurement of sphere displacement versus slow-time as determined from the change in Doppler echo phase. The sphere is embedded in a 3% gelatin gel. Region I is a time period before radiation force is applied and the sphere is at rest. Region II is a time period that the source transducer is transmitting a 1-MHz cw burst and the sphere is displaced away from the source. Oscillations indicate cross talk between the source and Doppler probes. Region III is the time period after the source is turned off and the sphere returns to its original position.

From the data of Fig. 2, we can illustrate the process for a specific experiment. The spectral Doppler acquisition was initiated (region I). After approximately 1.26 s, the source transducer was turned on for 200 ms (region II). The phase of the Doppler echo from the sphere changed as the sphere was displaced by the acoustic force. On the time axis of the figure at 1.46 s, the source transducer is turned off [this time is set to $t=0$ in Eq. (5)] and the sphere returns to the equilibrium position with the response of a slightly underdamped oscillator. We analyzed sphere displacement data as the

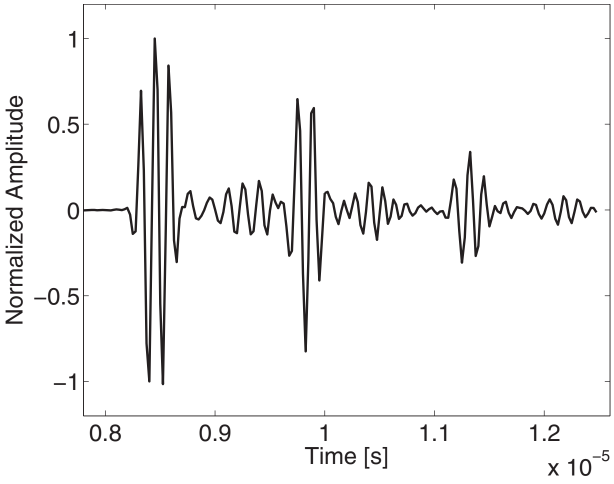


FIG. 3. Example of a broadband Doppler echo waveform versus fast-time. A single transmitted pulse is reflected from a steel sphere in 3% gelatin gel. Multiple echoes indicate ringing of the sphere.

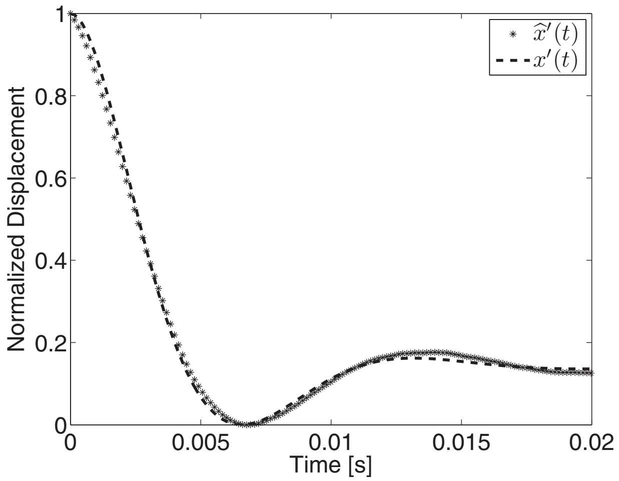


FIG. 4. Normalized sphere displacement measurements $\hat{x}'(t)$ from region III in Fig. 2 are compared with the model equation $x'(t)$ from Eq. (5). The minimum least-squares fit (r^2=0.996) was obtained for 3% gelatin gel aged 1 day to find μ=317 Pa and η=0.57 Pa s.

source pulse was turned off rather than turned on to avoid cross talk between the source and Doppler probes as seen in Fig. 2, region II.

Figure 4 is an example of a comparison between a measured displacement time series $\hat{x}'[m]$ and samples from the best-fit model $x'[m]$ as a function of slow-time, $2mT_s$. For an M-point displacement time series with normally distributed random error, the material parameters μ and η are chosen to give the smallest residual sum of squares,[29]

$$r^2 = 1 - \frac{\sum_{m=1}^{M}(\hat{x}'[m]-x'[m])^2}{\sum_{m=1}^{M}(\hat{x}'[m]-\bar{x}')^2}, \quad \bar{x}' = \frac{1}{M}\sum_{m=1}^{M}\hat{x}'[m]. \tag{8}$$

r^2 is bounded from above by 1 (perfect agreement between data and model) and from below by zero, although it can be negative.

For small displacements, there is close agreement between measurements and the model, suggesting that gel deformation is linear as required by Eq. (5). Furthermore, if the normalized displacement is time invariant, then we may express the model as a linear system

$$x(t) = \int_{-\infty}^{\infty} dt h(t-t')(1-\mathrm{step}(t'))$$

with impulse response

$$h(t) = -\frac{dx}{dt} = Ae^{-\alpha t}(\alpha\cos(\omega_d t+\varphi)+\omega_d\sin(\omega_d t+\varphi)). \tag{9}$$

Equation (9) enables prediction of the displacement for any time-varying applied load for which the gel responds linearly.

Measurements of μ and η for 3% and 4% gelatin gels conducted over 4 days are presented in Figs. 5 and 6, respectively. Without adding a strong chemical cross linker, gelatin gels slowly increase their cross-link density, and thus gels continue to stiffen over days. Although gelatin gel responses are not strictly time invariant, the change in the impulse response is negligible over the duration of any experiment. Estimated values of the modulus and shear damping for gels with C=4% gelatin concentration are larger than that at 3% for each day of the study. Gilsenan and Ross-Murphy[30] found that the shear modulus varies with the square of gelatin concentration, $\mu \propto C^2$, between 1% and 5%. Our data in Fig. 5 give a concentration dependence of $C^{2.7}$ on day 1 and $C^{2.4}$ on day 3.

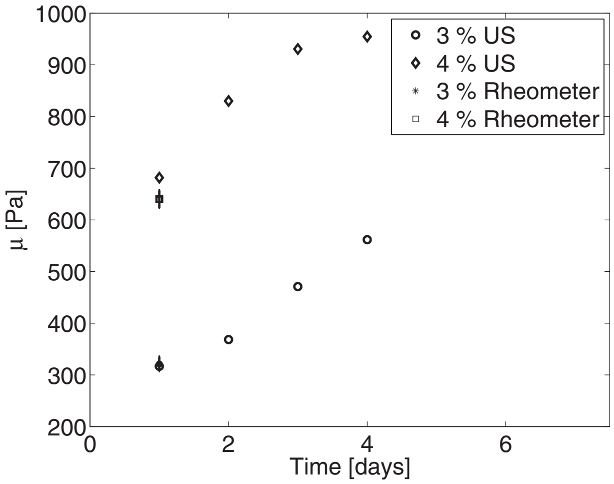


FIG. 5. Shear modulus as a function of gel age for 3% and 4% gelatin concentrations. Rheometer estimates of μ made on day 1 are also shown with error bars indicating ±1 sd.

As indicated in Fig. 5, rheometer measurements of the shear storage modulus were also made on day 1 for both gelatin concentrations. Five rheometer measurements were made on five different 3% gelatin samples to yield a mean and standard deviation of μ_r=321 ± 14 Pa. The comparable radiation force estimate was 317 Pa. Three measurements were made on three different 4% samples to find μ_r =640 ± 17 Pa. The comparable radiation force estimate is 681 Pa. Considering the rheometer measurements as a standard, radiation force estimates of shear modulus are accurate well within the observed day-to-day change in mean values. We were unable to obtain independent estimates of shear viscosity for the gels.

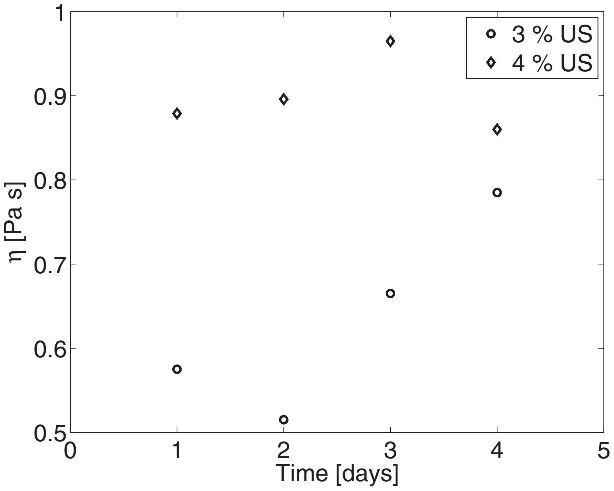


FIG. 6. Shear damping parameter as a function of gel age for 3% and 4% gelatin concentrations.

Radiation force measurements may also be used to estimate the shear speed c_s and shear viscosity μ_2; both are defined in the Appendix. At the end of the Appendix, we show that $\mu_1 \simeq \mu$, $\mu_2 = \eta - \mu_1 a/c_s$, and at low force frequencies where $\omega^2 \mu_2^2 \ll \mu_1^2$ we obtain the elastic result, $c_s \simeq \sqrt{\mu_1/\rho}$. Applying the 3% gelatin sample results at 24 h following gelation, $\mu = 317$ Pa and $\eta = 0.57$ Pa s, we estimate $c_s = 0.56$ m s^{-1} and $\mu_2 = 0.14$ Pa s. Our estimates are comparable to those reported by others using similar acoustic radiation force techniques.[31,32]

Intra-sample precision variability was estimated by measuring μ multiple times for a single sphere in one gelatin sample. The percent standard deviation was found to be approximately 3.5% of the mean, for example, $\mu = 317 \pm 11$ Pa. Boundary variability, i.e., proximity of each steel sphere to the gel sample surfaces, was examined by averaging μ measurements for different spheres placed in one gelatin sample. That standard deviation was approximately 7% of the mean. Inter-sample variability for μ was larger, 20% of the mean, primarily because of differences in gel preparation. The relatively small random experimental error is a consequence of the high echo signal-to-noise ratio.

IV. DISCUSSION

Mechanical parameter values are primary factors determining the ultrasonic sampling rate for pulsed-Doppler velocity estimation. Discussion near Eqs. (5)–(7) explains that the time-varying displacement amplitude, the frequency, and the phase are functions of μ and η. Estimation accuracy and precision will vary with the sampling rate depending on the bandwidth of the displacement spectrum. For linear gels, the displacement spectrum is the spectrum of the applied force filtered by the mechanical system response of the gel, $H(\omega;\mu,\eta)$. $H(\omega;\mu,\eta)$ is the temporal Fourier transform of Eq. (9) parametrized by the material properties.

The model spectrum of interest is the squared magnitude of the temporal Fourier transform of $x'(t)$ from Eq. (5). It has the Lorentz form

$$|X'(\omega)|^2 = \frac{1}{\alpha^2 + (\omega - \omega_d)^2}.$$

The 3-, 6-, and 20-dB bandwidths of the displacement spectrum are, respectively, $\Delta\omega = R/M_t$, $\sqrt{3}R/M_t$, and $\sqrt{99}R/M_t$. Therefore the upper limit on angular frequency is

$$\omega_{\max} = \omega_d + \Delta\omega/2 = \sqrt{\omega_0^2 - \alpha^2} + B\alpha, \qquad (10)$$

where $B = 1$, $\sqrt{3}$, or $\sqrt{99}$ for the 3-, 6-, or 20-dB bandwidths.

To illustrate, Fig. 7 displays the displacement spectrum corresponding to the parameters for measurements on day 1 for 3% gelatin-concentration samples. The highest frequency in the 3-dB bandwidth is found from Eq. (10) to be $f_{\max} = \omega_{\max}/2\pi = 120$ Hz. The highest frequencies in the 6- and 20-dB bandwidths are 152 and 510 Hz, respectively.

The sampling theorem for bandlimited signals states that the minimum sampling rate needed to avoid aliasing is twice the value of the maximum frequency in the bandwidth. However, we must further increase the rate by the number of pulses in the velocity estimator ensemble, M_e. That is,

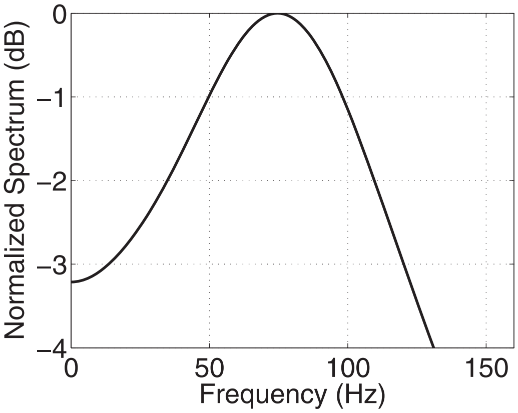


FIG. 7. Power spectrum of displacement for model parameters from 3% gelatin 24 h after gelation: $\mu = 317$ Pa and $\eta = 0.57$ Pa s. Characteristic parameters are the natural frequency $w_d/2\pi = 76.1$ Hz, half bandwidth $\Delta w/4\pi = 44$ Hz, and maximum frequency at the 3-dB limit $f_{\max} = 120.1$ Hz.

$$\begin{aligned} f_s &\geqslant 2M_e f_{\max} \\ &= \frac{M_e}{\pi}\left(\sqrt{\omega_0^2 - \alpha^2} + B\alpha\right) \\ &= \frac{M_e}{\pi}\left(\sqrt{\frac{6\pi a\mu}{M_t} - \left(\frac{3\pi a\eta}{M_t}\right)^2} + \frac{3\pi B a\eta}{M_t}\right). \end{aligned} \qquad (11)$$

For the experiments described in the previous paragraph, where we adopt the 6-dB bandwidth limit and $M_e = 2$, the pulse-repetition frequency (PRF=f_s) must exceed 608 Hz to avoid aliasing.

To decide on an acceptable lower bound on the sampling frequency, we oversampled the Doppler measurements at $f_s = 13$ kHz. We then incrementally downsampled this waveform sequence, being careful to apply the appropriate low-pass anti-aliasing filter as the Nyquist frequency changed, before processing. We thus obtained μ and η estimates as a function of f_s. We observed that a 15-dB bandwidth ($B = \sqrt{31}$) was sufficient to eliminate estimation errors within the intra-sample random error range of 7%. If the echo signal-to-noise ratio was reduced, for example, in stiff gels where sphere displacement is small or for low-scattering spheres, M_e could be increased to compensate as given by Eq. (11). There is also a tradeoff between time resolution for velocity estimates and distance to the target, in our case the sphere depth. Increasing f_s reduces the depth for the maximum unambiguous range to $c/2f_s$.[21]

We have evaluated Eq. (11) for a range of μ and η values estimated for gels and for the typical experimental parameters $M_e = 2$, $M_t = 14.7$ mg, and $B = \sqrt{31}$ (15-dB bandwidth). The corresponding minimum Doppler-pulse sampling rates are plotted in Fig. 8. It is important to point out that Eq. (11) is valid only for the Lorentz spectrum characteristic of the Kelvin–Voigt model, with total mass M_t as defined above. Changing the model to, for example, a three-element Zener model,[33] would require a new analysis to establish the minimum sampling frequency.

Quick estimates of μ and η may be made for a well-calibrated experimental system. If M_t and a are known, then η can be found directly from the 3-dB bandwidth of the step

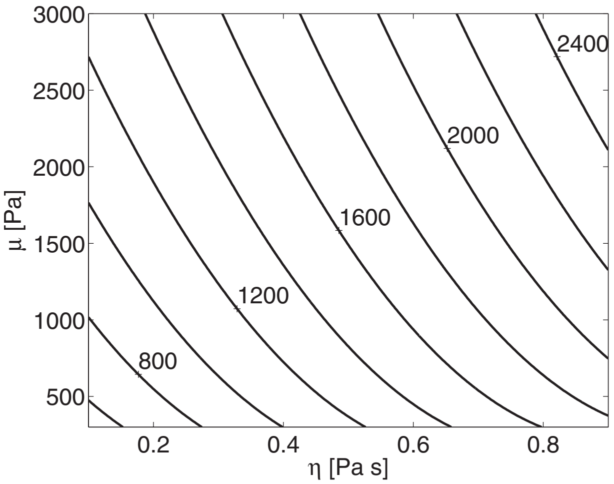


FIG. 8. Contour plot of the minimum sample frequency (i.e., PRF) in hertz from Eq. (11) as required to estimate μ and η as a function of these same material properties. We used a fixed ensemble size $M_e=2$, $B=\sqrt{31}$ (15-dB bandwidth), and $M_t=14.7$ mg. For example, for $\mu=1.5$ kPa and $\eta=0.5$ Pa s, $f_s \gtrsim 1.6$ kHz.

response, $\eta=M_t\Delta\omega_{3\text{ dB}}/6\pi a$. Applying this result and an estimate of the spectral peak to the expression for resonant frequency ω_d, we can estimate μ.

V. CONCLUSION

A damped harmonic oscillator model accurately predicts the movement of a hard sphere embedded in a congealed hydrogel to a sudden change in acoustic radiation force. This result suggests that the gel responds linearly to the force. We show how to relate parameters of the harmonic oscillator to the mechanical impedance of the system and material parameters. We estimated the coefficients of the complex shear modulus (shear elasticity and viscosity) with 7% intrasample random experimental error by interpreting model parameters in terms of rheological elements. The radiation force estimates of modulus at two gel concentrations closely agree with independent measurements of the gels using a rheometer. This simple but accurate technique is designed to measure viscoelastic properties of 3D cell cultures remotely to maintain sterile conditions.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Rousselot Inc., Dubuque, IA, for providing a generous supply of gelatin in support of this research, and the Autonomic Materials Laboratory at the Beckman Institute for use of a rheometer. This work was supported by the National Cancer Institute under Award No. R01 CA082497.

APPENDIX: MATERIAL PROPERTIES FROM MECHANICAL IMPEDANCE

The viscoelastic material properties of the gel medium surrounding a rigid sphere are frequently characterized by the mechanical impedance, Z, which relates, in three spatial dimensions, $\mathbf{y}=y_1,y_2,y_3$, and time, t, the resistance force of the sphere to motion, $\mathbf{F}(t,\mathbf{y})$, and the resulting sphere velocity, $\mathbf{v}_s(t,\mathbf{y})$, via the Ohm's law-like expression[34]

$$\mathbf{F}(t,\mathbf{y}) = -Z \times \mathbf{v}_s(t,\mathbf{y}). \quad \text{(A1)}$$

Force and velocity are vector quantities, impedance is a scalar, and all three are complex quantities in this expression. Stationary harmonic forces at radial frequency ω applied along the y_1 axis and of the form $F(t)=F_{\omega,1}\exp(-i\omega t)$ generate sphere velocities of the form $v_s(t)=v_{\omega,1}\exp(-i\omega t)$. In that situation, the material properties that influence Z are gel density $\rho \simeq 1$ g/cm^3, sphere radius $a=0.75$ mm, and the complex Lamè moduli $\mu'=\mu_1-i\omega\mu_2$ and $\lambda'=\lambda_1-i\omega\lambda_2$. μ_1 is shear elasticity, μ_2 is shear viscosity, λ_1 is volume compressibility, and λ_2 is volumetric viscosity. For the small forces used in our experiments, it is assumed that the gel responds linearly so the sphere velocity for an arbitrary time-varying force is a weighted linear superposition of velocities at each frequency in the force bandwidth. Further, it is assumed that the sphere is bound to the continuous, homogeneous, and isotropic gel.

Of course, the force and velocity vectors also vary spatially. For harmonic, compressional, plane waves traveling along the y_1 axis, $F(t,\mathbf{y})\hat{\mathbf{y}}_1=F_{\omega,1}\exp(ky_1-\omega t)$, the impedance is straightforward to find from Eq. (A1) and the one-dimensional wave equation, as shown in standard texts.[34]

The mechanical impedance Z for the case of an oscillating sphere was found by Oestreicher.[35] Making use of the spherical symmetry, he separately solved for the irrotational and incompressible components of the wave equation relating pressure and displacement in terms of spherical harmonics.[26] Integrating the pressure over the sphere to find force and for $v_s(t)=-i\omega x(t)$, he found

$$Z = -\frac{4}{3}\pi a^3\rho i\omega\left\{1-\left[\frac{1}{3}\left(1-\frac{3(1-ik_ca)}{k_c^2a^2}\right)^{-1} - \frac{2}{3}\left(1-\frac{3(1-ik_sa)}{k_s^2a^2}\right)^{-1}\right]^{-1}\right\}. \quad \text{(A2)}$$

The above expression is the form given by Norris [see Eq. (5) in Ref. 36]. In Eq. (A2), $k_c=(\rho\omega^2/(2\mu'+\lambda'))^{1/2}$ and $k_s=(\rho\omega^2/\mu')^{1/2}$ are, respectively, the compressional and shear complex wave numbers. The wave number $k_s=\omega/c_s+i\alpha_s$ may also be written as a function of the shear wave speed and shear wave attenuation constant, respectively,[37]

$$c_s = \omega/\Re\{k_s\} = \sqrt{\frac{2(\mu_1^2+\omega^2\mu_2^2)}{\rho(\mu_1+\sqrt{\mu_1^2+\omega^2\mu_2^2})}} \quad \text{(A3)}$$

and

$$\alpha_s = \Im\{k_s\} = \sqrt{\frac{\rho\omega^2(\sqrt{\mu_1^2+\omega^2\mu_2^2}-\mu_1)}{2(\mu_1^2+\omega^2\mu_2^2)}}. \quad \text{(A4)}$$

Oestreicher[35] commented that the number of constants in the Lamè moduli increases if time derivatives of order greater than 1 are required to model the data. For harmonic oscillations, the corresponding Lamè moduli will have added terms multiplied by increasing powers of $-i\omega$. Higher-order time derivatives generate frequency dependent Lamè moduli that appear experimentally as dispersion, i.e., frequency dependent wave speeds. It was shown experimentally[24] that gelatin is non-dispersive for compressional waves between 1

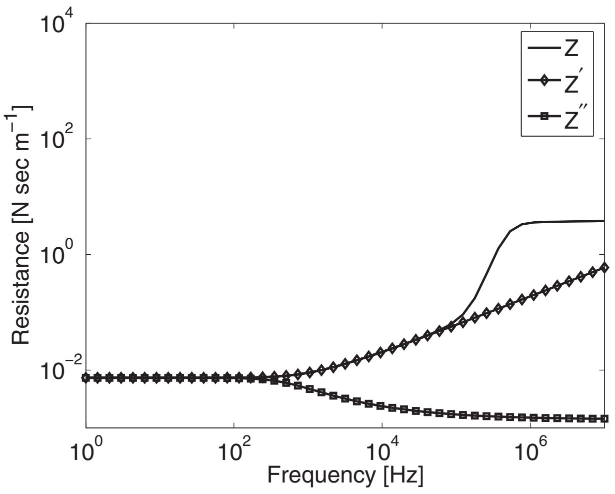


FIG. 9. Mechanical resistance (real part of impedance) for an oscillating sphere in 3% gelatin gel.

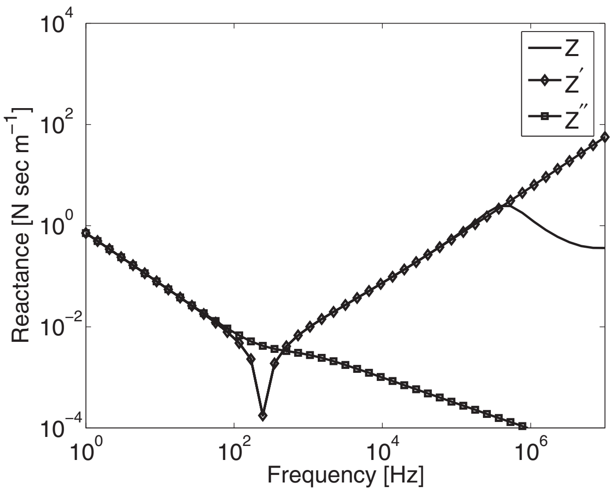


FIG. 10. Mechanical reactance (imaginary part of impedance) for an oscillating sphere in 3% gelatin gel.

and 10 MHz with and without particle scatterers. Applying Eq. (A3) and the values of μ_1 and μ_2 reported in Sec. III, it can be seen that c_s varies by less than 0.7% for clear gelatin gels at shear-wave frequencies less than 50 Hz. Consequently, it is reasonable to assume non-dispersive media for our low-frequency experiments.

In incompressible viscoelastic gels, the bulk modulus $\lambda + 2\mu/3$ becomes infinite while μ remains finite.[36] We measured $c_c = 1506$ m/s and $\mu_1 = 317$ Pa for clear 3% gelatin gels and adopt $\mu_2 = 0.1$ Pa s as Ilinskii *et al.*[28] Applying the expressions from the paragraph below Eq. (A2), we estimate that $c_s = 0.56$ m/s and $\lambda_1 = 2.25 \times 10^9$ Pa. Further, like Oestreicher,[35] we assume $\lambda_2 = 0$. Consequently, $k_c/k_s \ll 1$, and Eq. (A2) reduces to

$$Z' = -\frac{6\pi a\mu'}{i\omega}\left[\frac{k_s^2a^2}{9} - (1 - ik_s a)\right] \qquad \text{(A5)}$$

provided the sphere remains bound to the gelatin.[36] Expanding Eq. (A5) using $\mu' = \mu_1 - i\omega\mu_2$, we find

$$Z' = -6\pi a\left[\mu_2\left(1 - \frac{k_s^2a^2}{9}\right) + \frac{\mu_1}{\omega}k_s a\right] - i6\pi a\left[\frac{\mu_1}{\omega}\left(1 - \frac{k_s^2a^2}{9}\right) - \mu_2 k_s a\right]. \qquad \text{(A6)}$$

Noting that $\mu_2/\mu_1 \ll 1$ and $a = 7.5 \times 10^{-4}$, we neglect all terms $\mathcal{O}(a^3)$ and $\mathcal{O}(\mu_2 a^2)$ to find

$$Z'' \simeq -6\pi a\left(\mu_2 + \frac{\mu_1 k_s a}{\omega} - i\frac{\mu_1}{\omega}\right). \qquad \text{(A7)}$$

Finally, expanding k_s as a function of c_s and α_s we can rewrite Eq. (A7) as

$$Z'' = -6\pi a\left(\mu_2 + \frac{\mu a}{c_s} - i\frac{\mu_1}{\omega}(1 + \alpha_s a)\right). \qquad \text{(A8)}$$

Impedance, Z, and its approximations, Z' and Z'', were evaluated numerically using values for constants listed above. The real parts are plotted in Fig. 9 and the imaginary parts in Fig. 10. There is no significant difference among the three expressions provided $\omega/2\pi < 100$ Hz, where we are free to adopt Eq. (A8). The damping constant, R from Eq. (4), corresponds to the real part of the mechanical impedance, Z''. Comparing R in Eq. (6) with $\Re\{Z''\}$ in Eq. (A8), we find $\eta = \mu_2 + \mu_1 a/c_s$. Also, since $v_s(t) = -i\omega x(t)$, comparing μ from Eq. (7) with $\Im\{Z''\}$ in Eq. (A8) yields $\mu = \mu_1(1 + \alpha_s a) \simeq \mu_1$ for our experimental conditions.

[1] A. Itoh, E. Ueno, E. Tohno, H. Kamma, H. Takahashi, T. Shiina, M. Yamakawa, and T. Matsumur, "Breast disease: Clinical application of US elastography for diagnosis," Radiology **239**, 341–350 (2006).

[2] M. Fatemi and J. F. Greenleaf, "Ultrasound-stimulated vibro-acoustic spectrography," Science **280**, 82–85 (1998).

[3] E. A. Barannik, A. Girnyk, V. Tovstiak, A. I. Marusenko, S. Y. Emilianov, and A. P. Sarvazyan, "Doppler ultrasound detection of shear waves remotely induced in tissue phantoms and tissue in vitro," Ultrasonics **40**, 849–852 (2002).

[4] E. E. Konofagou and K. Hynynen, "Localized harmonic motion imaging: Theory, simulations and experiments," Ultrasound Med. Biol. **29**, 1405–1413 (2003).

[5] K. Nightingale, M. S. Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility," Ultrasound Med. Biol. **28**, 625–634 (2001).

[6] R. Sinkus, K. Siegmann, T. Xydeas, M. Tanter, C. Claussen, and M. Fink, "MR elastography of breast lesions: Understanding the solid/liquid duality can improve the specificity of contrast-enhanced MR mammography," Magn. Reson. Med. **58**, 1135–1144 (2007).

[7] Y. Qiu, M. Sridhar, J. K. Tsou, K. K. Lindfors, and M. F. Insana, "Ultrasonic viscoelasticity imaging of nonpalpable breast tumors: Preliminary results," Acad. Radiol. **15**, 1526–1533 (2008).

[8] D. E. Discher, P. Janmey, and Y. L. Wang, "Tissue cells feel and respond to the stiffness of their substrate," Science **310**, 1139–1143 (2005).

[9] T. A. Krouskop, P. S. Younes, S. Srinivasan, T. Wheeler, and J. Ophir, "Differences in the compressive stress-strain response of infiltrating ductal carcinomas with and without lobular features—Implications for mammography and elastography," Ultrason. Imaging **25**, 162–170 (2003).

[10] S. Kalyanam, R. D. Yapp, and M. F. Insana, "Poroviscoelastic behavior of gelatin hydrogels under compression: Implications for bioelasticity imaging," ASME J. Biomech. Eng. **131**, 1–21 (2009).

[11] M. Sridhar, J. Liu, and M. F. Insana, "Elasticity imaging of polymeric media," ASME J. Biomech. Eng. **129**, 259–272 (2007).

[12] M. Sridhar and M. F. Insana, "Ultrasonic measurements of breast viscoelasticity," Med. Phys. **34**, 4757–4767 (2007).

[13] R. D. Cardiff and N. Kenney, "Mouse mammary tumor biology: A short history," Adv. Cancer Res. **98**, 53–116 (2007).

[14] C. L. Shaffer, D. Hernando, J. Stastny, S. Kalyanam, J. Haldar, E. Chaney, X. Liang, and M. F. Insana, "Multimodality imaging development using

3-D gel cultures," in BMES Annual Fall Conference, Los Angeles CA (2007), p. 374.

[15] J. Xu, R. Kong, R. Bhargava, and M. F. Insana, "3-D cell co-cultures to develop multimodality breast imaging," in BMES Annual Fall Conference, St. Louis, MO (October 2008).

[16] J. D. Ferry, *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley, New York, 1980).

[17] B. Fabry, G. N. Maksym, J. P. Butler, M. Glogauer, D. Navajas, N. A. Taback, E. J. Millet, and J. J. Freedberg, "Time scale and other invariants of integrative mechanical behavior in living cells," Phys. Rev. E **68**, 041914 (2003).

[18] T. Hasegawa and K. Yosioka, "Acoustic-radiation force on a solid elastic sphere," J. Acoust. Soc. Am. **46**, 1139–1143 (1969).

[19] M. Orescanin and M. F. Insana, "Ultrasonic radiation forces for elasticity imaging of 3-D tissue models," Proc. SPIE **6513**, OH1–OH11 (2007).

[20] S. S. Brunke, M. F. Insana, J. J. Dahl, C. Hansen, M. Ashfaq, and H. Ermert, "An ultrasound research interface for a clinical system," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 198–210 (2007).

[21] J. A. Jensen, *Estimation of Blood Velocities Using Ultrasound: A Signal Processing Approach* (Cambridge University Press, New York, 1996).

[22] R. J. Doviak and D. S. Zrnic, *Doppler Radar and Weather Observations*, 3rd ed. (Academic, San Diego, 1993).

[23] A. W. Al-Khafaji and J. R. Tooley, *Numerical Methods in Engineering Practice* (Oxford University Press, Oxford, 1985).

[24] E. L. Madsen, J. A. Zagzebski, and G. R. Frank, "Oil-in-gelatin dispersions for use as ultrasonically tissue-mimicking materials," Ultrasound Med. Biol. **8**, 277–287 (1982).

[25] We propose a simple engineering model for which two parameters μ and η may be found. To interpret these parameters as physical quantities, we outline in the Appendix a classic physical model from the literature, and we give the conditions where the parameters may be interpreted as material properties.

[26] S. H. Lamb, *Hydrodynamics*, 6th ed. (Cambridge University Press, Cambridge, UK, 1932).

[27] P. Kundu, I. M. Cohen, and H. H. Hu, *Fluid Mechanics* (Academic, New York, 2004).

[28] Y. A. Ilinskii, G. D. Meegan, E. A. Zabolotskaya, and S. Y. Emilianov, "Gas bubble and solid sphere motion in elastic media in response to acoustic radiation force," J. Acoust. Soc. Am. **117**, 2338–2346 (2005).

[29] A. C. Cameron and F. A. G. Windmeijer, "R-squared measures for count regression models with applications to health-care utilization," J. Bus. Econ. Stat. **14**, 209–220 (1996).

[30] P. M. Gilsenan and S. B. Ross-Murphy, "Shear creep of gelatin gels from mammalian and piscine collagens," Int. J. Biol. Macromol. **29**, 53–61 (2001).

[31] J. Bercoff, M. Tanter, M. Muller, and M. Fink, "The role of viscosity in the impulse diffraction field of elastic waves induced by the acoustic radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1523–1536 (2004).

[32] S. Chen, M. Fatemi, and J. F. Greenleaf, "Remote measurement of material properties from radiation force induced vibration of an embedded sphere," J. Acoust. Soc. Am. **112**, 884–889 (2002).

[33] M. Ahearne, Y. Yang, A. J. El Haj, K. Y. Then, and K. K. Liu, "Characterizing the viscoelastic properties of thin hydrogel-based constructs for tissue engineering applications," J. R. Soc., Interface **2**, 455–463 (2005).

[34] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustic*, 4th ed. (Wiley, New York, 2000).

[35] H. L. Oestreicher, "Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics," J. Acoust. Soc. Am. **23**, 707–714 (1951).

[36] A. N. Norris, "Impedance of a sphere oscillating in an elastic medium with and without a slip," J. Acoust. Soc. Am. **119**, 2062–2066 (2006).

[37] E. L. Madsen, H. J. Sathoff, and J. A. Zagzebski, "Ultrasonic shear wave properties of soft tissues and tissuelike materials," J. Acoust. Soc. Am. **74**, 1346–1355 (1983).

Thermoacoustic mixture separation with an axial temperature gradient

D. A. Geller[a] and G. W. Swift
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, MS K764, Los Alamos, New Mexico 87545

(Received 2 October 2008; revised 17 February 2009; accepted 19 February 2009)

The theory of thermoacoustic mixture separation is extended to include the effect of a nonzero axial temperature gradient. The analysis yields a new term in the second-order mole flux that is proportional to the temperature gradient and to the square of the volumetric velocity and is independent of the phasing of the wave. Because of this new term, thermoacoustic separation stops at a critical temperature gradient and changes direction above that gradient. For a traveling wave, this gradient is somewhat higher than that predicted by a simple four-step model. An experiment tests the theory for temperature gradients from 0 to 416 K/m in 50–50 He–Ar mixtures.
© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097767]

PACS number(s): 43.35.Ud, 43.20.Mv, 43.35.Ty [RR] Pages: 2937–2945

I. INTRODUCTION

When sound propagates through a gas mixture confined within a duct, the mixture partially separates, creating gradients in the concentrations of its components along the length of the duct.[1–4] This thermoacoustic separation is due to the oscillating, combined effects of viscosity and thermal diffusion, with appropriate phasing, in the acoustic boundary layer. The thermoacoustic separation mechanism is similar to that employed in a conventional thermal-diffusion column,[5] except for three major differences. First, the radial temperature gradient and thermal diffusion are oscillating in the case of thermoacoustic separation but are steady in the case of conventional thermal diffusion. Second, the temperature excursions δT in thermoacoustic separation, which are due to the adiabatic compressions and rarefactions in the acoustic wave, are small compared to the absolute mean temperature of the gas; in thermal-diffusion columns, though, the temperature difference δT between the walls of the column is generally comparable to the absolute mean temperature. Third, the bulk gas motion that yields large mole-fraction differences $\Delta n \gg k_T \delta T$ (where k_T is the thermal-diffusion ratio) occurs due to acoustically driven, oscillating motion of the gas in thermoacoustic separation, instead of being driven steadily by natural convection as in a thermal-diffusion column, allowing thermoacoustic separation to work with any orientation of the duct, in a coiled or folded duct, or even in the absence of gravity.

Since thermoacoustic separation depends on thermal diffusion, it is important to investigate the effect of an axial temperature gradient. For example, if a separation duct is to be operated at elevated temperatures, one would need to know whether the entire duct must be subjected to high temperatures or whether it is possible to have the feedstock inlet, product outlets, or acoustic drivers reside outside the hot region.

In Secs. II–IV, we revisit the previous derivations[1,2,4] of thermoacoustic separation including now a finite temperature gradient lengthwise along the duct. We show in Sec. II by a simple model that there is a limiting temperature gradient at which the separation process ceases. In Secs. III and IV, we derive the mathematical details of the temperature gradient's effect on the separation. Then, in Sec. V, we describe an experimental apparatus designed for testing the theory, and we conclude in Sec. VI by showing data for separation in 50–50 He–Ar mixtures compared with the theory. The effect of the nonzero axial temperature gradient is small, but not negligible, for the conditions in these experiments.

II. THE BUCKET-BRIGADE MODEL

As has been shown previously,[1–4] the time-averaged thermoacoustic mixture-separation flux occurs due to processes in the acoustic boundary layer. To gain some intuitive understanding of the process in the presence of an axially imposed temperature gradient, we present an updated caricature of the process in Fig. 1. This figure is similar to Fig. 1 of Ref. 2, except that now the boundary carries a time-averaged axial temperature gradient dT_m/dx, whereas it was spatially isothermal in Ref. 2. The duct wall still has sufficiently large specific heat and thermal conductivity with respect to the gas that its temperature is considered to be fixed in time for any location x along the duct.

When thermoacoustic separation occurs, the time-averaged separation flux is carried by the gas parcels located approximately one thermal penetration depth δ_κ from the boundary, because those parcels experience both thermal diffusion and motion, with appropriate phasing. For such a parcel, the oscillating temperature amplitude is approximately $|T_1| = |p_1|/\rho_m c_p = (\gamma-1) T_m |p_1|/\gamma p_m$ due to the approximately adiabatic compressions and rarefactions of the gas. In this expression, $|p_1|$ is the pressure amplitude of the sound wave, ρ_m and p_m are the mean density and pressure of the gas, c_p is the isobaric specific heat, and γ is the ratio of the isobaric and isochoric specific heats. The resulting lateral thermal

[a] Author to whom correspondence should be addressed. Electronic mail: dgeller@lanl.gov

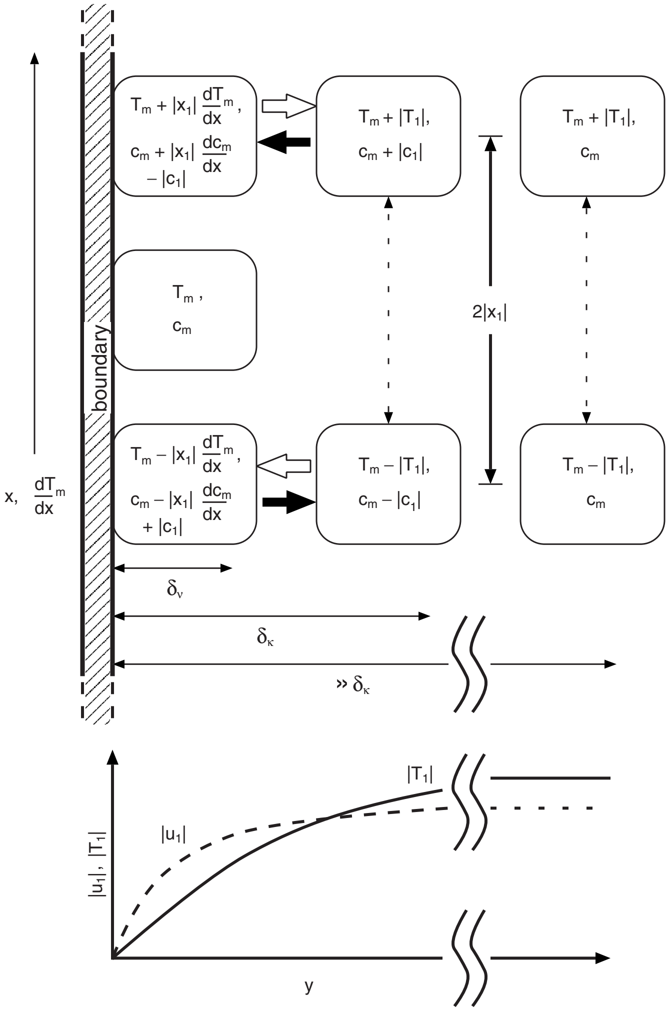


FIG. 1. Discrete time-step model of thermoacoustic separation in a binary mixture, assuming standing-wave phasing and including the effect of a temperature gradient along the duct. The three parcels of gas closest to the wall are locked in place by viscosity, but they each have different mean temperatures due to the thermal gradient along the duct. On the right-hand side of the figure, far from the wall of the duct, one parcel of gas is shown at the extrema of its motion. This parcel experiences no lateral temperature gradient during the motion, because it is outside the thermal boundary layer, as depicted in the bottom portion of the figure. In the middle of the figure, another parcel of gas is shown at the extrema of its motion near the edge of the thermoviscous boundary layer. At the extremes of its motion, there is a temperature gradient between this parcel and the parcels adjacent to the wall, driving thermal diffusion between this parcel and those at the wall. If $dT_m/dx > 0$ for the phasing between motion and pressure considered here, then the lateral thermal gradient, and therefore also the thermal diffusion between the parcels, is lower than it would be for an isothermal boundary.

gradient is determined by this temperature amplitude, by the parcel's distance from the wall, and by the axial displacement amplitude $|x_1|$ of the parcel:

$$\frac{\partial T}{\partial y} \sim \frac{|T_1| - |x_1| \frac{dT_m}{dx}}{\delta_\kappa}. \tag{1}$$

Thermal diffusion will not take place between the moving parcel and a parcel at the same x-location at the boundary when there is no temperature difference between these parcels, i.e., when Eq. (1) is zero. Therefore, thermoacoustic separation ceases when the axial thermal gradient is approximately

$$\left.\frac{dT_m}{dx}\right|_{\text{crit}} = \frac{|T_1|}{|x_1|} = \frac{\gamma - 1}{\gamma} T_m \frac{|p_1|}{p_m} \frac{\omega}{|\langle u_1 \rangle|}, \tag{2}$$

where ω is the angular frequency and $|\langle u_1 \rangle| = \omega |x_1|$ is the amplitude of the spatially averaged velocity. This is exactly the condition defining the critical temperature gradient that differentiates thermoacoustic engines and refrigerators.[6] In the simple model of Fig. 1, we are following the oscillating flow of molecules instead of heat. However, since both of these flows are driven by a lateral temperature gradient, the axial gradient dT_m/dx at which both processes stop must be the same. For higher gradients, thermoacoustic separation will work in the opposite direction, e.g., typically causing a time-averaged flux of the lighter component opposite to the direction of acoustic power flow, analogous to the change in direction of time-averaged enthalpy flow as a thermoacoustic refrigerator is taken past its critical temperature gradient to become an engine.

Similar to our previous calculation of the saturation gradient in Ref. 2, we define $(dT_m/dx)_{\text{crit}}$ exactly as in Eq. (2), and we define

$$\Gamma_T = \frac{dT_m/dx}{(dT_m/dx)_{\text{crit}}} \tag{3}$$

as a dimensionless measure of the temperature gradients in what follows. Although in general this real number could have any magnitude, in our experiments $-1 < \Gamma_T < 1$.

III. DEVELOPMENT OF THE FIRST-ORDER EQUATIONS

To calculate the effect of an axial temperature gradient, we follow the notation and approach of Swift and Spoor[1] by expanding the thermoacoustic variables to first order in the time dependence

$$p = p_m + \Re[p_1(x) e^{\imath\omega t}], \tag{4}$$

$$u = \Re[u_1(x,r) e^{\imath\omega t}], \tag{5}$$

$$T = T_m(x) + \Re[T_1(x,r) e^{\imath\omega t}], \tag{6}$$

$$\rho, c, s\text{: similar to } T, \tag{7}$$

where x is the axial coordinate along the duct, s is the entropy per unit mass, c is the mass fraction of the heavier component, $\Re[]$ denotes the real part, and $\imath = \sqrt{-1}$. The mean temperature T_m was independent of x in Ref. 1.

The diffusive mass-flux density vector is given by Landau and Lifshitz[7] as

$$\mathbf{i} = -\rho D_{12} \left[\nabla c + \frac{k_T'}{T} \nabla T \right], \tag{8}$$

where D_{12} is the mutual diffusion coefficient and k_T' is the mass-scaled thermal-diffusion ratio. The convection and diffusion of the mass fraction c is then

$$\rho\left(\frac{\partial c}{\partial t}+\mathbf{u}\cdot\nabla c\right)=-\nabla\cdot\mathbf{i}=\nabla\cdot\left[\rho D_{12}\left(\nabla c+\frac{k_T'}{T}\nabla T\right)\right]. \tag{9}$$

Inserting the expansions of Eqs. (4)–(7) and keeping only the first-order oscillating quantities, one might consider new terms containing spatial derivatives of T_m or ρ_m that are nonzero due to the gradients in temperature and concentration, such as $-\rho_1 D_{12} k_T'(1/T_m \cdot dT_m/dx)^2$. Those new terms can be neglected because they are of order $(\delta_\kappa/\lambda)^2$ smaller than other terms, where λ is the wavelength, leaving

$$c_1+\frac{u_1}{\iota\omega}\frac{dc_m}{dx}=\frac{\delta_D^2}{2\iota}\left[\nabla_r^2 c_1+\frac{k_T'}{T_m}\nabla_r^2 T_1\right], \tag{10}$$

where $\delta_D=\sqrt{2D_{12}/\omega}$ is the mass diffusion length, which is unchanged from Eq. (18) of Ref. 2. Although the temperature gradient is nonzero, it does not affect the convection and diffusion in the mixture through this first-order equation. (Nevertheless, the gradient dT_m/dx does influence mass flow at *zeroth* order through the steady thermal diffusion in the x direction; this will be seen in the expression for the total separation flux in Sec. IV.)

The oscillating heat transfer in the mixture is given by Eq. (20) of Ref. 1:

$$\rho_m T_m\left(\iota\omega s_1+u_1\frac{ds_m}{dx}\right)=k\nabla_r^2 T_1-\left[k_T'\left(\frac{\partial g}{\partial c}\right)_{p,T}-T_m\left(\frac{\partial g}{\partial T}\right)_{p,c}\right]\nabla\cdot\mathbf{i}_1, \tag{11}$$

where k is the thermal conductivity and g is the Gibbs free energy per unit mass of the mixture, and the divergence of the first-order oscillating mass flux is

$$\nabla\cdot\mathbf{i}_1=-\iota\omega\rho_m c_1-\rho_m u_1\frac{dc_m}{dx} \tag{12}$$

from the left-hand side of Eq. (9). The entropy terms are replaced using the differential identity

$$ds=\frac{c_p}{T}dT-\left(\frac{\partial g}{\partial T}\right)_{p,c}dc-\frac{1}{\rho T}dp \tag{13}$$

from Eq. (22) of Ref. 1. In the case of ds_m/dx, though, we now must keep both the dT_m/dx and dc_m/dx contributions, while the dp_m/dx part remains zero since there is no mean pressure gradient along the duct. Using Eq. (13) to replace ds_m/dx and s_1 in Eq. (11), and using

$$\varepsilon\equiv\frac{(k_T')^2}{T_m c_p}\left(\frac{\partial g}{\partial c}\right)_{p,T}=\frac{\gamma-1}{\gamma}\frac{k_T^2}{n_H(1-n_H)} \tag{14}$$

as originally defined by Ref. 1 and in which n_H is the mole fraction of the heavier component and k_T is the thermal-diffusion ratio, we find

$$T_1=\frac{p_1}{\rho_m c_p}+\frac{\varepsilon T_m}{k_T'}c_1+\frac{\delta_\kappa^2}{2\iota}\nabla_r^2 T_1+\frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx}\frac{u_1}{\iota\omega}-\frac{dT_m}{dx}\frac{u_1}{\iota\omega} \tag{15}$$

after some cancellations. There is only one new term, featuring dT_m/dx, and it is in phase with the concentration-gradient term previously derived.[2]

Solving Eq. (15) for c_1, we substitute the result in Eq. (10) to find

$$T_1=\frac{p_1}{\rho_m c_p}+\frac{1}{2\iota}[\delta_\kappa^2+\delta_D^2(1+\varepsilon)]\nabla_r^2 T_1+\frac{\delta_\kappa^2\delta_D^2}{4}\nabla_r^4 T_1-\frac{dT_m}{dx}\frac{u_1}{\iota\omega}-\frac{\delta_D^2}{2\iota}\left(\frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx}-\frac{dT_m}{dx}\right)\frac{1}{\iota\omega}\nabla_r^2 u_1. \tag{16}$$

Making use of the general expression for the velocity in the duct,

$$u_1=\frac{\langle u_1\rangle}{1-f_\nu}(1-h_\nu), \tag{17}$$

where h_ν is a function describing the shape of the velocity profile across the duct and $f_\nu=\langle h_\nu\rangle$ is the average of this function over the cross section, we can rewrite Eq. (16) as

$$T_1=\frac{p_1}{\rho_m c_p}-\frac{\langle u_1\rangle}{\iota\omega}\frac{1}{1-f_\nu}\frac{dT_m}{dx}+\frac{1}{2\iota}[\delta_\kappa^2+\delta_D^2(1+\varepsilon)]\nabla_r^2 T_1+\frac{\delta_\kappa^2\delta_D^2}{4}\nabla_r^4 T_1+\frac{\langle u_1\rangle}{\iota\omega}\frac{1}{1-f_\nu}\frac{dT_m}{dx}h_\nu+\frac{\delta_D^2}{2\iota}\left(\frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx}-\frac{dT_m}{dx}\right)\frac{\langle u_1\rangle}{\iota\omega}\frac{1}{1-f_\nu}\nabla_r^2 h_\nu, \tag{18}$$

a result which is not limited to a particular duct geometry if ∇_r^2 is taken to be the two-dimensional Laplacian operator in the plane perpendicular to x. This equation is the same as that derived in Ref. 2, except that two new terms appear in the inhomogeneous part that are proportional to dT_m/dx. One of these terms is proportional to the velocity u_1, and the other is proportional to $\nabla^2 u_1$.

This fourth-order differential equation is to be solved for the same boundary conditions as in the previous work. Although the boundary is no longer spatially isothermal, we still assume that its specific heat and thermal conductivity are much larger than that of the gas, so that the boundary at any position x is temporally isothermal. In this case, the amplitude $T_1=0$ at the solid wall, just as it was in the case of the isothermal boundary. The other boundary conditions are that T_1 remains finite everywhere in the duct, and that the diffusive flux into the solid wall is zero:

$$i_r|_{\text{wall}}\propto\left[\nabla_r c_1+\frac{k_T'}{T_m}\nabla_r T_1\right]_{\text{wall}}=0. \tag{19}$$

Using Eqs. (15) and (17), this zero-flux boundary condition becomes

$$\left[(1+\varepsilon)\nabla_r T_1 - \frac{\delta_\kappa^2}{2\imath}\nabla_r^3 T_1 + \left(\frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx} - \frac{dT_m}{dx}\right)\frac{\langle u_1\rangle}{\imath\omega}\frac{1}{1-f_\nu}\nabla_r h_\nu\right]_{\text{wall}} = 0. \qquad (20)$$

In order to calculate the derivatives needed for solving the differential equation, we now have to specify the geometry of the duct. For direct comparison with Ref. 2, it is useful to consider the boundary-layer limit, for which

$$h_\nu = e^{-(1+\imath)y/\delta_\nu} \quad \text{and} \quad f_\nu = \frac{(1-\imath)\delta_\nu}{2r_h}, \qquad (21)$$

where r_h is the hydraulic radius,[8] $\delta_\nu = \sqrt{2\mu/\omega\rho_m}$ is the viscous penetration depth, and μ is the dynamic viscosity. Because it is of more practical interest, though, we also solve the problem for the case of a circular tube of arbitrary radius R. In that case,

$$h_\nu = \frac{J_0[(\imath-1)r/\delta_\nu]}{J_0[(\imath-1)R/\delta_\nu]} \quad \text{and}$$

$$f_\nu = \frac{2J_1[(\imath-1)R/\delta_\nu]}{J_0[(\imath-1)R/\delta_\nu](\imath-1)R/\delta_\nu}, \qquad (22)$$

where the J_i are the usual cylindrical Bessel functions. Calculations based on the Bessel-function solution will be compared with experimental data in Sec. VI.

One may show by substitution that the solution of Eq. (18) is of the form

$$T_1 = \frac{p_1}{\rho_m c_p}\left[1 - \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega}\frac{1}{1-f_\nu}\frac{dT_m}{dx} - Bh_\nu - Ch_{\kappa D} - \left(1 - \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega}\frac{1}{1-f_\nu}\frac{dT_m}{dx} - B - C\right)h_{D\kappa}\right], \qquad (23)$$

where the coefficient of the last term was determined by the boundary condition $T_1|_{\text{wall}}=0$, and the h_i are defined as in Eq. (21) or Eq. (22) but with different length scales δ_i defined below. In the previous solution[2] without a temperature gradient, the r- or y-independent part of T_1 was simply the oscillating temperature due to the adiabatic pressure oscillations of the gas in the sound wave. Here, we have an additional term due to the motion of gas along the temperature gradient. Since $\langle u_1\rangle/\imath\omega \sim \langle x_1\rangle$, the amplitude of the motion, this term modifies the temperature excursion T_1 to include the fact that the gas instantaneously in the particular control volume at location x (in the Eulerian sense) came from a mean position of higher or lower mean temperature. Depending on the relative phasing of p_1 and u_1, this extra term may either raise or lower the amplitude of oscillating temperature at a given point along the duct.

Since the homogenous part of Eq. (18) is unchanged from Ref. 1, the current solution contains the same length scales

$$\delta_{\kappa D}^2 = \tfrac{1}{2}\delta_\kappa^2\left[1 + (1+\varepsilon)/L + \sqrt{[1+(1+\varepsilon)/L]^2 - 4/L}\right], \qquad (24)$$

$$\delta_{D\kappa}^2 = \tfrac{1}{2}\delta_\kappa^2\left[1 + (1+\varepsilon)/L - \sqrt{[1+(1+\varepsilon)/L]^2 - 4/L}\right], \qquad (25)$$

with $L \equiv (\delta_\kappa/\delta_D)^2 = k/\rho_m c_p D_{12}$, appearing in $h_{\kappa D}$ and $h_{D\kappa}$ due to the diffusion of heat and of mass for either the boundary-layer limit or a cylindrical duct; the functional forms of $h_{\kappa D}$ and $h_{D\kappa}$ are the same as those of h_ν, either exponential or Bessel depending on the geometry. In the boundary-layer limit, the δ’s must be positive in order to satisfy the boundary condition that the solution remains finite as $y \to \infty$. For a cylindrical duct, the condition that the solution is finite at the center eliminates the Y_0 Bessel-function solutions. In that case, we can conveniently choose the δ’s to be positive because J_0 is an even function of r.

Substitution of Eq. (23) in Eq. (18) and matching of h_ν terms gives

$$B = -\frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega}\frac{1}{1-f_\nu}\frac{\sigma}{(1-\sigma)(1-\sigma L) - \varepsilon\sigma} \times \left[(\sigma L - 1)\frac{dT_m}{dx} + \frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx}\right] \qquad (26)$$

for arbitrary duct geometry, and with the Prandtl number $\sigma = (\delta_\nu/\delta_\kappa)^2 = \mu c_p/k$. For comparison with Eq. (33) of Ref. 2, this can be recast as

$$B = \frac{\imath e^{-\imath\theta}}{1-f_\nu}\frac{\sigma}{(1-\sigma)(1-\sigma L) - \varepsilon\sigma}\left[(\sigma L - 1)\Gamma_T + \varepsilon\Gamma_c\right], \qquad (27)$$

where θ is the phase by which p_1 leads $\langle u_1\rangle$ and Γ_c is the same as in Ref. 2,

$$\Gamma_c = \frac{dc_m/dx}{(dc_m/dx)_{\text{sat}}} = \frac{dn_H/dx}{(dn_H/dx)_{\text{sat}}}, \qquad (28)$$

where

$$\left(\frac{dc_m}{dx}\right)_{\text{sat}} = \frac{\gamma-1}{\gamma}k_T'\frac{|p_1|}{p_m}\frac{\omega}{|\langle u_1\rangle|}, \qquad (29)$$

$$\left(\frac{dn_H}{dx}\right)_{\text{sat}} = \frac{\gamma-1}{\gamma}k_T\frac{|p_1|}{p_m}\frac{\omega}{|\langle u_1\rangle|}, \qquad (30)$$

and Γ_T is given by Eqs. (2) and (3). For a parcel of gas executing standing-wave motion in a critical gradient of the correct sign with respect to the relative phasing of p_1 and u_1, the temperature of the gas would be approximately constant in time, ignoring the effect of viscosity on the velocity profile.

Finally, the coefficient C is obtained by substitution into the zero-flux boundary condition. For any duct geometry,

$$C = \left\{B\left[f_\nu\frac{(1-\sigma)}{\sigma} - f_{D\kappa}\left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2} - 1\right)\right] + f_{D\kappa}\left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2} - 1\right) + \frac{\imath e^{-\imath\theta}}{1-f_\nu}\Gamma_T\left[f_\nu + f_{D\kappa}\left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2} - 1\right)\right]\right\} \Big/ \left[f_{D\kappa}\left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2} - 1\right) - f_{\kappa D}\left(\frac{\delta_\kappa^2}{\delta_{\kappa D}^2} - 1\right)\right]. \qquad (31)$$

This result can be made more compact for the boundary-layer limit as

$$C = C_{\mathrm{S\&S}}\left\{1 - B\left[1 + \left(\frac{\sigma - 1}{\sqrt{\sigma}}\right)\frac{\delta_\kappa}{\sqrt{L}\delta_{\kappa D} - \delta_{D\kappa}}\right] + \frac{\imath e^{-\imath\theta}}{1 - f_\nu}\Gamma_T\left(1 + \sqrt{\sigma}\frac{\delta_\kappa}{\sqrt{L}\delta_{\kappa D} - \delta_{D\kappa}}\right)\right\}. \tag{32}$$

When $dT_m/dx=0$, this reverts directly to C as written in Eq. (35) of Ref. 2. When the concentration gradient is also set to zero, this expression clearly also reverts to the version of C found in the earlier article, Ref. 1.

Substituting T_1 into Eq. (15), we now obtain the oscillating concentration

$$c_1 = -\frac{p_1}{\rho_m c_p}\frac{k_T'}{\varepsilon T_m}\left\{\left(1 - \frac{1}{\sigma}\right)Bh_\nu + \left(1 - \frac{\delta_\kappa^2}{\delta_{\kappa D}^2}\right)Ch_{\kappa D} + \left(1 - \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega(1 - f_\nu)}\frac{dT_m}{dx} - B - C\right)\left(1 - \frac{\delta_\kappa^2}{\delta_{D\kappa}^2}\right)h_{D\kappa} + \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega(1 - f_\nu)}\frac{dT_m}{dx}h_\nu + \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega(1 - f_\nu)}\frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx}(1 - h_\nu)\right\}, \tag{33}$$

including the effect of a temperature gradient along the duct. This expression is correct for either the boundary-layer limit or the finite cylindrical tube, because Eq. (15) contains only even derivatives of T_1.

IV. THE SEPARATION FLUX TO SECOND ORDER

From Eq. (8) expressed in heavy mole fraction n_H instead of concentration c, the zeroth-order diffusive mole flux of the heavy component is

$$\dot{N}_{H,m} = -NAD_{12}\left(\frac{dn_H}{dx} + \frac{k_T}{T_m}\frac{dT_m}{dx}\right), \tag{34}$$

where A is the cross section of the duct and N is the molar density of the mixture. In the absence of bulk steady flow, the total mole flux of the heavy component through second order, including the effect of the axial temperature gradient, is then

$$\dot{N}_H = \dot{N}_{H,m} + \dot{N}_{H,2} = -NAD_{12}\left(\frac{dn_H}{dx} + \frac{k_T}{T_m}\frac{dT_m}{dx}\right) + \frac{\delta_\kappa}{4r_h}\frac{\gamma - 1}{\gamma}\frac{k_T}{R_{\mathrm{univ}}T_m}|p_1||U_1|[F_{\mathrm{trav}}\cos\theta + F_{\mathrm{stand}}\sin\theta + F_{\nabla c}\Gamma_c + F_{\nabla T}\Gamma_T], \tag{35}$$

where R_{univ} is the universal gas constant, U_1 is the oscillating volumetric velocity, F_{trav} and F_{stand} were first introduced in Ref. 1, and $F_{\nabla c}$ is just F_{grad} from Ref. 2. Using the results of Sec. III, we can calculate the time-averaged second-order mole flux of the heavy component from

$$\dot{N}_{H,2} = \frac{m_{\mathrm{avg}}}{m_H m_L}\frac{A\rho_m}{2}\Re\{\langle c_1\tilde{u}_1\rangle\}, \tag{36}$$

where tilde denotes the complex conjugate, m_H and m_L are the heavy and light molar masses, respectively, and m_{avg} $=n_H m_H+(1-n_H)m_L$ is the average molar mass. Because u_1 depends only on the momentum equation, all new dT_m/dx-dependent terms enter the mole flux linearly through c_1. Writing out all the terms explicitly,

$$\frac{A\rho_m}{2}\Re\{\langle c_1\tilde{u}_1\rangle\} = \frac{1}{2}\frac{k_T'/\varepsilon}{c_p T_m}\Re\left\{\frac{p_1\tilde{U}_1}{1 - \tilde{f}_\nu}\left[-\left(\frac{\sigma - 1}{\sigma}\right)B\langle h_\nu(1 - \tilde{h}_\nu)\rangle - \left(1 - \frac{\delta_\kappa^2}{\delta_{\kappa D}^2}\right)C\langle h_{\kappa D}(1 - \tilde{h}_\nu)\rangle - \left(1 - \frac{\delta_\kappa^2}{\delta_{D\kappa}^2}\right) \times \left(1 - \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega(1 - f_\nu)}\frac{dT_m}{dx} - B - C\right)\langle h_{D\kappa}(1 - \tilde{h}_\nu)\rangle - \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega(1 - f_\nu)}\frac{dT_m}{dx}\langle h_\nu(1 - \tilde{h}_\nu)\rangle - \frac{\rho_m c_p}{p_1}\frac{\langle u_1\rangle}{\imath\omega(1 - f_\nu)}\frac{\varepsilon T_m}{k_T'}\frac{dc_m}{dx}\langle|1 - h_\nu|^2\rangle\right]\right\}. \tag{37}$$

The last term in brackets, times $p_1\tilde{U}_1/(1-\tilde{f}_\nu)$, is purely imaginary so that it evaluates to zero in the $\Re\{\}$ and can be ignored. If $dT_m/dx \to 0$, then this expression reduces to Eq. (48) of Ref. 2. Since those contributions to the mole flux are already known, we can simplify the algebra by subtracting them from Eq. (37) in order to derive the contribution to the mole flux due to the temperature gradient alone. However, when $dT_m/dx \neq 0$, the definitions of B and C given here by Eqs. (27) and (31) contain dT_m/dx-dependent terms, so that one must take care in subtracting Eq. (48) of Ref. 2 from Eq. (37) above.

The flux can be evaluated making use of the identity

$$\langle h_i(1 - \tilde{h}_j)\rangle = \frac{\delta_i^2}{\delta_i^2 + \delta_j^2}(f_i - \tilde{f}_j), \tag{38}$$

which holds true for any channel geometry.[9] If one considers only terms containing dT_m/dx, then one obtains from the portion in brackets above, after considerable work,

$$F_{\nabla T} = \frac{R}{\delta_\kappa}\frac{1}{|1 - f_\nu|^2}\frac{\sigma}{(1 - \sigma)(1 - \sigma L) - \sigma\varepsilon} \times \Im\left\{\frac{f_\nu}{S}\left(\frac{\delta_{D\kappa}^2 - \delta_\kappa^2}{\delta_{D\kappa}^2 + \delta_\nu^2}f_{D\kappa} - \frac{\delta_{\kappa D}^2 - \delta_\kappa^2}{\delta_{\kappa D}^2 + \delta_\nu^2}f_{\kappa D} + \frac{(1 + \sigma)LQ}{M}\tilde{f}_\nu + S\right) + \frac{1 - \sigma L - \sigma\varepsilon}{\sigma}G\right\}, \tag{39}$$

where R is the radius of the circular duct, and

$$S = \left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2} - 1\right)f_{D\kappa} - \left(\frac{\delta_\kappa^2}{\delta_{\kappa D}^2} - 1\right)f_{\kappa D}, \tag{40}$$

$$Q = \frac{\delta_{\kappa D}^2 - \delta_{D\kappa}^2}{\delta_\kappa^2}, \tag{41}$$

$$M = (1 + \sigma)(1 + \sigma L) + \varepsilon\sigma, \tag{42}$$

$$G = \frac{\sigma L Q}{S M} f_{\kappa D} f_{D\kappa} + \frac{\widetilde{f}_{\nu}}{S}\left(\frac{f_{\kappa D}}{1 + \delta_{\nu}^{2}/\delta_{D\kappa}^{2}} - \frac{f_{D\kappa}}{1 + \delta_{\nu}^{2}/\delta_{\kappa D}^{2}}\right). \tag{43}$$

For parallel-plate geometry, use the hyperbolic-tangent expressions[6] instead of the Bessel expressions for f_i.

In the boundary-layer limit, $f_i = (1 - \imath)\delta_i / 2r_h$, so the result for $F_{\nabla T}$ reduces to just

$$-F_{\nabla T} = \frac{\sqrt{\sigma L}(1 + \sqrt{L})(1 + \sigma^{2}\sqrt{L}) + \sigma^{3/2}L + \sqrt{\sigma}\varepsilon - \frac{1}{\sqrt{\sigma}} + [\sqrt{L}(1 - \sigma L - \sigma\varepsilon) - \sigma L(1 + \sigma)]\frac{\delta_{\kappa D} + \delta_{D\kappa}}{\delta_{\kappa}}}{(1 + \sqrt{L})[(1 + \sigma)(1 + \sigma L) + \sigma\varepsilon][(1 - \sigma)(1 - \sigma L) - \sigma\varepsilon]/\sigma}. \tag{44}$$

The various F parameters are plotted in Fig. 2(a) as a function of Ar mole fraction in a He–Ar mixture in the boundary-layer limit. Evidently $F_{\nabla T}$ is a fairly small contributor to the time-averaged mole flux for $|\Gamma_T| < 1$. Figure 2(b) shows the expected effect of this small contribution in our experiment's finite-diameter circular tube, for values of temperature gradient corresponding to $-0.8 < \Gamma_T < 0.8$, which is the range covered by our experiments.

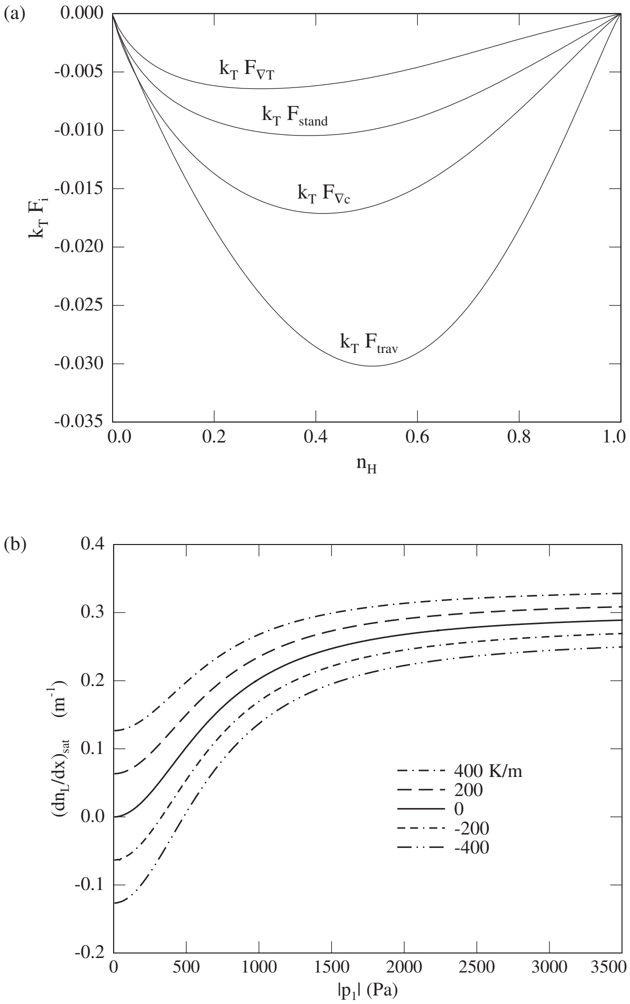


FIG. 2. (a) Comparison of $F_{\nabla T}$ to previously defined F's, for He–Ar in the boundary-layer limit. (b) The calculated mole-fraction gradient at which thermoacoustic mixture separation saturates (i.e., $\dot{N}_H = 0$) versus oscillating pressure amplitude for various values of the temperature gradient along a duct similar to those used in the experiments. These curves are calculated for 80 kPa 50–50 He–Ar mixtures in a 3.3 mm diameter tube at a frequency of 200 Hz. The calculation assumes traveling-wave phasing and a mean temperature of 300 K, and it uses the functional forms for a finite-diameter circular tube.

V. EXPERIMENTAL APPARATUS

The apparatus used to study the behavior of thermoacoustic mixture separation with an axial temperature gradient is a modified version of the apparatus used in the experiments on separation of neon isotopes[3] and on separation with continuous flow[4] of feedstock and products. Its geometry is shown in Fig. 3. In the current arrangement, the separation duct is a single 0.965 m length of stainless-steel tubing and fittings, with an inner diameter of 3.33 mm. The tubing was cut at three locations and joined back together by soldering into custom copper unions. These unions include side taps for attachment of microcapillaries carrying a small flux of the sample gas to a residual gas analyzer[10] (RGA) for concentration analysis. The unions also have side taps for connection to pressure transducers, which were used to verify the acoustic field in the duct. Small holes were drilled in the copper unions and thermocouple transducers were inserted to measure the local temperatures along the duct.

At the middle of the duct—its hottest point—it was necessary to thermally isolate the pressure transducer from the duct for two reasons. First, operating a transducer at elevated temperatures would have required calibration to evaluate its internal temperature compensation. Second, the temperature in the middle of the duct would often exceed the maximum operating temperature of the available transducer.[11] Therefore, this transducer was instead screwed into a small brass chamber with low dead volume, which was connected to the duct through 6 cm of stainless-steel capillary of 0.5 mm inner diameter. The brass chamber was water-cooled to maintain a temperature of about 300 K on the transducer. Calculations showed that this capillary and chamber did not significantly attenuate p_1 at the transducer.

The high operating temperatures in the middle of the duct also required that the microcapillaries to the RGA used to measure the He and Ar mole fractions be designed to withstand temperatures up to 575 K. The 5 cm lengths of glass microcapillary, with internal diameter 10 μm, were ep-

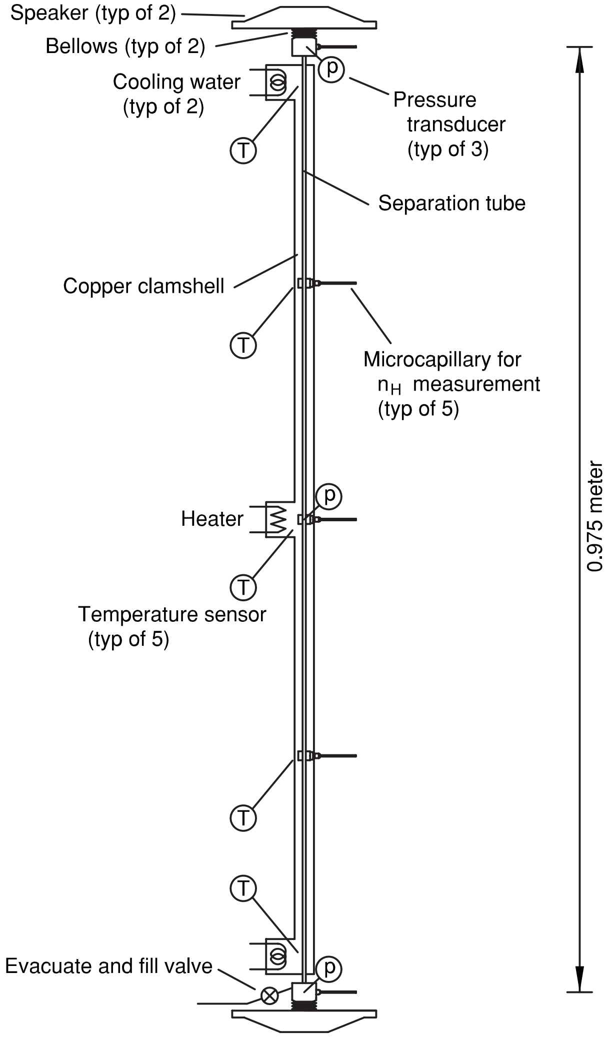


FIG. 3. The mixture-separation apparatus consists of a 3.3 mm diameter, 0.965 m long stainless steel tube, with hermetically sealed acoustic drivers at both ends. The connections to these drivers have ports for measuring the oscillating pressure and for withdrawing the mixture to a RGA, such that the length of the duct across which measurements are made is 0.975 m, slightly longer than the steel tube. The steel tube is sandwiched between two copper bars through which a circular groove was machined. This copper clamshell establishes the nearly linear thermal gradient along the duct, and it is held in place by the copper heatsinks at either end and by a copper clamp in the middle around which the heater is wound. The entire length of the copper clamshell is surrounded by fiberglass insulation (not shown) in order to minimize heat leaks to the surrounding room.

oxied into stainless-steel capillaries using a high-temperature epoxy.[12] These assemblies are attached to the copper unions using Swagelok connectors.

The duct, including the three copper unions, is enclosed in a clamshell formed from two 0.94 m long bars of 1 cm × 1 cm square copper. This clamshell serves as a thermal conductor to enforce a nearly linear temperature gradient along the duct. A ball-end mill was used to cut a groove lengthwise along one side of each bar to accommodate the circular duct. This clamshell was clamped onto the duct with water-cooled heat sinks at each end. The heat sinks cool the ends of the tube approximately to room temperature where they enter the acoustic drivers. Another copper clamp holds the clamshell together at the midpoint of the duct, and a heater tape is wound around this clamp. Heat is thus applied at the midpoint of the duct, so that the copper clamshell generates equal temperature gradients from the center to both ends of the tube, so we can study gradients both along and opposing the traveling wave at the same time in a single experiment.

Each end of the duct is connected to an acoustic driver composed of a hermetically sealed bellows attached to the dome of an electromagnetic speaker. The volume in each bellows is much larger than that of the separation tube itself. The duct is oriented vertically, and the gas mixture sample is introduced through a plug valve near the lower acoustic driver. Pressure transducers screwed directly into the duct near the drivers are used to set the drive voltages of the two speakers in order to achieve the desired pressure amplitudes and relative phases at the ends of the duct. The duct-end conditions on the oscillating pressure are chosen to give a traveling wave with a desired amplitude at the midpoint of the duct. These duct-end pressures are calculated numerically for a desired wave using DeltaEC.[13]

VI. RESULTS

Separation experiments were performed, starting from a spectroscopically verified 50–50 He–Ar mixture, for nine values of the temperature gradient—0, ±116, ±216, ±316, and ±416 K/m—corresponding to five mid-duct temperatures. For each value of the gradient, experiments were performed with traveling waves in the duct with nominal values for the oscillating pressure $|p_1|$ of about 1.0, 1.5, and 3.0 kPa at the midpoint of the duct. Before each experimental run, the temperature profile was established, the residual gas in the apparatus was pumped away to a pressure below 50 μm Hg, and the duct was filled with a fresh gas mixture to a pressure of 80 kPa, which is slightly above the typical atmospheric pressure in Los Alamos. When the duct's fill valve was left open too long, thermal diffusion between the duct and the fill manifold was observed to alter the mean concentrations in the duct over several minutes. This would result in the average concentration of the charge of gas in the duct not being 50–50 after the valve was closed. To minimize this effect, we first filled the filling manifold alone to a pressure calculated such that when the valve to the duct was momentarily opened, the duct would reach the desired equilibrium pressure given the applied thermal gradient. Then the final fill valve was opened just until the panel pressure stopped changing, which took less than 5 s.

After introducing the mixture, the acoustic wave was applied and the sample was allowed to separate for 1–3 h, depending on how long the concentrations were observed to change at the ends of the duct, which in turn depended on the amplitude of the wave and the temperature gradients along the duct. The mole fractions of He and Ar were then measured with the RGA for each of the five microcapillaries along the duct. The partial pressure of nitrogen was also recorded as a diagnostic for detecting leaks of air into the duct.

As noted in Ref. 4, the RGA is much more sensitive to argon than to helium, and the RGA-pumping system responds nonlinearly to the flow rate of the sample gas. In our experiments, the relative sensitivity of the RGA system to

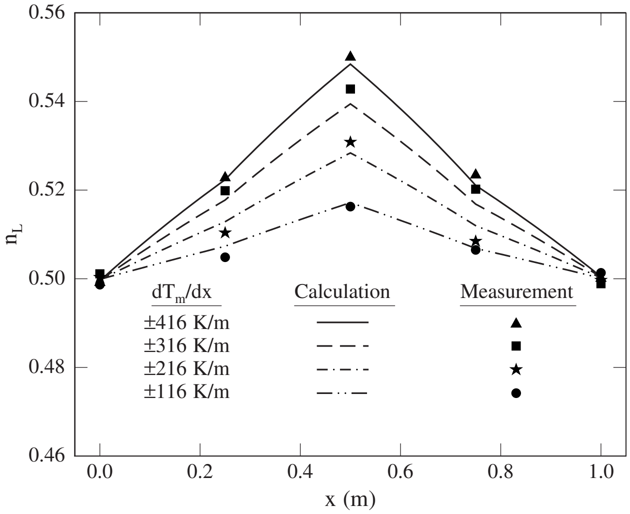


FIG. 4. Thermal diffusion for several applied gradients without acoustic excitation. Symbols are measurements and lines are calculations. At the ends of the duct, the temperature was held at 290 K by a recirculating chiller.

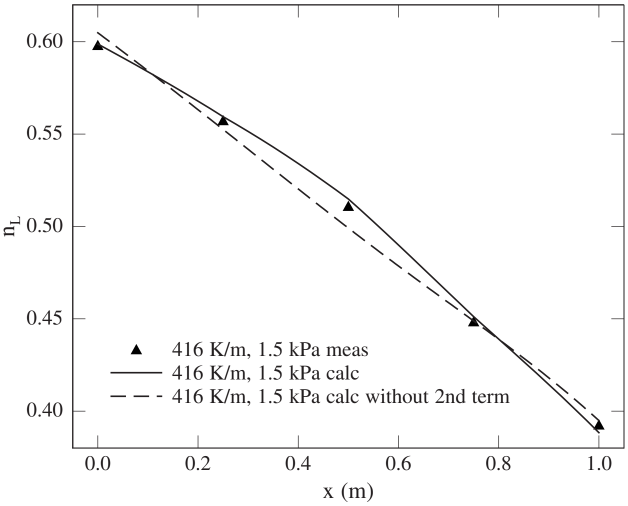


FIG. 5. Concentration versus position for a traveling wave propagating in the $-x$ direction with pressure amplitude $|p_1|=1.5$ kPa at the midpoint and a temperature gradient of ±416 K/m on either side of the midpoint. The solid triangles are measurements of helium mole fraction n_L at the five microcapillaries along the duct. The solid curve is a calculation (Ref. 13) using Eq. (45) and the dotted curve is a corresponding calculation using the theory of Ref. 2, which omits the dT_m/dx term in Eq. (45). The curves were calculated using as boundary conditions the values of acoustic pressure p_1 at each end of the duct and the requirement that the total helium concentration integrated over the apparatus was 0.5, because the fill valve was closed at the beginning of each experiment. Comparing the data and models in this way highlights the small differences in slope arising from the term with $F_{\nabla T}$.

helium versus argon is a function both of the flow rate of the gas through a microcapillary and of the true ratio of the concentrations of the gas mixture being sampled. For the present work, all five microcapillaries have very similar flow impedances so a single calibration might be expected to work for them all. However, the wide range of temperatures required by the experiment cause different flow rates through different microcapillaries, and therefore different relative sensitivities, in a single experiment. To account for this via calibration, the ratio of component partial pressures pp(Ar)/pp(He) reported by the RGA as gas flowed through a single, typical microcapillary was recorded with no sound wave and no temperature gradients, for different mixtures of He–Ar from 70–30 to 30–70 at 80 kPa and for mean pressures from 0 to 87 kPa in the 50–50 mixture. The relative sensitivity depended weakly, if at all, on argon concentration at fixed pressure, observed differences being less than the 0.01–0.02 accuracy of the measurement for the range of concentrations ($0.38 \leq n_L \leq 0.62$) used in our separation experiments. However, the relative sensitivity varied by as much as 0.07 over the range of RGA partial pressures seen in our separation experiments. The relative sensitivity was nonlinear in the argon partial pressure, and at any given pressure it varied from day to day. To account for these problems, we used a fit to the calibration data to scale the relative sensitivity for each microcapillary based on the argon partial pressure measured there, and we forced a weighted average of the mole fractions (weighted by the volume distribution in the entire apparatus) to be exactly 0.5. This procedure reduced the uncertainty in the mole-fraction results to about 0.01, as confirmed in Fig. 4 for thermal diffusion without a sound wave. The thermal-diffusion ratio for He–Ar mixtures is well known,[14] so the calculated curves in Fig. 4 are a standard against which the data can be confidently compared.

When the ends of the tube are closed and the separation is allowed to run until it saturates, then $\dot{N}_H=0$ and the concentration gradient calculated from Eq. (35) is

$$\left(\frac{dn_H}{dx}\right)_{\text{sat}} = \left[\frac{\delta_\kappa}{4r_h}\frac{\gamma-1}{\gamma}\frac{k_T}{R_{\text{univ}}T_m}|p_1||U_1|(F_{\text{trav}}\cos\theta + F_{\text{stand}}\sin\theta) - \left(NAD_{12}-\frac{\delta_\kappa}{4r_h}NA\frac{|\langle u_1\rangle|^2}{\omega}F_{\nabla T}\right)\frac{k_T}{T_m}\frac{dT_m}{dx}\right] \Big/ \left[NAD_{12}-\frac{\delta_\kappa}{4r_h}NA\frac{|\langle u_1\rangle|^2}{\omega}F_{\nabla c}\right]. \qquad (45)$$

The change in the concentration gradient at saturation due to $dT_m/dx \neq 0$ therefore depends on the amplitude of the sound wave in the duct:

$$\Delta\left(\frac{dn_H}{dx}\right)_{\text{sat}} = -\frac{1-\frac{\delta_\kappa}{4r_h}\frac{|\langle u_1\rangle|^2}{\omega D_{12}}F_{\nabla T}}{1-\frac{\delta_\kappa}{4r_h}\frac{|\langle u_1\rangle|^2}{\omega D_{12}}F_{\nabla c}}\left(\frac{k_T}{T_m}\frac{dT_m}{dx}\right). \qquad (46)$$

The temperature gradient has the largest absolute effect when $|\langle u_1\rangle|=0$, because there is no thermoacoustic separation in that case but ordinary thermal diffusion still produces a concentration gradient. In that case, the dimensionless ratio in Eq. (46) is 1. As $|\langle u_1\rangle| \to \infty$, the acoustics dominates and the dimensionless ratio approaches $F_{\nabla T}/F_{\nabla c}$, which for a 50–50 He–Ar mixture is approximately 1/3, as shown in Fig. 2. For a 1.5-kPa traveling wave in our duct, the dimensionless ratio in Eq. (46) is about midway between these two extremes, so the thermoacoustic consequence of nonzero dT_m/dx is comparable with that of ordinary thermal diffusion. Figure 5 shows experimental data and calculations for this case.

For extensive comparison with the theory, separations were performed over a range of temperature gradients span-

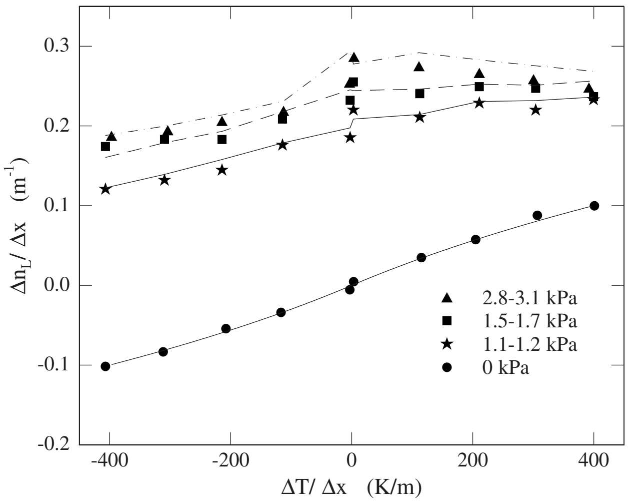


FIG. 6. A representation of all data and corresponding calculations, over the ranges $0 \leq |p_1| \leq 3.0$ kPa and $0 \leq |dT_m/dx| \leq 416$ K/m. The filled symbols are data for the finite-difference gradient in helium concentration versus the finite-difference gradient in temperature. The differences are from the ends of the duct to the middle, ignoring the measurements at the intermediate microcapillaries numbered 2 and 4. The curves are corresponding calculations, matched to the actual pressure amplitude of each measured point.

ning 0–416 K/m and over a range of oscillating pressure amplitudes $|p_1|$ from 0 to 3 kPa. The results are summarized in Fig. 6. This plot shows the gradients in concentration n_L versus gradients in temperature calculated between the middle of the duct, which was usually heated, and the end points of the duct, which were held at room temperature. In this way, each experiment at a different midpoint temperature or amplitude $|p_1|$ contributes two points to the graph. These points can be compared against the curves calculated using Eqs. (34)–(43) as implemented in DeltaEC,[13] for the experimentally measured temperatures and the pressure amplitudes recorded for each run. In general, the data roughly match the calculated curves.

An uncertainty of 0.01 in the measurements of n_L can result in an error of as much as $0.02/(0.5\ \mathrm{m}) = 0.04\ \mathrm{m}^{-1}$ in $\Delta n_L/\Delta x$. However, the highest-amplitude data show deviation from calculated values a little higher than this for $\Delta T/\Delta x > 0$, possibly due to turbulence. The Reynolds number of the oscillating flow is as high as 1600 at the high-amplitude end of the separation tube (the $\Delta T/\Delta x > 0$ end) based on $|U_1|$ and tube diameter. This is near the expected transition to turbulence for oscillating flow[15] in a tube with a diameter of the order of $10\delta_\nu$. The fact that the 3 kPa, $\Delta T/\Delta x > 0$ data here deviate from calculations more than do the 3 kPa, $\Delta T/\Delta x = 0$ data here and the 3 kPa, $\Delta T/\Delta x = 0$ data of Ref. 4 suggests that nonzero axial temperature gradients may affect the transition to turbulence.

Finally, evaluating Eq. (2) for the conditions of this experiment gives $(dT_m/dx)_{\mathrm{crit}} \simeq 500$ K/m, only slightly above the experiment's highest gradients, 416 K/m. Thus, at temperature gradients near critical, we might have expected the acoustic separation to differ very little from the zero-acoustics separation for which axial thermal diffusion alone is responsible. The resolution of this paradox becomes apparent by setting $\dot{N}_H = 0$ and $dn_H/dx = (k_T/T_m) dT_m/dx$ in Eq. (35) and solving for dT_m/dx, obtaining

$$\frac{dT_m}{dx} = \frac{\gamma - 1}{\gamma} T_m \frac{|p_1|}{p_m} \frac{\omega}{|\langle u_1 \rangle|} \frac{F_{\mathrm{trav}} \cos\theta + F_{\mathrm{stand}} \sin\theta}{F_{\nabla c} - F_{\nabla T}}, \qquad (47)$$

the actual temperature gradient for which the presence of the sound wave does not change dn_H/dx. Equation (47) differs from Eq. (2) by a factor depending on the four F's. For a standing wave in 50–50 He–Ar in the boundary-layer limit, Eq. (47) is nearly equal to Eq. (2), because $F_{\mathrm{stand}} \simeq F_{\nabla c} - F_{\nabla T}$. But for a traveling wave in 50–50 He–Ar in the boundary-layer limit, Eq. (47) is 1500 K/m, about three times the value given by Eq. (2). This is the gradient at which all four sets of data in Fig. 6 would approach one another.

We retain Eq. (2) as the formal definition of $(dT_m/dx)_{\mathrm{crit}}$ because it is simple, independent of tube size, and independent of the transport properties of the gas.

ACKNOWLEDGMENTS

This work was supported by Locally Directed R&D funds at Los Alamos National Laboratory. We are grateful to Mike Torrez and Carmen Espinoza for fabrication and assembly of the apparatus.

[1] G. W. Swift and P. S. Spoor, "Thermal diffusion and mixture separation in the acoustic boundary layer," J. Acoust. Soc. Am. **106**, 1794–1800 (1999); **107**(4), 2299 (2000); **109**(3) 1261 (2001).
[2] D. A. Geller and G. W. Swift, "Saturation of thermoacoustic mixture separation," J. Acoust. Soc. Am. **111**, 1675–1684 (2002).
[3] D. A. Geller and G. W. Swift, "Thermoacoustic enrichment of the isotopes of neon," J. Acoust. Soc. Am. **115**, 2059–2070 (2004).
[4] G. W. Swift and D. A. Geller, "Continuous thermoacoustic mixture separation," J. Acoust. Soc. Am. **120**, 2648–2657 (2006).
[5] R. C. Jones and W. H. Furry, "The separation of isotopes by thermal diffusion," Rev. Mod. Phys. **18**, 151–224 (1946).
[6] G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[7] L. D. Landau and E. M. Lifshitz *Fluid Mechanics* (Pergamon, New York, 1982).
[8] The hydraulic radius of a duct is defined as the ratio of its cross section to its perimeter. For a right circular cylinder, the hydraulic radius is equal to one-half of the cylinder radius.
[9] W. Pat Arnott, H. E. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991). See Eq. (45).
[10] RGA 100, Stanford Research Systems, Sunnyvale, CA, www.thinksrs.com (Last viewed February, 2009).
[11] Endevco Model 8510B-5, San Juan Capistrano, CA, www.endevco.com (Last viewed February, 2009).
[12] Aremco 526N, Valley Cottage, NY, www.aremco.com (Last viewed February, 2009).
[13] W. C. Ward, J. P. Clark, and G. W. Swift "Interactive analysis, design, and teaching for thermoacoustics using DeltaEC," J. Acoust. Soc. Am. **123**(5), 3546 (2008); software and user's guide available at www.lanl.gov/thermoacoustics/DeltaEC.html (Last viewed February, 2009). Version 6.2b3 included the segment type MIXTCIRC, which uses the equations derived in this paper.
[14] K. E. Grew and T. L. Ibbs, *Thermal Diffusion in Gases* (Cambridge University Press, Cambridge, 1952).
[15] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Melville, NY, 2002), Fig. 7.4.

Correction for partial reflection in ultrasonic attenuation measurements using contact transducers

Martin Treiber and Jin-Yeon Kim[a)]
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Laurence J. Jacobs
School of Civil and Environmental Engineering and GWW School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Jianmin Qu
GWW School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 28 October 2008; revised 23 February 2009; accepted 24 February 2009)

This research investigates the influence of partial reflection on the measurement of the absolute ultrasonic attenuation coefficient using contact transducers. The partial, frequency-dependent reflection arises from the thin fluid-layer interface formed between the transducer and specimen surface. It is experimentally shown that neglecting this reflection effect leads to a significant overestimation in the measured attenuation coefficient. A systematic measurement procedure is proposed that simultaneously obtains the ultrasonic signals needed to calculate both the reflection coefficient of the interface and the attenuation coefficient, without disturbing the existing coupling conditions. The true attenuation coefficient includes a correction based on the measured reflection coefficient—this is called the *reflection correction*. It is shown that including the reflection correction also reduces the variation (random error) in the measured attenuation coefficient. The accuracy of the proposed method is demonstrated for a material with a known attenuation coefficient. The proposed method is then used to measure the high attenuation coefficient of a cement-based material. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106125]

PACS number(s): 43.35.Yb, 43.20.Ye [YHB] Pages: 2946–2953

I. INTRODUCTION

Together with the wave speeds and the acoustic nonlinearity parameter, the attenuation coefficient is one of the fundamental acoustic parameters of a material. This macroscopic parameter contains information on a material's microstructure such as grain structure, dislocations, mesoscale inhomogeneity, etc., and thus can often be related to the damage that evolves from the microstructural changes during fatigue, creep, and other damage processes.[1–3] For this reason, an accurate measurement of the ultrasonic attenuation coefficient of a solid material is important.

A number of different techniques have been proposed based on different measurement principles. Hartmann and Jarzynski[4] developed an immersion technique in which both the sample and the transducers are immersed in a bath filled with water or other liquid that is used as the couplant. This technique has the advantage that the coupling between the sample and transducers is perfect and an exact acoustic reflection at the water-sample interface can be calculated. Toksoz *et al.*[5] and Sears and Bonner[6] used a reference-based method. A material sample that has a known or very low attenuation is taken as a reference sample. Two transducers are attached to both sides of a sample to measure the first transmitted signal. Frequency spectra of the transmitted signals from the reference and current samples are compared to obtain the attenuation of the current sample. This technique is simple but the conditions at the interfaces between the sample and transducers cannot be explicitly considered. When the sample is thin such that its thickness is equivalent to only a few wavelengths of ultrasound, pulses from the sample are not separate in the time domain. In this case, the so-called buffer-rod technique proposed by Papadakis[7] can be used. A buffer-rod that is much thicker than the sample is bonded to one of the sample's surfaces. The first echo signal from the sample buffer-rod interface and two following signals (once and twice reflected in the sample thickness) are compared to obtain the attenuation coefficient. In this technique, one needs to know an exact value of the transmission or reflection coefficient at the sample buffer-rod interface and the bond thickness. Papadakis[7] analyzed influences of the bond between the sample and the buffer-rod on the attenuation measurement. However, this analysis uses an *a priori* knowledge on the elastic properties and thickness of the bond. In practice, it is almost impossible to predict these parameters precisely. This technique has been further developed by Kushibiki *et al.*[8] for determining the attenuation in a very high frequency range. Redwood and Lamb[9,10] and McSkimin[11] used guided waves for measuring the attenuation in a cylindrical rod sample. The sample boundary confines the acoustic energy along the acoustic wave path, which leads to a need for a different correction for losses due to mode conversion upon multiple reflections from the side wall. As pointed out by Truell *et al.*,[12] the determination of the losses caused by the mode conversion is difficult and

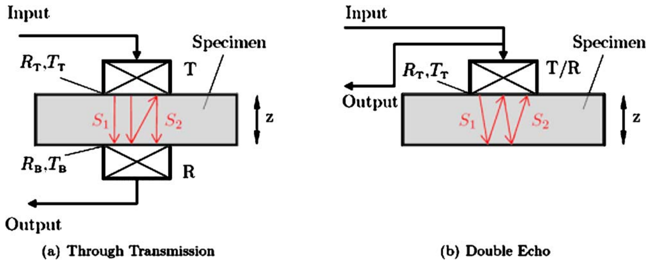
[a)]Author to whom correspondence should be addressed. Electronic mail: jk290@mail.gatech.edu depends on sample geometry and elastic properties, and the error also depends on the value of attenuation that is being measured. A more elaborated method is the pulse interference technique[13] that was modified from McSkimin's pulse overlap technique[14] for measuring wave speed. Among others, this paper considers, in particular, a contact measurement technique that uses a short pulse signal.

Ultrasonic attenuation measurement techniques using contact transducers[15] that are coupled to a material sample with a coupling agent (liquid or a solid-state bond) are widely employed in the laboratory and field. These techniques have the advantage that they can be applied in situations where a high amount of incident wave energy is required, or where the nature of the material or measurement setup does not allow for the immersion of the specimens into water. A disadvantage may be that the coupling condition between the transducer and the sample is not completely reproducible, which significantly influences the measured attenuation coefficient, and can potentially produce a large random and/or bias error in the measurement results. In contact attenuation measurement techniques, two time-domain ultrasonic pulse signals are experimentally obtained, and the ratio of their spectra is taken to obtain the attenuation spectrum. The influences of transducer-sample contact conditions in these two experimentally measured time-domain signals are assumed to be common and cancel out when calculating the ratio of their spectra. As will be shown in Sec. IV, the effects of the contact conditions are not completely removed, and must be quantitatively accounted for. While contact measurement techniques are widely used, and it is well recognized that the interfacial condition can significantly influence the ultrasonic measurement results, a systematic way to remove or reduce this influence in the attenuation measurement has not previously been presented.

In this paper, the effects of partial reflection from the interface between the specimen and the transducer surfaces are experimentally evaluated and compared with the well-known beam diffraction effect[16,17] to show the importance of this effect. A systematic procedure is developed in which the reflection coefficient of the interfaces is measured *in-situ* during the attenuation measurement without disturbing the current coupling condition. This measured reflection coefficient is used to develop a reflection correction to the attenuation coefficient. It is shown that the reflection correction can significantly reduce the experimental scatter in the measured attenuation coefficient. The accuracy and robustness of the proposed method are demonstrated by making measurements on a well-known material, polymethyl methacrylate (PMMA). Finally, the proposed method is used to measure the longitudinal wave attenuation coefficient of a cement paste sample.

II. ATTENUATION MEASUREMENT USING CONTACT TRANSDUCERS

In order to measure the ultrasonic attenuation coefficient of a material, the spectral amplitudes of two ultrasonic pulse signals that have propagated different distances are compared; Fig. 1 shows two experimental setups using contact transducers that are frequently employed to measure attenuation. The first setup, called a through transmission technique, uses two ultrasonic transducers, one as a transmitter and the other as a receiver, that are acoustically coupled to both sides of the sample [Fig. 1(a)] with a thin liquid layer of couplant. Sufficient clamping force should be applied to the transducers to secure good contact with the sample surface. The first (S_1) and second (S_2) through-transmitted signals are measured by the receiving transducer and used to calculate the attenuation coefficient. These signals travel one (z) and three ($3z$) times the sample thickness. The second setup, called a double echo technique, uses a single transducer that is acoustically coupled to one side of the sample with a thin layer of liquid couplant [Fig. 1(b)]. This transducer transmits an ultrasonic pulse into the sample, and also receives the echoes that are reflected from the other side of the sample, which is left stress-free. Usually, the first (S_1) and second (S_2) reflected signals, which travel twice ($2z$) and four ($4z$) times the sample thickness, are used to calculate the attenuation coefficient. The through transmission technique requires the signal to travel a relatively shorter distance (three times the sample thickness) than the double echo technique (four times the sample thickness). Therefore, when the attenuation is high and/or the sample is thick, the through transmission technique is likely to have better signal-to-noise-ratio than the double echo technique.



(a) Through Transmission

(b) Double Echo

FIG. 1. (Color online) Attenuation measurement setups using contact transducers. (a) Through transmission technique and (b) double echo technique.

Two surfaces of the sample are carefully polished such that they are smooth and perfectly parallel. In general, to predict the reflection coefficient of an interface between two solid materials, one should know accurately the thickness and acoustic properties of the interface, which, however, can only be obtained from a precision measurement.[18,19] For convenience, the boundary condition of the interface between the ultrasonic transducer and the sample surface is usually assumed to be that of the liquid (couplant)-solid material interface or approximately "-1" neglecting any loss at the interface. The latter approximation implies that most of the ultrasonic energy reflects at the interface (a near total reflection) and the transducer detects very small energy leaked on the surface. Intuitively, the reflection from this interface will involve two effects: partial reflection from the thin liquid layer between two solids[18,19] and diffraction from the finite-size aperture of a reflector (transducer). This means that the reflection coefficient will be frequency-dependent, and its magnitude will be less than unity. Therefore, whenever a contact type transducer is used to measure the attenuation coefficient and other acoustic parameters, the effects of partial reflection from the transducer-material interface should be taken into account.

Neglecting losses at the interfaces under the assumption of a near total reflection, the attenuation coefficient in either the through transmission or the double echo setup is

$$\alpha(f) = \frac{1}{2z}\left[\ln\left(\frac{S_1(f)}{S_2(f)}\right) - \ln\left(\frac{D_1(f)}{D_2(f)}\right)\right], \tag{1}$$

where $S_1(f)$ and $S_2(f)$ are the magnitudes of the complex frequency spectra of the first and second signals, and $D_1(f)$ and $D_2(f)$ are the magnitudes of the complex diffraction correction functions[20] corresponding to the propagation distances of these signals. In many cases, the attenuation coefficient has been calculated using this formula. However, as will be shown Secs. III and IV, use of this formula can lead to large errors (overestimation) in the measured attenuation coefficient.

III. THEORY FOR ATTENUATION MEASUREMENTS

Consider a through-the-thickness transmission ultrasonic measurement setup in which two ultrasonic transducers are fluid-coupled to both sides of a material sample as shown in Fig. 1(a). The material is assumed to be macroscopically homogeneous and the signal distortion due to the coherent scattering noise is relatively small. The frequency characteristics of the material for acoustic beam propagation along the $+z$ axis from an acoustic source can be written as

$$\mathbf{H}(f;z) = e^{ikz-\alpha z}\mathbf{D}(f;z), \tag{2}$$

where $\mathbf{D}(f;z)$ is the complex diffraction correction function[20] and k $(=2\pi f/c)$ is the wave number in the material having a wave speed c. In Eq. (2), the time dependence $e^{-i\omega t}$ is omitted for brevity. The acoustic properties at the interfaces between the transducers and sample surfaces can be characterized with the reflection and transmission coefficients, which are denoted by $\mathbf{R}_T$ and $\mathbf{T}_T$ for the top surface and by $\mathbf{R}_B$ and $\mathbf{T}_B$ for the bottom surface [Fig. 1(a)]. More specifically, the transmission coefficients are defined for the case in which the ultrasonic pulse is transmitted from the transducer into the material sample while the reflection coefficients are defined for the case in which the incident pulse from the material is reflected off the material-transducer interface. Since these coefficients are, in general, complex quantities (and thus frequency-dependent), they are denoted here by boldface letters.

The spectra of the first and second signals in the through transmission setup [Fig. 1(a)] can be written as

$$\mathbf{S}_1(f) = \mathbf{I}\mathbf{G}_T\mathbf{T}_T(-\mathbf{T}_B)\mathbf{G}_B\mathbf{D}(f;z)e^{-\alpha z+ikz}, \tag{3}$$

$$\mathbf{S}_2(f) = \mathbf{I}\mathbf{G}_T\mathbf{T}_T\mathbf{R}_B\mathbf{R}_T(-\mathbf{T}_B)\mathbf{G}_B\mathbf{D}(f;3z)e^{-3\alpha z+3ikz}, \tag{4}$$

where $\mathbf{I}(f)$ denotes the spectrum of the input signal fed to the transmitting transducer and $\mathbf{G}_T$ and $\mathbf{G}_B$ are the transfer functions of the transducers on the top and bottom surfaces. Taking a natural logarithm of the ratio between Eqs. (3) and (4) yields the expression for the attenuation coefficient

$$\alpha(f) = \frac{1}{2z}\left[\ln\left(\left|\frac{\mathbf{S}_1(f)}{\mathbf{S}_2(f)}\right|\right) - \ln\left(\left|\frac{\mathbf{D}(f;z)}{\mathbf{D}(f;3z)}\right|\right) + \ln(|\mathbf{R}_B\mathbf{R}_T|)\right]. \tag{5}$$

It is seen that the reflection coefficients of both interfaces are involved in this expression while the transducers' transfer functions and the transmission coefficients are not. Of course, if the reflection coefficients are assumed to be -1, Eq. (5) is identical to Eq. (1). This means that the retention of the reflection term will introduce some correction to the measured attenuation coefficient. The question is how significant is it?

In a similar fashion, the spectra of the first and second signals in the double echo setup [Fig. 1(b)] can be written as

$$\mathbf{S}_1(f) = \mathbf{I}(\mathbf{G}_T\mathbf{T}_T)^2\mathbf{D}(f;2z)e^{-2\alpha z+2ikz}, \tag{6}$$

$$\mathbf{S}_2(f) = -\,\mathbf{I}(\mathbf{G}_T\mathbf{T}_T)^2\mathbf{R}_T\mathbf{D}(f;4z)e^{-4\alpha z+4ikz}. \tag{7}$$

Note that it is assumed that the transducer acts in a reciprocal fashion, that is, its reception and transmission frequency characteristics are identical, and that the reflection coefficient at the free surface is assumed to be $\mathbf{R}_B=-1$. The expression of the attenuation coefficient is

$$\alpha(f) = \frac{1}{2z}\left[\ln\left(\left|\frac{\mathbf{S}_1(f)}{\mathbf{S}_2(f)}\right|\right) - \ln\left(\left|\frac{\mathbf{D}(f;2z)}{\mathbf{D}(f;4z)}\right|\right) + \ln(|\mathbf{R}_T|)\right]. \tag{8}$$

It is seen that this expression also involves the reflection coefficient term. Note that most previous research did not take these reflection effects into account, but instead implicitly assumed free surface reflections on both sides of the sample. Section IV examines the significance of these reflection coefficients on the attenuation measurement.

IV. INFLUENCE OF PARTIAL REFLECTION

Equations (5) and (8) both contain the term of the reflection coefficient that defines the acoustic characteristics of the interface between the sample and the transducer. Two influences of the partial reflection are as follows. First, since the reflection coefficients are chiefly determined by the coupling conditions of the measuring transducers, variations in contact conditions lead to variations in the measured reflection coefficients from measurement to measurement; this can cause a large scatter in the attenuation coefficient when multiple measurements are performed. Second, taking reflection effects into account in the analysis of the attenuation coefficient prevents the attenuation coefficient from being overestimated.

As an example, Fig. 2 shows the reflection coefficients for the top and bottom surfaces for a cement paste material sample. Broadband contact transducers with a nominal center frequency of 5 MHz and a diameter of 12.7 mm are used in this measurement. The transducers are coupled to the sample surfaces with light lubrication oil, and the sample surfaces are flat and smooth. The reflection signals from the free surfaces (top and bottom) are measured first to get the

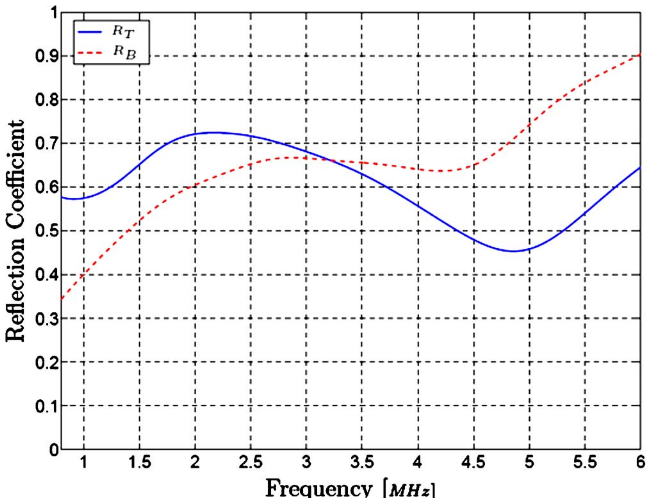


FIG. 2. (Color online) Reflection coefficient magnitudes of the top and bottom interfaces.

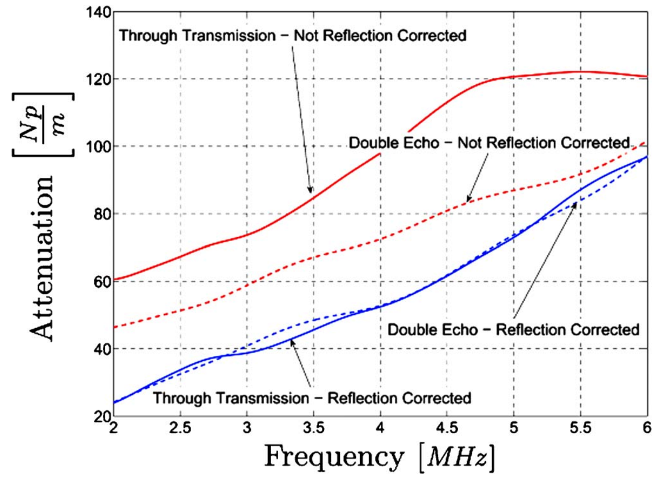


FIG. 3. (Color online) Influence of the reflection coefficient.

transducers' spectra, and the reflection signals from the transducer-mounted surfaces are then measured. The reflection coefficients are calculated[18,19] using

$$\mathbf{R} = \frac{\mathbf{S}(f)}{\mathbf{S}(f)_{\text{free}}}, \tag{9}$$

where $\mathbf{S}(f)$ and $\mathbf{S}(f)_{\text{free}}$ are the spectra of the ultrasonic pulses reflected from the transducer-mounted and free surfaces, respectively. As described in Sec. V, this reflection coefficient measurement is combined with the attenuation measurement procedure. The reflection coefficients shown in Fig. 2 are obtained from the combined measurement procedure.

It is observed that $|\mathbf{R}_T| < 1$ and $|\mathbf{R}_B| < 1$, so the assumption of a free surface is not valid, and the reflection coefficients of the top and bottom surfaces are different even though the mechanical parameters of the transducers are the same, and the clamping forces on the transducers are quite similar. While both curves show similar trends, indicating similar coupling conditions on the top and bottom sides of the sample, they are not exactly the same, which signifies that it is impossible to reproduce the exact same coupling situation every time. This unrepeatability inevitably introduces random errors in the measured attenuation coefficients. The variance of the random errors will depend on various factors such as applied pressure, amount of couplant and so on, which cannot be fully controlled to be the same in every measurement. Note also that the reflection coefficients are frequency-dependent (and thus complex quantities). All of this means that the reflection coefficient of the transducer-sample interface has to be measured *in-situ* while the attenuation coefficient is being measured; this partial reflection coefficient cannot be measured separately or simply assumed to be a specific value. Any overestimation of the attenuation coefficient will be due to partial reflection, as can be observed by inspection of Eqs. (5) and (8). Since $|\mathbf{R}_T| < 1$ and $|\mathbf{R}_B| < 1$, it always holds true that $\ln|\mathbf{R}_T\mathbf{R}_B| < 0$ and $\ln|\mathbf{R}_T| < 0$ for the through transmission and the double echo modes, respectively. Consequently, the last terms in Eqs. (5) and (8) are negative, which causes a decreasing correction to the attenuation coefficients in both cases. Physically this decreasing correction corresponds to some energy absorption by the thin viscoelastic couplant layer and some energy transmission into the transducer material.

Figure 3 shows the influence of this partial reflection on the attenuation coefficient of the cement paste sample for the two measurement setups of Fig. 1. Note that these results are obtained using the measurement procedure described in Sec. V in which the reflection and attenuation coefficients are measured simultaneously, and average values from three repeated measurements. Figure 3 compares attenuation coefficients with and without these reflection effects taken into account. The effect of the reflection coefficient is very pronounced; when the reflection effects are not accounted for, the attenuation coefficient of this material is overestimated by as much as about 40 Np/m in the through transmission setup, and by about 20 Np/m in the double echo setup. The overestimation is stronger in the through transmission mode because the reflection coefficients of both sides are involved, rather than only one side as is the case of the double echo mode. Even though the position and sample are exactly the same in both measurements, the uncorrected curves show a large deviation, while the corrected attenuation curves coincide almost exactly. This agreement indirectly proves the validity of the correction concept and method proposed in this paper. Theoretically, the corrected curves should be identical. However, due to the measurement uncertainty, small discrepancies between the two reflection curves still exist even after the corrections are made. The effects of the partial reflection will of course depend on the material (the acoustic impedance mismatch). In the case of a metal specimen that has higher acoustic impedance, the effects will be even more significant than in the cement paste specimen.

Figure 4 compares the effect of beam diffraction to that of reflection; beam diffraction describes the spatial variation and decay in amplitude of an acoustic beam radiated from a baffled piston source.[17,20] Figure 4 shows the results of an attenuation measurement on a cement paste sample, where the attenuation coefficients are evaluated first with both diffraction and reflection effects taken into account, then with

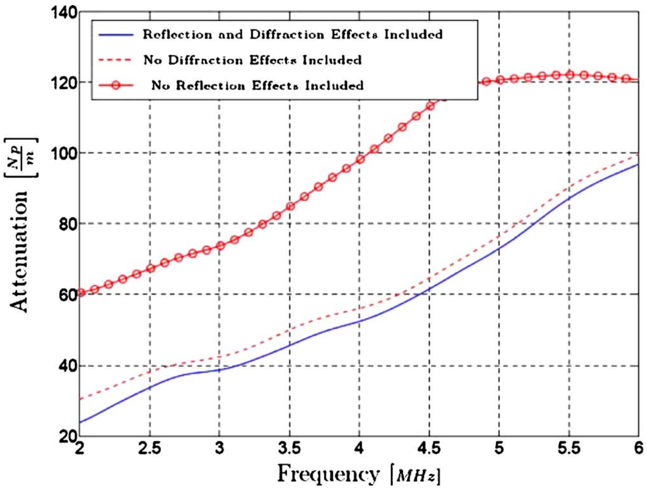


FIG. 4. (Color online) A comparison of the influences of partial reflection and beam diffraction.

the diffraction effect neglected, and finally with the reflection effect neglected. Both effects commonly produce an overestimation of the attenuation coefficient when they are neglected. However, the influence of the reflection effect is much stronger than the diffraction effect for the material and frequency range presented here.

V. MEASUREMENT PROCEDURE

The reflection coefficient quantitatively describes the *current* coupling state of the transducer to the sample. While one may attempt to precisely control the coupling conditions with a clamping device, this control will be extremely cumbersome and based on trial-and-error. An easier and more straightforward approach is to measure the current coupling condition in each measurement being performed, and then make a correction to the attenuation coefficient—this procedure will require a few more measurement steps in the fundamental attenuation measurement procedure shown in Fig. 1. Since the coupling state is unique (not reproducible once disturbed), these additional measurements should be done *in-situ* during the attenuation measurement. Here, a measurement procedure that has been used in the present research is described. The procedure integrates the measurement of the current reflection coefficients into the attenuation measurement without disturbing the coupling conditions at the interface. In addition, both measurement techniques (through transmission and double echo) are combined into this single procedure as shown in Fig. 5. Note that while it is not the only procedure possible and one may develop another procedure that can achieve the same goal in a different manner, the proposed methodology has been found to be quite useful and easy to implement.

In the first step M1, a reflection signal from the free top surface $[s^R_{\mathrm{free;top}}(t)]$ is measured with transducer 1. Then, transducer 2 is mounted on the top surface. In the next step M2, a reflection signal from the sample-transducer 2 interface $[s^R_{\mathrm{interf;top}}(t)]$ is obtained. In M3, a signal transmitted to transducer 2 from transducer 1 $[s^T_{\mathrm{bott}\rightarrow\mathrm{top}}(t)]$ is collected.

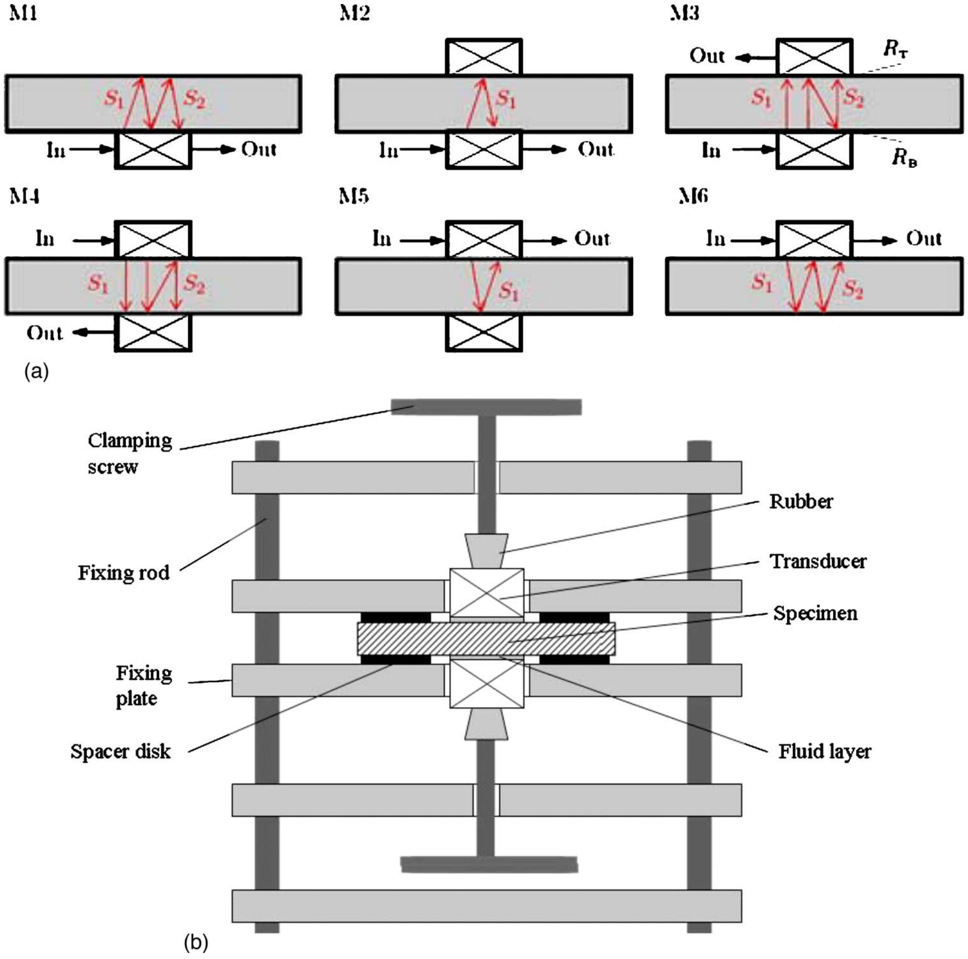


FIG. 5. (Color online) A six-step measurement procedure (a) and a schematic of fixture used in the present research (b).

Then, the two transducers are switched. In M4, the transmission in the opposite direction [$s^{T}_{\text{top}\to\text{bott}}(t)$] is measured. Finally, reflection signals with and without transducer 1 on the bottom surface [$s^{R}_{\text{interf;bott}}(t)$ and $s^{R}_{\text{free;bott}}(t)$] are taken in steps M5 and M6, respectively. Spectra of two echo signals in $s^{R}_{\text{free;top}}(t)$ and $s^{R}_{\text{free;bott}}(t)$ are denoted $\mathbf{S}^{R1}_{\text{free;top}}(f)$, $\mathbf{S}^{R2}_{\text{free;top}}(f)$, $\mathbf{S}^{R1}_{\text{free;bott}}(f)$, and $\mathbf{S}^{R2}_{\text{free;bott}}(f)$, respectively. Those of the first echo signals in $s^{R}_{\text{interf;top}}(t)$ and $s^{R}_{\text{interf;bott}}(t)$ are $\mathbf{S}^{R}_{\text{interf;top}}(f)$ and $\mathbf{S}^{R}_{\text{interf;bott}}(f)$. The spectra of transmitted echo signals in steps M3 and M4 are $\mathbf{S}^{T1}_{\text{bott}\to\text{top}}(f)$, $\mathbf{S}^{T2}_{\text{bott}\to\text{top}}(f)$, $\mathbf{S}^{T1}_{\text{top}\to\text{bott}}(f)$, and $\mathbf{S}^{T2}_{\text{top}\to\text{bott}}(f)$.

The reflection coefficients of the transducer-mounted top and bottom surfaces are calculated $\mathbf{R}_T=\mathbf{S}^{R}_{\text{interf;bott}}(f)/\mathbf{S}^{R1}_{\text{free;top}}(f)$ and $\mathbf{R}_B=\mathbf{S}^{R}_{\text{interf;bott}}(f)/\mathbf{S}^{R1}_{\text{free;bott}}(f)$. With these reflection coefficients, one can determine the attenuation coefficients in the double echo mode in two ways: one with [$\mathbf{S}^{R1}_{\text{free;top}}(f)$ and $\mathbf{S}^{R2}_{\text{free;top}}(f)$] and the other with [$\mathbf{S}^{R1}_{\text{free;bott}}(f)$ and $\mathbf{S}^{R2}_{\text{free;bott}}(f)$], using Eq. (8). In a similar fashion, there are also two ways to determine the attenuation coefficients from the signals obtained in the through transmission mode: one with [$\mathbf{S}^{T1}_{\text{bott}\to\text{top}}(f)$ and $\mathbf{S}^{T2}_{\text{bott}\to\text{top}}(f)$] and the other with [$\mathbf{S}^{T1}_{\text{top}\to\text{bott}}(f)$ and $\mathbf{S}^{T2}_{\text{top}\to\text{bott}}(f)$], using Eq. (5). This procedure yields the attenuation coefficients measured by four different ways.

The measurements in this research use a specially designed fixture that enables mounting or demounting the transducer on one side without disturbing the coupling condition on the other side. A schematic is shown in Fig. 5(b).

VI. APPLICATIONS

A. Reference measurement

To demonstrate the robustness and accuracy of the proposed measurement technique, reference measurements on two PMMA samples (Lucite) with thicknesses of 25.4 and 9.2 mm are performed. This material is chosen because its attenuation characteristics (the level and linear dependence on frequency) are similar to those of materials under investigation in our current research. As a signal (pulse) source, the pulser/receiver Panametrics 5072 PR with a 30 MHz frequency band is used. The transducers are coupled to the sample using low viscosity oil (Bel-Ray AW Lube 10). The ultrasonic transducers used are a broadband longitudinal pair with center frequencies of 5 MHz, and a diameter of 12.7 mm $\left(\frac{1}{2}\right.$ in.$\left.\right)$. The transducers are selected such that their frequency spectra are as close to Gaussian as possible. A rectangular window is used to extract the ultrasonic pulses out of a whole length signal prior to performing the fast Fourier transform, avoiding any undesirable windowing artifacts being introduced. The diffraction correction of Rogers and Van Buren[20] is performed. The result obtained for the longitudinal wave attenuation coefficient is shown in Fig. 6. As seen in Fig. 6, the attenuation coefficient of PMMA can be well approximated by a linear function, $\alpha=12.8f-2.68$ with f in MHz. Similar linear behavior has been observed in many polymeric materials.[21] The attenuation mechanism in polymers is explained with the hysteresis motions of long molecular chains—due to their length and the complex molecular structure in polymers, these molecular chains do not return to their initial locations once they are dislocated by ultrasonic waves and thus some portion of work done by the ultrasonic waves is not stored elastically, causing the hysteresis cycle and reduction in the wave amplitude. The linear behavior may be characterized by the attenuation per wavelength $\alpha\lambda=\text{const}$, where λ denotes the wavelength of the propagating wave. Considering the first term of the linear regression, the attenuation per wavelength is obtained as follows:

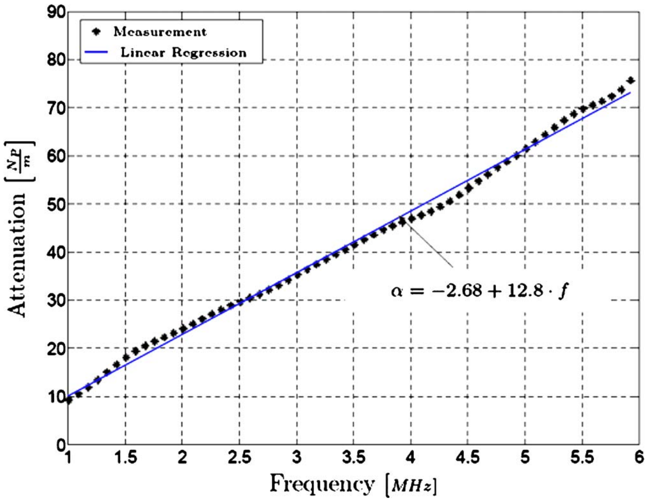


FIG. 6. (Color online) Measured longitudinal wave attenuation coefficient of PMMA and linear regression.

$$\alpha \cdot \lambda = 12.8f\frac{c_L}{f} = 12.8c_L = 0.036 \ \text{Np}. \tag{10}$$

This result falls in the range of published values that show a large variation: Hartman and Jarzynski[21] who used the immersion technique reported $\alpha\lambda=0.022$ Np, Asay *et al.*[13] measured $\alpha\lambda=0.020$ Np using the pulse interference technique, and Kono[22] found $\alpha\lambda=0.044$ Np also using the immersion technique. It is noted that the difference in these measurement techniques does not seem to cause this large variation; probably the large variation in the physical properties and chemical compositions of polymeric materials is responsible.

There are a number of factors that may influence the attenuation measurement result, including the viscosity of couplant, contact pressure, and roughness and parallelism of sample. A full discussion on the influences of all these factors will be given in a separate paper; here, only the effects of couplant and contact pressure are briefly discussed. Figure 7 shows attenuation coefficients measured using two different couplants and loose and tight contact conditions. The two couplants used are the low viscosity oil (45.4 cST at 40 °C, couplant 1) and vacuum grease having a viscosity of 2 000 000 cST at 25 °C (couplant 2). The transducers and sample are hand-tightened. Figure 7 shows that the attenuations for the two different couplants are very close in the frequency range 2–5 MHz while they are a bit different at frequencies out of this range. This result demonstrates that the proposed method can successfully compensate the variation of couplant (viscosity). The dotted line in Fig. 7 corre-

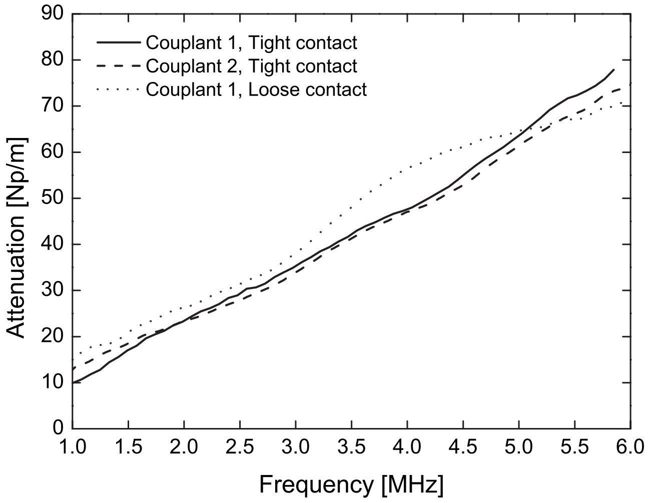


FIG. 7. Measured longitudinal wave attenuation coefficient of PMMA using two different couplants and under loose and tight contact conditions.

sponds to the attenuation measured with a loose contact condition. This condition is produced by untightening the clamping screw such that the transducers are still at the same position, but can slide on the sample surface by a small force (a quantitative device is not used to measure the contact pressure). The attenuation measured under this condition visibly deviates from the other two and exhibits undulation as frequency increases. This deviation may be attributed to some perturbation in alignment due to the loose clamping. Theoretically, any boundary condition can be compensated in the proposed method; however, there seems to be certain limitations in applying the method, which requires further investigation. This result provides one simple instruction that a transducer should be in a tight contact with the sample under examination in order to get a consistent and reliable result. These performed reference measurements demonstrate partially the accuracy of the proposed measurement procedure.

B. Attenuation coefficient of cement paste

The overall objective of the present research is to find the correlation between the measured ultrasonic attenuation and the microstructure of cement-based materials. The proposed measurement procedure has been used to assess the high longitudinal wave attenuation of pure cement paste samples and cement paste samples with different amounts of sand inclusions. This paper presents the result for the pure cement paste sample. Details about the sample used in these measurements can be found in Ref. 23. The attenuations for the cement paste used are input parameters for simulating the attenuation in these materials, defining the matrix material absorption of the composites (concrete) considered in this research.

As seen in Fig. 8, the longitudinal wave attenuation of the cement paste increases linearly with frequency. This behavior is similar to the hysteresis absorption phenomenon in polymeric materials[21] and is observed in cement paste by Punurai *et al.*[24] The attenuation of cement paste that is fitted by a linear regression is $\alpha = 16.18f - 10.19$.

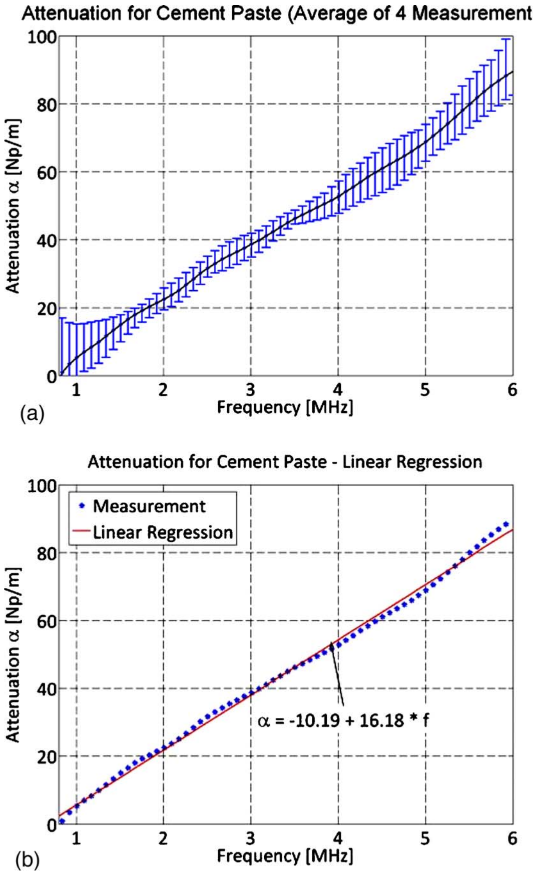


FIG. 8. (Color online) Longitudinal attenuation coefficient for cement paste. (a) Result from four independent measurements. Error bars represent the maximum and minimum values at each measurement frequency. (b) Average of the four measurements with the linear regression.

VII. CONCLUSION

This paper describes the influence of the partial reflection at the specimen and transducer interface in contact attenuation measurements, and proposes an experimental method to measure the actual reflection coefficient and to include a reflection correction in the attenuation coefficient. The reflection coefficient is presented as a quantitative acoustic measure of the current coupling condition of the measuring transducers. It is demonstrated that failure to account for these reflection effects causes a large overestimation, and a large variation in the measured material attenuation coefficient. This paper presents a combined measurement procedure that integrates the reflection coefficient measurement into the attenuation measurement, which removes the overestimation of the attenuation coefficient, and significantly reduces variations in the resulting attenuation coefficient. The accuracy and robustness of the measurement procedure are demonstrated for PMMA, a reference material with a known attenuation coefficient. The proposed measurement technique has shown to improve the accuracy and reliability of contact attenuation measurements.

ACKNOWLEDGMENTS

This work was partially supported by the Georgia Department of Transportation, GTI Project "In Situ Measurement of Air Content in Rigid Pavements" and by the German Academic Exchange Service (DAAD).

[1] A. G. Evans, B. R. Tittman, L. Ahlberg, B. T. Khuri-Yakub, and G. S. Kino, “Ultrasonic attenuation in ceramics,” J. Appl. Phys. **49**, 2669–2679 (1978).

[2] S. Kenderian, T. P. Berndt, R. E. Green, Jr., and B. B. Djordjevic, “Ultrasonic monitoring of dislocations during fatigue of pearlitic rail steel,” Mater. Sci. Eng., A **348**, 90–99 (2003).

[3] H. Ogi, M. Hirao, and K. Minoura, “Noncontact measurement of ultrasonic attenuation during rotation fatigue test of steel,” J. Appl. Phys. **81**, 3677–3684 (1997).

[4] B. Hartmann and J. Jarzynski, “Immersion apparatus for ultrasonic measurements in polymers,” J. Acoust. Soc. Am. **56**, 1469–1477 (1974).

[5] M. N. Toksoz, D. H. Johnston, and A. Timur, “Attenuation of seismic waves in dry and saturated rocks: I. Laboratory measurements,” Geophysics **44**, 681–690 (1979).

[6] F. M. Sears and B. P. Bonner, “Ultrasonic attenuation measurement by spectral ratios utilizing signal processing technique,” IEEE Trans. Geosci. Remote Sens. **GE-19**, 95–99 (1981).

[7] E. P. Papadakis, “Ultrasonic attenuation in thin specimens driven through buffer rods,” J. Acoust. Soc. Am. **44**, 724–734 (1968).

[8] J. Kushibiki, R. Okabe, and M. Arakawa, “Precise measurements of bulk-wave ultrasonic velocity dispersion and attenuation in solid materials in the VHF range,” J. Acoust. Soc. Am. **113**, 3171–3178 (2003).

[9] M. Redwood and J. Lamb, “On the propagation of high-frequency compressional waves in isotropic cylinders,” Proc. Phys. Soc. London, Sect. B **70**, 136–143 (1957).

[10] M. Redwood, “Dispersion effects in ultrasonic waveguides and their importance in the measurement of attenuation,” Proc. Phys. Soc. London, Sect. B **70**, 721–737 (1957).

[11] H. J. McSkimin, “Propagation of longitudinal waves and shear waves in cylindrical rod at high frequencies,” J. Acoust. Soc. Am. **28**, 484–493 (1956).

[12] R. Treull, C. Elbaum, and B. B. Chick, *Ultrasonic Methods in Solid State Physics* (Academic, New York, 1968).

[13] J. R. Asay, D. L. Lamberson, and A. H. Guenther, “Pressure and temperature dependence of the acoustic velocities in polymethylmethacrylate,” J. Appl. Phys. **40**, 1768–1783 (1969).

[14] H. L. McSkimin, “Pulse superposition method for measuring ultrasonic velocity in solids,” J. Acoust. Soc. Am. **33**, 12–23 (1961).

[15] W. P. Mason and H. J. McSkimin, “Attenuation and scattering of high frequency sound waves in metals and glasses,” J. Acoust. Soc. Am. **19**, 464–473 (1947).

[16] H. Seki, A. Granato, and R. Truell, “Diffraction effects in the ultrasonic field of a piston source and their importance in the accurate measurement of attenuation,” J. Acoust. Soc. Am. **28**, 230–238 (1956).

[17] E. P. Papadakis, “Correction for diffraction losses in the ultrasonic field of a piston source,” J. Acoust. Soc. Am. **31**, 150–152 (1959).

[18] A. I. Lavrentyev and S. I. Rokhlin, “Ultrasonic spectroscopy of imperfect contact interfaces between a layer and two solids,” J. Acoust. Soc. Am. **103**, 657–664 (1998).

[19] J. Zhang, B. W. Drinkwater, and R. S. Dwyer-Joyce, “Acoustic measurement of lubrication-film thickness distribution in ball bearings,” J. Acoust. Soc. Am. **119**, 863–871 (2006).

[20] P. H. Rogers and A. L. Van Buren, “Exact expression of Lommel diffraction correction integral,” J. Acoust. Soc. Am. **55**, 724–728 (1974).

[21] B. Hartmann and J. Jarzynski, “Ultrasonic hysteresis absorption in polymers,” J. Appl. Phys. **43**, 4304–4312 (1972).

[22] R. Kono, “The dynamic bulk viscosity of polystyrene and polymethyl methacrylate,” J. Phys. Soc. Jpn. **15**, 718–725 (1960).

[23] The cement paste sample was cast from commercial type I Portland cement powder into cylinders of 76.2 mm diameter. The powder was mixed with water, at a water to cement mass ratio of 0.4, utilizing a Hobart mixer before a vibration table is used to diminish the amount of entrapped air. The cylinder samples were demolded after 24 h and then left in a water bath for hydration for 14 days. The cylinders were finally cut into disks with different thicknesses and surfaces were treated with a diamond polishing paper.

[24] W. Punurai, J. Jarzynski, J. Qu, K. E. Kurtis, and L. Jacobs, “Characterization of entrained air voids with scattered ultrasound,” NDT & E Int. **39**, 514–524 (2006).

Dynamic equations for fluid-loaded porous plates using approximate boundary conditions

Peter D. Folkow[a] and Martin Johansson
Department of Applied Mechanics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

(Received 19 August 2008; revised 27 January 2009; accepted 28 January 2009)

Systematically derived equations for fluid-loaded thin poroelastic layers are presented for time-harmonic conditions. The layer is modeled according to Biot theory for both open and closed pores. Series expansion techniques in the thickness variable are used, resulting in separate symmetric and antisymmetric plate equations. These equations, which are believed to be asymptotically correct, are expressed in terms of approximate boundary conditions and can be truncated to arbitrary order. Analytical and numerical results are presented and compared to the exact three dimensional theory and a flexural plate theory. Numerical comparisons are made for two material configurations and two thicknesses. The results show that the presented theory predicts the plate behavior accurately. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3086267]

PACS number(s): 43.40.Dx, 43.20.Gp, 43.20.Tb, 43.40.Rj [KVH] Pages: 2954–2966

I. INTRODUCTION

The behavior of wave propagation phenomena in interface layers is of great interest in many engineering areas, for instance, acoustics, electromagnetics, and elastodynamics. When the layers are thin, i.e., the layer thickness is small in comparison to the wavelengths involved, there are many approximate theories that may be applicable. As these theories may greatly simplify the analysis, thin layer theories have been studied extensively in the literature.

In elastodynamics, more or less systematically derived approximate thin layer theories have been developed for all sorts of geometries and material configurations, e.g., isotropic, anisotropic, piezoelectric, and porous media. There is much debate on the correctness of the simplifying kinematic assumptions which constitute the base for these theories. For more complex layer structures, such as porous layers, it becomes more crucial to choose the most appropriate simplifications. This calls for a systematic derivation approach, which is the main purpose of the present paper.

Porous layers are found in several disciplines such as biomechanics, seismology, geomechanics, and acoustics; in the latter case the noise absorbing effects are of special importance in engineering. The structure of porous media is very complex and irregular, being composite and multiphase. For macroscopic scales, i.e., when the relevant length scales are much larger than the pore sizes, a widely used theory is due to Biot.[1] It is a three dimensional (3D) linear theory that treats the porous material as a continuum in terms of averaged macroscopic displacement fields. The equations are formulated using a set of mechanical and geometrical parameters and much work has been devoted to the determination of these parameters from experiments.

The complexity of porous materials has motivated approximations and simplified methods; one such being the rigid frame model.[2,3] However, for thin layers several different elastic plate theories of various complexities have been derived. To our knowledge, all existing models are based on the Kirchhoff plate hypothesis combined with the Biot theory, see, for example, Refs. 4–6. Numerical solutions to such porous plate equations using bending theories for finite plates are obtained by Leclaire *et al.*[7] for acoustical excitation, Etchassahar *et al.*[8] for a concentrated force, and Aygun *et al.*[9] for fluid loading. Moreover, the special case of in-plane fluid flow was treated by Li *et al.*[10] Hence, so far there seems to be a lack in the literature on more advanced porous plate theories, e.g., based on the Mindlin hypothesis or some other more refined theory.

In this paper, the attention is on the derivation of the governing dynamic equations for a fluid-loaded saturated porous layer. The method used is based on a series expansion technique, previously used on elastic rods,[11] elastic plates,[12–15] and piezoelectric plates.[16,17] Starting from the 3D Biot theory, the displacement fields and boundary conditions are expanded using power series in the thickness coordinate. This procedure is performed in a systematic manner, resulting in symmetric and antisymmetric plate equations involving terms up to an (in principle) arbitrary order. There are good reasons to believe that the approach leads to asymptotically correct equations, without using *ad hoc* assumptions.[18] Explicit expressions suitable for thin layer approximations, including terms up to order 3 in the thickness, are presented analytically and compared to the corresponding expressions using the Theodorakopoulos and Beskos (TB) theory.[5] In order to illustrate the results, numerical examples are presented for two material combinations: water and a quartz-fiber, and air in combination with a plastic foam.

The plate equations in question are written in terms of so called approximate (effective) boundary conditions. Hereby, the effects from the porous layer are replaced by differential equations in the surface plane, expressed in terms of the exterior fluid variables. These differential equations that account for the influence from the porous layer constitute the approximate boundary conditions. It is thus not necessary to model and solve equations for the porous medium separately,

[a] Electronic mail: peter.folkow@chalmers.se

© 2009 Acoustical Society of America

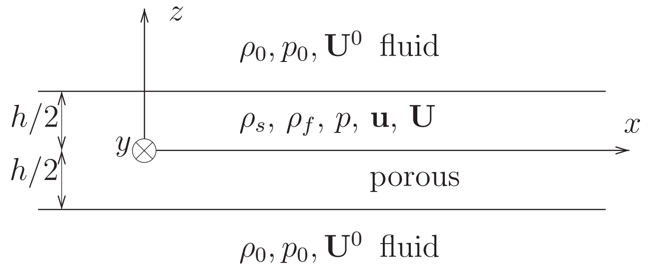

z
$\rho_0, p_0, \mathbf{U}^0$ fluid
$h/2$
$\rho_s,\ \rho_f,\ p,\ \mathbf{u},\ \mathbf{U}$
x
y
$h/2$
porous
$\rho_0, p_0, \mathbf{U}^0$ fluid

FIG. 1. Problem geometry.

which simplifies the analysis of porous-fluid coupled problems. Approximate boundary conditions have been used in many areas, for instance, acoustics,[19] electromagnetics,[20] and elastodynamics.[15,16,21,22]

This paper is organized as follows: The governing equations of the Biot theory are presented in Sec. II. In Sec. III an outline of the derivation method is given, followed by the derivation of the asymptotic approximate boundary conditions for the porous layer in the case of closed and open pores, respectively. Numerical results are presented in Sec. IV, where the transmission and damping coefficients are calculated for incident plane waves. The results are compared to the full 3D Biot theory as well as the TB theory. This paper is summarized in Sec. V. The Appendix shows the asymptotic plate equations for the antisymmetric case in terms of the plate displacement.

II. ASSUMPTIONS AND GOVERNING EQUATIONS

Consider an infinite, isotropic porous plate of thickness h immersed in a fluid, see Fig. 1. The saturated plate is modeled according to the linear Biot theory.[1] For the continuum theory to hold, the plate thickness is large when compared to the pore size. However, for the simplified plate equations to hold, the layer is assumed to be thin, i.e., the plate thickness is assumed to be smaller than the shortest wavelength considered. These assumptions are in line with the low-frequency theory[1] where the flow in the pores is considered to obey Poiseuille flow.

The porous plate densities of the solid and fluid parts are ρ_s and ρ_f, respectively. The state of the plate may be expressed through the solid macroscopic displacement $\mathbf{u}$ together with the average displacement $\mathbf{U}$ and the pressure p of the liquid phase. The state of the surrounding fluid, density ρ_0, is modeled through the pressure p_0 and the displacement $\mathbf{U}^0$. Viscous effects are neglected in the exterior fluid, but they are important in the porous domain.

The pores at the plate boundaries are either closed or fully open; the intermediary states are not considered here. With closed pores it is understood that a thin, impermeable massless layer separates the exterior fluid from the fluid in the plate. The properties of the saturating fluid and the surrounding fluid may in such a case differ. For open pores, fluid may flow between the plate and the surroundings through the boundaries. It is then understood that the saturating fluid is identical to the surrounding fluid.

Biot theory supports three bulk wave types. One shear wave (wave speed c_s) and two compressional waves referred to as the fast and slow compressional waves (wave speeds c_{p1} and c_{p2}, respectively). Here, it is the slowest wave that is the critical one for thin layer approximations. Which of the three waves that is the slowest depends on the frame and saturating fluid materials as well as the frequency.

A. Governing equations

Considering time-harmonic conditions, the equations of motion governing the displacements in porous media may be written as[1]

$$\begin{aligned} &N\nabla^2\mathbf{u} + \nabla\nabla\cdot((N+A)\mathbf{u}+Q\mathbf{U}) = -\omega^2(\rho_{11}\mathbf{u}+\rho_{12}\mathbf{U}),\\ &\nabla\nabla\cdot(Q\mathbf{u}+R\mathbf{U}) = -\omega^2(\rho_{12}\mathbf{u}+\rho_{22}\mathbf{U}), \end{aligned} \tag{1}$$

where the factor $e^{-i\omega t}$ has been omitted. Here N is the shear modulus of the frame and Q, A, and R are generalized elastic coefficients. These latter factors are related to the measurable quantities: the porosity Φ, the bulk modulus of the frame K_b, the bulk modulus of the solid material K_s, and the bulk modulus of the fluid K_f, see Ref. 2:

$$\begin{aligned} &A = \frac{(1-\Phi)(1-\Phi-K_b/K_s)K_s + \Phi K_s K_b/K_f}{C_K} - \tfrac{2}{3}N,\\ &Q = \frac{\Phi(1-\Phi-K_b/K_s)K_s}{C_K}, \quad R = \frac{\Phi^2 K_s}{C_K}, \end{aligned} \tag{2}$$

where

$$C_K = 1-\Phi-K_b/K_s+\Phi K_s/K_f. \tag{3}$$

The densities ρ_{ij} are defined by

$$\begin{aligned} &\rho_{12} = (1-\alpha)\Phi\rho_f, \quad \rho_{11} = (1-\Phi)\rho_s - \rho_{12},\\ &\rho_{22} = \Phi\rho_f - \rho_{12}, \end{aligned} \tag{4}$$

where the density ρ_{12} represents added mass due to coupling of the solid and fluid motions. Here α is the dynamic tortuosity, related to both the porosity and the geometry of the interconnected pores. For viscous flow in the pores causing damping, α is complex and frequency dependent according to the model by Johnson *et al.*[23]

$$\alpha = \alpha_\infty + \frac{i\eta\Phi}{\rho_f\kappa_0\omega}\left(1-\frac{4i\alpha_\infty^2\kappa_0^2\rho_f\omega}{\eta\Phi^2\Lambda^2}\right)^{1/2}, \tag{5}$$

where η is the viscosity, Λ is a characteristic length of the pore size, and κ_0 is the static permeability which may be written as the quotient between the viscosity and the flow resistivity. For non-viscous fluids the tortuosity is given by the constant value $\alpha=\alpha_\infty$. This choice of notation is due to the limit $\omega\to\infty$ for α in the viscous case. Due to the behavior of α for viscous flow, the densities ρ_{ij} are hereby also complex and frequency dependent. Note that when the saturating fluid is a gas, thermal effects should be considered as the thermal conduction modifies the bulk modulus of the fluid.

The constitutive equations for the stresses in the porous plate are

$$\boldsymbol{\sigma}^s = N(\nabla\mathbf{u}+(\nabla\mathbf{u})^T) + \mathbf{I}\,\nabla\cdot(A\mathbf{u}+Q\mathbf{U}), \tag{6}$$

$$\boldsymbol{\sigma}^f = (-\mathbf{I}\Phi p) = \mathbf{I}\,\nabla\cdot(Q\mathbf{u}+R\mathbf{U}), \tag{7}$$

where $\mathbf{I}$ is the 3×3 identity tensor, $\boldsymbol{\sigma}^s$ is the averaged stress of the solid frame, and $\boldsymbol{\sigma}^f$ is the averaged stress of the fluid portion. The sum of these stresses forms the total macroscopic stress $\boldsymbol{\sigma}=\boldsymbol{\sigma}^s+\boldsymbol{\sigma}^f$.

In the exterior fluid the governing equation for the pressure is according to the simple wave equation

$$(\nabla^2+k_0^2)p_0=0, \quad k_0=\omega/c_0, \tag{8}$$

where c_0 is the wave velocity and k_0 is the wave number for waves in the fluid. The relation between the pressure p_0 and the displacement $\mathbf{U}^0$ in the surrounding fluid is given by the momentum equation

$$\nabla p_0 = \rho_0\omega^2\mathbf{U}^0. \tag{9}$$

B. Boundary conditions

The boundary conditions at $z=\pm h/2$ for a porous-fluid interface stem from continuity of traction, continuity of filtration velocity, as well as Darcy's law.[24] The continuity requirements for the surface tractions in tangential and normal directions are obtained using Eqs. (6) and (7),

$$N(\partial_x u_z+\partial_z u_x)=0, \tag{10}$$

$$N(\partial_y u_z+\partial_z u_y)=0, \tag{11}$$

$$2N\partial_z u_z+\nabla\cdot((A+Q)\mathbf{u}+(Q+R)\mathbf{U})=-p_0. \tag{12}$$

Here, the partial derivatives are expressed as $\partial_x=\partial/\partial_x$ and so on. The fluid flow across the interfaces $z=\pm h/2$ is governed by continuity of the normal component of the filtration vector and by Darcy's law. Under time-harmonic conditions these relations are

$$\Phi(U_z-u_z)=U_z^0-u_z, \tag{13}$$

$$\pm i\omega\Phi(U_z-u_z)=\kappa_s(p_0-p). \tag{14}$$

Here the parameter κ_s characterizes the permeability of the interface and is essentially the static permeability per unit length divided by the viscosity. The permeability is related to dissipation effects at the boundaries and is as such inverse proportional to the flow resistivity. Here the open pore case implies free flow which results in infinite permeability, while a sealed interface results in zero permeability. In practice,[24] most interfaces are partially open where $0<\kappa_s<\infty$. The state of the porous plate at the boundaries may be given in terms of either the fields $\{\mathbf{u},\mathbf{U}\}$ or $\{\mathbf{u},p\}$, while for the surrounding fluid the boundary conditions are given by either $\mathbf{U}^0$ or p_0. The choice of representation is a matter of convenience, and here the $\{\mathbf{u},\mathbf{U},p_0\}$-formulation is used in conformity with the governing equations for the porous plate [Eq. (1)] and the surrounding fluid [Eq. (8)]. Thus, the pressure of the liquid phase p appearing in Darcy's law [Eq. (14)] may be rewritten using Eq. (7). In a similar manner, the surrounding fluid displacement U_z^0 in Eq. (13) may be written in terms of the pressure field p_0 adopting Eq. (9). Since the results of the fluid flow boundary conditions [Eqs. (13) and (14)] are different for closed and open pores, these cases are discussed separately below.

For *closed pores* the permeability κ_s is zero as no mass transport is allowed between the exterior fluid and the fluid of the porous medium. Darcy's law [Eq. (14)] then reduces to $U_z=u_z$ at the surfaces. This simplifies the filtration condition [Eq. (13)] to $U_z^0=u_z$ as expected. By using Eq. (9), the interface fluid flow relations become

$$\partial_z p_0=\rho_0\omega^2 u_z, \tag{15}$$

$$0=U_z-u_z. \tag{16}$$

In the case of *open pores* the permeability κ_s is infinite. Due to finiteness of the flow, Darcy's law [Eq. (14)] gives $p=p_0$ at the boundaries. Adopting Eq. (9) in Eq. (13), as well as Eq. (7), gives the interface fluid flow relations

$$\partial_z p_0=\rho_0\omega^2((1-\Phi)u_z+\Phi U_z), \tag{17}$$

$$p_0=-\Phi^{-1}\,\nabla\cdot(Q\mathbf{u}+R\mathbf{U}). \tag{18}$$

Here, $\rho_f=\rho_0$ as the same fluid is assumed inside and outside the plate.

III. PLATE EQUATIONS

A. Series expansions

There are different ways to obtain plate theories, either based on different kinematical assumptions in line with the classical theories for elastic plates or in a more rigorous fashion adopting the 3D equations of motion. Here, a systematic approach will be used based on the 3D equations of motion together with power series expansions of the physical fields with respect to the thickness coordinate z. This method is believed to be asymptotically correct without any *ad hoc* assumptions, resulting in a hierarchy of higher order plate theories that can (in principle) be truncated to any order. One such expansion method is to expand the plate displacement fields around the midplane $z=0$, which has been done for an isotropic elastic plate, see Ref. 14. Another method is to expand the boundary conditions around the midplane $z=0$. This approach has been adopted by Johansson *et al.*[15] for a fluid-loaded elastic plate, where the results are expressed in terms of so called approximate boundary conditions. These two methods seem to be analogous, which is discussed in the case of an elastic plate.[15]

In the present case for fluid-loaded porous plates, the latter approach is used in order to benefit from the results from fluid-loaded elastic plates.[15] As in the cited work, the plate equations are expressed in terms of approximate boundary conditions where the plate fields are eliminated so as to obtain differential equations in terms of the fluid pressure p_0 and its normal derivative $\partial_z p_0$ at the boundaries. By doing this, the influence from the porous plate is present implicitly in the differential equations, without the need of solving the plate equations explicitly in terms of the plate fields. For completeness, the antisymmetric plate equations are also given in terms of the standard plate displacement, see Appendix. Now, when using series expansion of the boundary conditions in order to eliminate the plate fields, it is convenient to proceed in terms of the differences and sums of the boundary fields

$$\begin{aligned}\Delta f &= f(x,y,h/2) - f(x,y,-h/2),\\ \Sigma f &= f(x,y,h/2) + f(x,y,-h/2),\end{aligned} \tag{19}$$

where $f=\{\mathbf{u},\mathbf{U},p_0\}$ and their spatial derivatives. For the *plate fields* $\mathbf{u}$ and $\mathbf{U}$ and their spatial derivatives, the sums and differences are expanded in Maclaurin series in the thickness coordinate z,

$$\Delta f = 2\sum_{j=0}^{n-1} \frac{\partial_z^{2j+1} f_c}{(2j+1)!}\left(\frac{h}{2}\right)^{2j+1} + \mathcal{O}(h^{2n+1}), \tag{20}$$

$$\Sigma f = 2\sum_{j=0}^{n-1} \frac{\partial_z^{2j} f_c}{(2j)!}\left(\frac{h}{2}\right)^{2j} + \mathcal{O}(h^{2n}), \tag{21}$$

where $\partial_z^m f_c = \partial_z^m f(x,y,z)|_{z=0}$. The sums and differences of the boundary conditions for the traction [Eqs. (10)–(12)] and the fluid flow [Eqs. (15) and (16)] (closed pores) or Eqs. (17) and (18) (open pores) may then be written in terms of the sums and differences of the exterior fluid pressure p_0 and $\partial_z p_0$, together with the porous field variables at the center plane $\mathbf{u}_c$ and $\mathbf{U}_c$.

In conformity with Johansson *et al.*,[15] two separate cases are identified due to the differential orders in the normal direction of the plate displacements. For Δp_0 and $\Sigma \partial_z p_0$, even order z-derivatives ∂_z^{2m} act on the fields $\{u_{z,c}, U_{z,c}\}$ and their derivatives in the (xy)-plane, while odd order z-derivatives ∂_z^{2m+1} act on the fields $\{u_{x,c}, u_{y,c}, U_{x,c}, U_{y,c}\}$ and their derivatives in the (xy)-plane. Considering Σp_0 and $\Delta \partial_z p_0$ the opposite situation holds. The explicit representation of the plate fields $\{\mathbf{u}_c, \mathbf{U}_c\}$ and their spatial derivatives may then be eliminated in order to give separate differential equations in terms of Δp_0 and $\Sigma \partial_z p_0$ (antisymmetric case) and Σp_0 and $\Delta \partial_z p_0$ (symmetric case). For each such case, $8n+7$ equations must be solved if the expansions in the plate thickness h are to involve terms up to and including h^{2n+1}. Since there are only five boundary conditions, the remaining equations are found through the governing equations [Eq. (1)]. By performing normal derivatives of these equations evaluated at the center plane, $8n+2$ additional equations are obtained.

The solutions of the equation systems discussed above, resulting in the approximate boundary conditions, may be written in terms of differential operators. For closed pores these boundary conditions are expressed as

$$\left(\sum_{j=0}^{n-1} h^{2j}\omega^2 \mathcal{C}_{1,j}\right)\Delta p_0 = \left(\sum_{j=1}^{n} h^{2j-1}\mathcal{C}_{2,j}\right)\Sigma \partial_z p_0 + \mathcal{O}(h^{2n}), \tag{22}$$

$$\left(\sum_{j=2}^{n} h^{2j}\omega^2 \mathcal{C}_{3,j}\right)\Sigma p_0 = \left(\sum_{j=2}^{n} h^{2j-1}\mathcal{C}_{4,j}\right)\Delta \partial_z p_0 + \mathcal{O}(h^{2n+1}) \tag{23}$$

for the antisymmetric and symmetric cases, respectively, while for open pores the equations are

$$\left(\sum_{j=1}^{n-1} h^{2j}\mathcal{C}_{5,j}\right)\Delta p_0 = \left(\sum_{j=1}^{n-1} h^{2j+1}\mathcal{C}_{6,j}\right)\Sigma \partial_z p_0 + \mathcal{O}(h^{2n}), \tag{24}$$

$$\left(\sum_{j=2}^{n} h^{2j}\mathcal{C}_{7,j}\right)\Sigma p_0 = \left(\sum_{j=1}^{n-1} h^{2j+1}\mathcal{C}_{8,j}\right)\Delta \partial_z p_0 + \mathcal{O}(h^{2n+1}) \tag{25}$$

in the antisymmetric and symmetric cases, respectively. The differential operators $\mathcal{C}_{i,j}$ generally involve $\nabla_s^{2p}\omega^{2j-2p}$ for $p=0,\ldots,j$ with $\nabla_s^2 = \partial_x^2 + \partial_y^2$. Closed form representations of $\mathcal{C}_{i,j}$ involve a substantial number of different terms even at low order. It should be noted that for a series expansion involving terms up to and including order h^{2n+1} in Eqs. (20) and (21), even higher order terms appear in the elimination process. However, as these higher order terms are not stable when using even more terms in the series expansion, the equations are truncated. Thus, only terms up to and including order h^{2n+1} are present in the final result given above. In the present work, series expressions involving terms up to and including order h^3 are given explicitly. These closed form expressions are obtained through tedious calculations using the mathematical software MATHEMATICA.[25] It is straightforward to derive even higher order expansions, but such lengthy results are not presented here.

B. Closed pores

Expressing relations (22) and (23) explicitly for terms up to and including h^3, the approximate boundary condition for the *antisymmetric* case becomes

$$\begin{aligned}&\beta_1 k_t^2\left\{1 - \frac{1}{2}\left(\frac{h}{2}\right)^2[3\nabla_s^2 + 3k_{\mathrm{av}}^2]\right\}\Delta p_0\\ &\quad= \left\{\frac{h}{2}k_t^2 - \frac{1}{6}\left(\frac{h}{2}\right)^3[8(1-\gamma_1)\nabla_s^4 + (16k_s^2 + 7k_t^2 - 24k_{\mathrm{av}}^2\right.\\ &\qquad - 8k_1^2 + 8k_{p1}^2 k_{p2}^2/k_s^2)\nabla_s^2 + (2k_s^2 + 3k_{\mathrm{av}}^2)k_t^2\\ &\qquad\left. + 2\gamma_2 k_{p1}^2 k_{p2}^2]\right\}\Sigma \partial_z p_0,\end{aligned} \tag{26}$$

while the approximate boundary condition for the *symmetric* case is

$$\begin{aligned}&\beta_1 k_t^2(\nabla_s^2 + k_{p1}^2)(\nabla_s^2 + k_{p2}^2)\left\{\frac{h}{2} - \frac{1}{6}\left(\frac{h}{2}\right)^3[3\nabla_s^2 + 3k_{\mathrm{av}}^2]\right\}\Sigma p_0\\ &\quad= \left\{[-4(1-\gamma_1)\nabla_s^4 - (k_t^2 - 4k_1^2)\nabla_s^2 - \gamma_2 k_{p1}^2 k_{p2}^2]\right.\\ &\qquad + \frac{1}{6}\left(\frac{h}{2}\right)^2[12(1-\gamma_1)\nabla_s^6 + (5k_t^2 + 12(1-\gamma_1)k_{\mathrm{av}}^2\\ &\qquad - 12k_1^2)\nabla_s^4 + ((9k_{\mathrm{av}}^2 - 2k_s^2)k_t^2 - 12k_1^2 k_{\mathrm{av}}^2\\ &\qquad\left. + 3\gamma_2 k_{p1}^2 k_{p2}^2)\nabla_s^2 + (2k_t^2 + 3\gamma_2 k_{\mathrm{av}}^2)k_{p1}^2 k_{p2}^2]\right\}\Delta \partial_z p_0.\end{aligned} \tag{27}$$

Here k_s, k_{p1}, and k_{p2} are the bulk wave numbers of the shear wave, the fast compressional wave, and the slow compressional wave, respectively. They are given by

$$k_{p1,p2}^2 = \frac{\omega^2}{c_{p1,p2}^2} = \frac{2a_1\omega^2}{a_2 \mp \sqrt{a_2^2 - 4a_1a_3}}, \quad k_s^2 = \frac{\omega^2}{c_s^2} = \frac{a_1\omega^2}{N\rho_{22}}, \tag{28}$$

where the constants a_i are

$$a_1 = \rho_{11}\rho_{22} - \rho_{12}^2, \quad a_2 = R\rho_{11} - 2Q\rho_{12} + S\rho_{22},$$
$$a_3 = RS - Q^2, \tag{29}$$

using $S=2N+A$. The wave number k_t is a generalized shear wave number, k_{av} is the mean square of the bulk wave numbers, and k_1 is an auxiliary wave number. They are defined by

$$k_t^2 = \frac{\rho_t\omega^2}{N}, \quad k_{\mathrm{av}}^2 = \frac{k_{p1}^2 + k_{p2}^2 + k_s^2}{3},$$
$$k_1^2 = \frac{((Q+R)\rho_{12} - (Q+S-N)\rho_{22})\omega^2}{a_3}. \tag{30}$$

The density ρ_t is the total density of the saturated porous body according to

$$\rho_t = \rho_{11} + 2\rho_{12} + \rho_{22} = (1-\Phi)\rho_s + \Phi\rho_f. \tag{31}$$

All the squared wave numbers in Eqs. (26) and (27) have a positive real part. The symbols β_1, γ_1, and γ_2 are nondimensional numbers (with positive real part) expressed in terms of densities and generalized elastic coefficients according to

$$\beta_1 = \frac{\rho_0}{\rho_t}, \quad \gamma_1 = \frac{NR}{a_3}, \quad \gamma_2 = \frac{R+2Q+S}{N}. \tag{32}$$

Comparing Eqs. (26) and (27) with the corresponding approximate boundary equations for an elastic layer reported by Johansson *et al.*,[15] it is seen that the expressions are quite similar. The differences are most pronounced for the symmetric case, where the derivatives in the porous case are two orders higher than in the elastic case. This has to do with the fact that Biot's theory supports three bulk waves, and evidently all of these are important to order h^3 in Eq. (27). In conformity with Johansson *et al.*,[15] it is possible to factorize the compressional bulk wave operator in the symmetric case. As there are two such waves in the porous case, the factorized operator is $(\nabla_s^2+k_{p1}^2)(\nabla_s^2+k_{p2}^2)$ on the left-hand side of Eq. (27). This implies that when the surface component of the wave number of the fluid loading coincides with the wave numbers of any of the compressional bulk waves in the porous medium, no symmetric normal motion of the fluid-porous boundary is generated. In the porous layer a guided pressure wave may propagate with the wave speed of any of the compressional bulk wave speeds as if the porous domain occupied the full 3D space. Moreover, the symmetrical part of the fluid loading behaves as if the fluid-porous interfaces were rigid walls. It should be noted, though, that the compressional bulk wave numbers are complex when losses are included in the porous medium. This coincidence phenomenon thus does not occur here since viscosity is neglected in the exterior fluid.

It is not difficult to generalize Eqs. (26) and (27) to the case of different exterior fluids as this only influences the normal displacement continuity condition, Eq. (15). The right-hand side of this equation may still be expressed in terms of either the sum or the difference of the normal displacement of the porous solid. Denoting the density of the exterior fluids on the positive and negative z by ρ_0 and ρ_0', respectively, this is done by substituting the right-hand sides of Eqs. (26) and (27) according to

$$\Sigma\partial_z p_0 \to \tfrac{1}{2}((1+\rho_0/\rho_0')\Sigma\partial_z p_0 + (1-\rho_0/\rho_0')\Delta\partial_z p_0),$$

$$\Delta\partial_z p_0 \to \tfrac{1}{2}((1-\rho_0/\rho_0')\Sigma\partial_z p_0 + (1+\rho_0/\rho_0')\Delta\partial_z p_0).$$

It is instructive to see whether the results for a homogeneous plate derived by Johansson *et al.*[15] are retrieved when the porosity tends to zero, $\Phi \to 0$. Starting with the generalized elastic coefficients [Eq. (2)] and the densities [Eq. (4)], these become $N=\mu$, $A=\lambda$, and $\rho_{11}=\rho_s$ while $Q=R=0$ and $\rho_{12}=\rho_{22}=0$ in the limit. Here, μ and λ are the Lamé constants for a homogeneous plate. The bulk wave numbers in Eq. (28) are hereby $k_{p1}^2=\omega^2/c_p^2$, $k_s^2=\omega^2/c_s^2$, and $k_{p2}^2=0$, where now $c_p^2=(\lambda+2\mu)/\rho_s$ and $c_s^2=\mu/\rho_s$. Moreover, the auxiliary wave number in Eq. (30) becomes $k_1=0$ while the constants in Eq. (32) become $\gamma_1=1/\gamma_2=\gamma$, where $\gamma=\mu/(\lambda+2\mu)$. Comparing these equations to the ones given in Ref. 15, a few terms differ. However, by closer inspection of the resulting non-truncated porous equations using a h^3 series expansion, common differential operators may actually be factorized on both sides of the equations when the porosity tends to zero. Such factorized operators stem from the elimination procedure when reducing the number of porous plate parameters to the homogeneous case. Hence, by eliminating these common operators, identical results are obtained when the porosity tends to zero as for the homogeneous plate.[15] The factorized operators in question are

$$\left(1 - \frac{1}{2}\left(\frac{h}{2}\right)^2\nabla_s^2\right), \quad \nabla_s^2\left(1 - \frac{1}{6}\left(\frac{h}{2}\right)^2\nabla_s^2\right) \tag{33}$$

in the antisymmetric and symmetric cases, respectively.

The results using series expansions [Eqs. (26) and (27)] may be compared analytically to other approximate theories. Among the different models based on the Kirchhoff hypothesis, the theory due to TB (Ref. 5) seems to be the most detailed, albeit several ambiguities are found therein. As only bending modes are studied in the TB theory, the results may only be compared to the antisymmetric case [Eq. (26)]. Hence, written in terms of approximate boundary conditions, the TB theory becomes

$$\begin{aligned}\beta_1 k_t^2\{\nabla_s^2 + k_{\mathrm{TB}}^2\}\Delta p_0 &= \left\{\frac{h}{2}[k_s^2(\nabla_s^2 + k_{\mathrm{TB}}^2)] - \frac{1}{6}\left(\frac{h}{2}\right)^3[8(1-\gamma_1)\nabla_s^6\right.\\ &\left.\quad + (8(1-\gamma_1)k_{\mathrm{TB}}^2 - 8k_{p1}^2k_{p2}^2/k_s^2)\nabla_s^4]\right\}\Sigma\partial_z p_0,\end{aligned} \tag{34}$$

where

$$k_{\mathrm{TB}}^2 = k_s^2(1+\gamma_1) - 3k_{\mathrm{av}}^2.$$

It is quite clear that Eqs. (26) and (34) are not very similar. Being based on the Kirchhoff hypothesis, the TB equation does not include the terms of order h^2 on the left-hand side, see comparable situation in the homogeneous case.[15] Note that the TB theory is of order 6 at the highest, while the asymptotic theory goes up to derivatives of order 4. This is surprising since the latter equation may be seen as a generalized Mindlin equation; a statement based on the corresponding situation for a homogeneous plate where the similarities between the Mindlin equation and the asymptotic equation of order h^3 are shown. When the porosity goes to zero in the TB equation, the operator $\nabla_s^2 + k_{\mathrm{TB}}^2$ may be factorized and the approximate boundary condition for the Kirchhoff case is retrieved.[15]

C. Open pores

In the open pore case, relations (24) and (25) are written explicitly for terms up to and including h^3. The approximate boundary condition for the antisymmetric case becomes

$$\begin{aligned}
&\beta_2\Bigg\{k_2^2 - \frac{1}{6}\left(\frac{h}{2}\right)^2[8(1-\gamma_1)\Phi^2\nabla_s^4 + (16\Phi^2k_s^2 + 7k_2^2 - 24\Phi^2k_{\mathrm{av}}^2\\
&\quad - 8k_3^2 + 8\Phi^2k_{p1}^2k_{p2}^2/k_s^2)\nabla_s^2 + (2k_s^2+3k_{\mathrm{av}}^2)k_2^2\\
&\quad + 2\gamma_3k_{p1}^2k_{p2}^2]\Bigg\}\Delta p_0\\
&\quad = \Bigg\{\frac{h}{2}k_s^2 - \frac{1}{6}\left(\frac{h}{2}\right)^3[8(1-\gamma_1)\nabla_s^4 + (13-8\gamma_1)k_s^2\nabla_s^2\\
&\qquad + (2k_s^2+3k_{\mathrm{av}}^2)k_s^2]\Bigg\}\Sigma\partial_z p_0,
\end{aligned}\tag{35}$$

while the approximate boundary condition for the symmetric case is

$$\begin{aligned}
&\beta_2\Bigg\{\left(\frac{h}{2}\right)[4(1-\gamma_1)\Phi^2\nabla_s^4 + (k_2^2-4k_3^2)\nabla_s^2 + \gamma_3k_{p1}^2k_{p2}^2]\\
&\quad - \frac{1}{6}\left(\frac{h}{2}\right)^3[12(1-\gamma_1)\Phi^2\nabla_s^6 + (5k_2^2 + 12(1-\gamma_1)\Phi^2k_{\mathrm{av}}^2\\
&\quad - 12k_3^2)\nabla_s^4 + ((9k_{\mathrm{av}}^2 - 2k_s^2)k_2^2 - 12k_3^2k_{\mathrm{av}}^2 + 3\gamma_3k_{p1}^2k_{p2}^2)\nabla_s^2\\
&\quad + (2k_2^2 + 3\gamma_3k_{\mathrm{av}}^2)k_{p1}^2k_{p2}^2]\Bigg\}\Sigma p_0\\
&\quad = \Bigg\{[-4(1-\gamma_1)\nabla_s^2 - k_s^2] + \frac{1}{6}\left(\frac{h}{2}\right)^2[20(1-\gamma_1)\nabla_s^4\\
&\qquad + (12(3-\gamma_1)k_{\mathrm{av}}^2 - (1+8\gamma_1)k_s^2 - 8k_{p1}^2k_{p2}^2/k_s^2)\nabla_s^2\\
&\qquad + (9k_{\mathrm{av}}^2 - 2k_s^2)\ k_s^2)]\Bigg\}\Delta\partial_z p_0.
\end{aligned}\tag{36}$$

Some new auxiliary wave numbers and non-dimensional numbers enter here, in addition to those defined by Eqs. (28)–(32),

$$\beta_2 = \frac{\rho_0}{\rho_{22}}, \quad \gamma_3 = \frac{R - 2\Phi(Q+R)}{N} + \Phi^2\gamma_2,$$

$$k_2^2 = \frac{(\rho_{22} - 2\Phi^2\rho_f)\omega^2}{N} + \Phi^2k_t^2, \quad k_3^2 = \Phi k_4^2 + \Phi^2k_1^2,$$

$$k_4^2 = \frac{(Q\rho_{22} - R\rho_{12})\omega^2}{a_3}.$$

As in the closed pore case these terms have a positive real part. By inspection of Eqs. (35) and (36) it is seen that similar differential operators appear as in the closed pore case [Eqs. (26) and (27)], albeit in a somewhat more complicated way. Moreover, some of the left-hand side operators in the open pore case happen to be generalizations of the corresponding operators appearing on the right-hand side in the closed pore case. Contrary to Eq. (27), it is not possible to make factorization of the left-hand side of the symmetric equation [Eq. (36)] due to the filtration effects at the surfaces of the porous layer.

The result using series expansion [Eq. (35)] may now be compared analytically to the antisymmetric TB bending theory. Written as an approximate boundary condition, the TB theory becomes

$$\begin{aligned}
&\beta_1\beta_3\{(k_t^2 + 2\Phi\beta_4k_4^2)\nabla_s^2 + k_t^2k_{\mathrm{TB}}^2\}\Delta p_0\\
&\quad = \Bigg\{\frac{h}{2}[(k_s^2 + 2\Phi\beta_2\beta_3(1-\Phi^2\beta_3)k_4^2)\nabla_s^2 + k_s^2k_{\mathrm{TB}}^2]\\
&\qquad - \frac{1}{6}\left(\frac{h}{2}\right)^3[8(1-\gamma_1)\nabla_s^6 + (8(1-\gamma_1)k_{\mathrm{TB}}^2\\
&\qquad - 8k_{p1}^2k_{p2}^2/k_s^2)\nabla_s^4]\Bigg\}\Sigma\partial_z p_0,
\end{aligned}\tag{37}$$

where

$$\beta_3 = \frac{1 - \Phi^2\rho_f/\rho_{22}}{1 - \Phi^2\rho_0/\rho_{22}}, \quad \beta_4 = \rho_t/\rho_{22}.$$

As for the closed pore case, Eqs. (35) and (37) are not very similar where the latter is of higher differential order. Moreover, contrary to the asymptotic theory, there is a clear resemblance between the closed and open pore cases in the TB theory. Here, the open pore case is virtually obtained by adding extra terms involving the porosity. Hence, there are modest differences between these cases according to the TB theory, which is illustrated in Sec. IV.

IV. NUMERICAL RESULTS

In order to illustrate the behavior of the different theories described above, the asymptotic and TB theories are to be compared numerically to the 3D Biot theory. Two material combinations are considered: water and QF-20, a quartz-fiber studied by Johnson *et al.*,[26] and air in combination with a plastic foam studied by Allard *et al.*[2,27] These cited works present material data, of which those needed for the numerics are given in Table I. From now on the saturating and the exterior fluids are assumed to be the same for both boundary conditions (open/closed) for each material combination. Viscous effects are accounted for only in the saturating fluid as such effects are negligible in the surrounding fluid. In the

TABLE I. Material data for QF-20 and water (Ref. 26) and plastic foam and air (Refs. 2 and 27).

Symbol	Unit	QF-20/water	Foam/air
ρ_f	kg/m^3	1000	1.213
ρ_s	kg/m^3	2759	429
Φ	⋯	0.402	0.93
α_∞	⋯	1.89	3.2
N	MPa	7.63	$0.18(1-0.1i)$
K_f	MPa	2.22	Equation (40)
K_s	MPa	36.6	⋯
K_b	MPa	9.47	$0.84(1-0.1i)$
Λ	μm	19.0	28.5
η	Pa s	1.14×10^{-3}	1.84×10^{-5}
κ_0	m^2	1.68×10^{-11}	3.3×10^{-10}
p_{atm}	MPa	⋯	0.10
γ_h	⋯	⋯	1.4
ω_c	1/s	⋯	1760

foam-air case also thermal effects are considered for the saturating fluid as well as damping in the plastic foam.

Due to the viscous effects resulting in frequency dependent material parameters, all three bulk waves are dispersive; the higher the frequency, the higher the velocity. Dispersion is particularly important for the slow compressional wave, whereas the other two modes are virtually unaffected. This is manifested in Fig. 2 where curves are presented for the three waves when the real part of the wavelengths equals $2h$. For the plastic foam case, the slow compressional wave velocity actually becomes higher than the shear wave for higher frequencies. The choice of wavelength used in Fig. 2 corresponds to the first positive interference maximum for normal incidence, which could be seen as a reference upper frequency limit for the approximate theories. Besides being dispersive, the considered porous media are also dissipative. The damping is most pronounced for the slow compressional wave in the low-frequency region; the other waves show dissipation that generally increases with frequency. Due to these frequency dependent effects, it is thus not possible to introduce a non-dimensional frequency variable so that the results are independent of the thickness. Therefore two different thicknesses, $h=0.4$ m and $h=0.04$ m corresponding to different frequency intervals, have been chosen in the numerical results.[28] Hereby both the lower-frequency dissipative regime and the higher-frequency propagating regime of the slow compressional wave are investigated. These thicknesses are also indicated in Fig. 2 as solid lines.

In conformity with Johansson *et al.*,[15] the behavior of the approximate plate equations is studied by calculating the transmission and absorption coefficients. Consider $z<-h/2$ where plane waves propagate in the (xz)-plane and impinge at $z=-h/2$. This results in reflected and transmitted plane waves according to

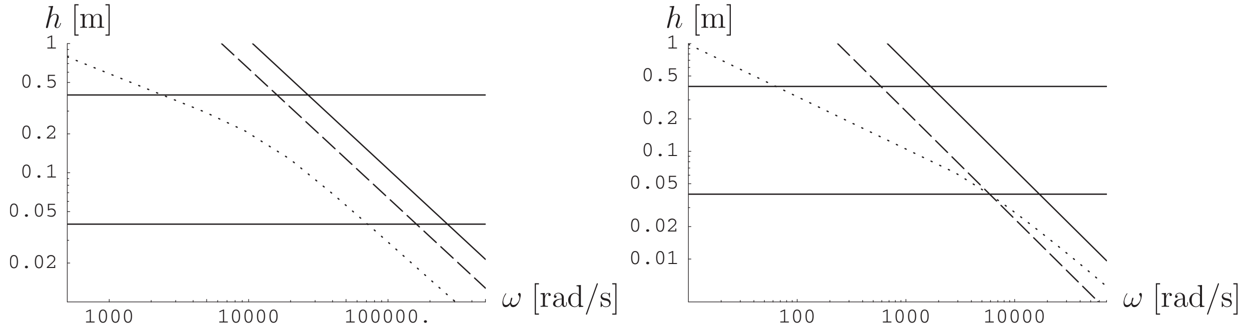
$$p_{0,i} = e^{i(k_0(\zeta x+\sqrt{1-\zeta^2}z)-\omega t)}, \quad z<-h/2,$$

$$p_{0,r} = Re^{i(k_0(\zeta x-\sqrt{1-\zeta^2}z)-\omega t)}, \quad z<-h/2,$$

$$p_{0,t} = Te^{i(k_0(\zeta x+\sqrt{1-\zeta^2}z)-\omega t)}, \quad z>h/2,$$

where k_0 is according to Eq. (9) and $\zeta=\sin\phi$; ϕ being the angle measured from the normal to the plate boundaries. Using the exact 3D equations of motion [Eq. (1)] with pertinent boundary conditions, it is straightforward to solve for the reflection coefficient R, the transmission coefficient T, and the absorption-like coefficient $A_d=1-|R|^2-|T|^2$ which defines the ratio between the energy dissipated by the porous plate and the energy in the incident acoustic wave. For the asymptotic equations, Eqs. (26) and (27) and Eqs. (35) and (36) are to be used for closed and open pores, respectively. Note that purely antisymmetric or symmetric modes will not be generated in the general case, so both the antisymmetric and symmetric equations are to be solved as a system. The numerical results involve solutions based on the asymptotic h and h^3 expansions, respectively. The range of applicability for each truncation level is hereby clearly visible. As for the TB theory, purely antisymmetric motions are considered. This implies that $\Delta\partial_z p_0=0$ resulting in a simple relation between R and T. Consequently, Eqs. (34) and (37) for closed and open pores, respectively, are readily solved.

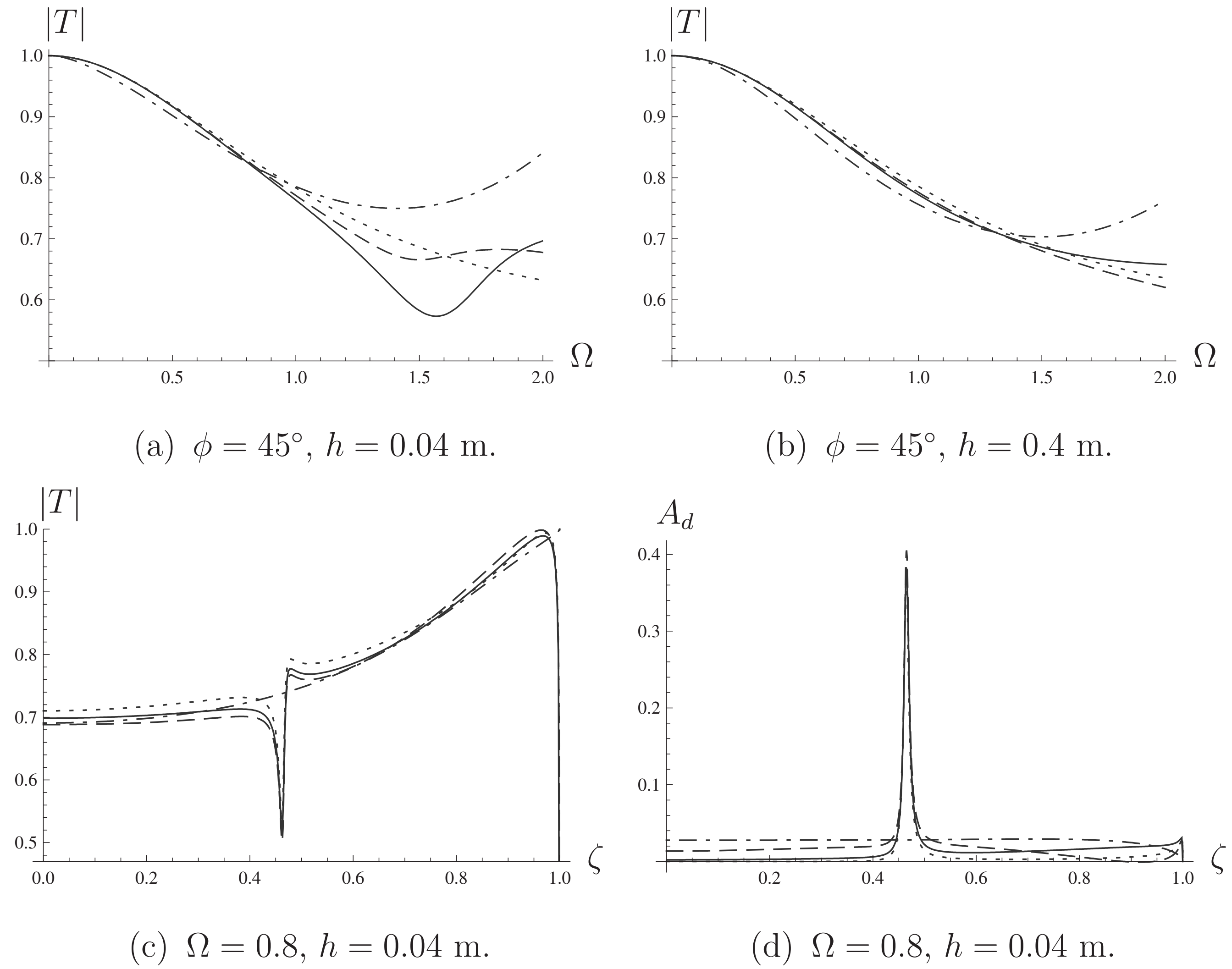
It is convenient to introduce the non-dimensional frequency $\Omega=k_{s,\infty}h$, where $k_{s,\infty}=\omega/c_{s,\infty}$. Here $c_{s,\infty}$ is the limit $\omega\to\infty$ for c_s which corresponds to the case of a non-viscous fluid. Note that the real part of $c_{s,\infty}$ is considered in the plastic foam case, as the complex shear modulus N results in a complex velocity.



(a)Quartz-fiber and water

(b)Plastic foam and air

FIG. 2. Curves for $kh=\pi$, where k is the real part of the wave numbers k_{p1} (solid), k_s (dashed), and k_{p2} (dotted). The horizontal solid lines indicate the chosen thicknesses.



(a) $\phi = 45°$, $h = 0.04$ m.

(b) $\phi = 45°$, $h = 0.4$ m.

(c) $\Omega = 0.8$, $h = 0.04$ m.

(d) $\Omega = 0.8$, $h = 0.04$ m.

FIG. 3. Closed pores for QF-20 and water. ——, exact; – – –, h^3; $\cdots$, h; and –·–, TB.

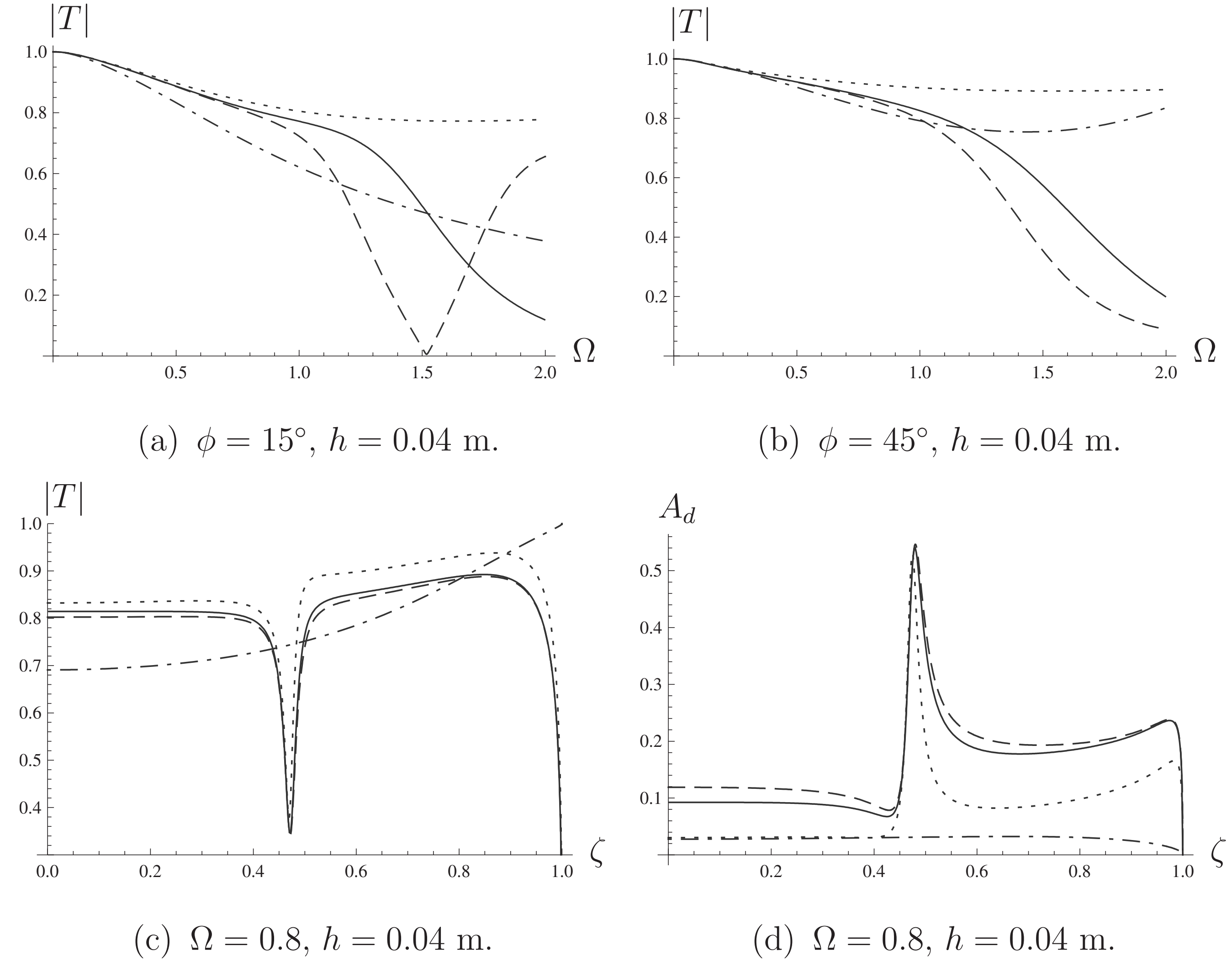
A. Quartz-fiber and water

In the case of QF-20 and water $c_{s,\infty} \approx 2036$ m s^{-1}. From Fig. 2(a) it is seen that for the slow compressional wave the limit $k_{p2}h = \pi$ occurs for $h = 0.4$ m when $\omega \approx 2250$ rad s^{-1} which corresponds to $\Omega \approx 0.44$, and for $h = 0.04$ m when $\omega \approx 70\,000$ rad s^{-1} corresponding to $\Omega \approx 1.38$. For the shear wave $k_s h = \pi$ occurs when $\Omega \approx 2.96$ for both thicknesses.

1. Closed pores

The results for closed pores are presented in Fig. 3. Figures 3(a) and 3(b), show the modulus of T when the angle of incidence is $\phi = 45°$. The results using the h^3 expansion are more accurate than the asymptotic h theory as expected. It is more surprising that the h theory is superior to the TB theory which involves terms of order h^3. This probably stems from the importance of the symmetric modes not modeled in the TB flexural plate theory. Considering the curves for the h^3 expansion, they deviate from the exact curves around $\Omega \approx 1$ for $h = 0.04$ m and around $\Omega \approx 1.5$ for $h = 0.4$ m, respectively. This latter case is far beyond the upper limit for the slow compressional wave discussed above. This accuracy may be connected to the fact that in this low-frequency interval the slow wave is much attenuated. It should also be stated that the slow compressional wave generally is not much excited when the pores are closed.[29] For other angles of incidence, similar results as depicted in Figs. 3(a) and 3(b) are obtained.

The dependence of $|T|$ on the angle of incidence is presented in Fig. 3(c) for the frequency $\Omega = 0.8$ when $h = 0.04$ m. The accuracy of the different approximations shows a similar behavior as in Figs. 3(a) and 3(b). Note the rapid transition of the transmission coefficient around $\zeta \approx 0.45$, that is, $\phi \approx 27°$. This is associated with the first compressional mode closely related to the zeroth symmetric Lamb mode; a similar behavior is reported by Johansson *et al.*[15] for a fluid-loaded elastic plate. The phenomenon is not captured by the TB theory since it does not take symmetric motion into account. Another minor incorrectness with the TB equation is that $|T| = 1$ for the grazing angle, $\zeta = 1$. The exact and asymptotic equations predict that the transmission modulus falls quickly to zero as ζ approaches unity. Similar plots are obtained for $h = 0.4$ m and for other frequencies, albeit the lower frequencies $\Omega < 0.5$ result in pronounced transmission for most angles as expected. Note that the angle of incidence corresponding to the rapid transition of the



(a) $\phi = 15°,\ h = 0.04$ m.

(b) $\phi = 45°,\ h = 0.04$ m.

(c) $\Omega = 0.8,\ h = 0.04$ m.

(d) $\Omega = 0.8,\ h = 0.04$ m.

FIG. 4. Open pores for QF-20 and water. ——, exact; – – –, h^3; ⋯, h; and –·–, TB.

transmission coefficient varies little with frequency as this symmetric wave guide mode exhibits small dispersion.

In Fig. 3(d) the absorption coefficient is plotted against the angle of incidence when $\Omega=0.8$ and $h=0.04$ m. Here the influence of the zeroth symmetrical mode is clearly visible, showing its significant influence on the wave absorption. As for the transmission coefficient the TB theory does not identify this behavior. Note here that the h-expansion seems to be more accurate than the asymptotic h^3 theory. This is merely by coincidence due to cancellation effects as the accuracies for both R and T are superior in the latter case. By studying the absorption coefficient as a function of frequency for different angles of incidence, the h^3 theory is seen to give more accurate results than the h theory in the lower-frequency interval. Similar plots are obtained for other frequencies as well as for $h=0.4$ m where in the latter case the absorption coefficient is slightly more pronounced.

2. Open pores

The results for open pores are presented in Fig. 4. Here Figs. 4(a) and 4(b), show the modulus of T when the angles of incidence are $\phi=15°$ and $\phi=45°$, respectively, for $h=0.04$ m. As for closed pores, the h^3 expansion is more accurate than the asymptotic h and the TB theories. All approximate theories deviate from the exact curves at lower frequencies when compared to the closed pore case. The main reason for this is probably due to that the slow compressional wave is more excited when the pores are open.[29] According to the h^3 expansion for $\phi=15°$ no transmission occurs at $\Omega\approx 1.5$, whereas the exact solution shows almost zero transmission at $\Omega\approx 2.2$. Such behavior is also present for $\phi=45°$. A similar situation appeared for heavy loading on an elastic plate.[15] However, this does not imply total reflection in the porous case as absorption is present. For thicker plates, less dramatic curves are obtained and the transmission coefficient is generally larger than for thin plates for a given non-dimensional frequency Ω. Here, the results due to the TB theory do not differ much for different thicknesses h for constant Ω, contrary to the exact theory and the h^3 expansion theory.

The dependence of $|T|$ on the angle of incidence is presented in Fig. 4(c) for the frequency $\Omega=0.8$ when $h=0.04$ m. The accuracy of the different approximations and the rapid transition show a similar behavior as for closed pores, clearly resembling the results for heavy loading on an elastic plate.[15] It is noted that the transition jump is more pronounced while there is less variation with the angle of incidence in the open pore case.

In Fig. 4(d) the absorption coefficient is plotted against the angle of incidence when $\Omega=0.8$ and $h=0.04$ m. Here the h^3 expansion theory is visibly superior to the h theory. When compared to the closed pore case, the absorption is more pronounced here, which probably is due to that the slow compressional wave is much more excited in the latter case. Similar plots are obtained for $h=0.4$ m as well as for other frequencies where in the latter case the absorption coefficient is larger for higher frequencies.

Note from the results in Figs. 3 and 4 that the TB solutions scarcely depend on the boundary conditions; virtually the same results are obtained for closed and open pores.

B. Plastic foam and air

Consider next the material configuration where the frame is made of a plastic foam of high flow resistivity saturated with air. This combination is studied by Allard *et al.*[27] In addition to viscous effects, viscoelastic and thermal effects are included as well. In this case, the material parameters in Eq. (2) may be simplified as $K_f/K_s \ll 1$ and $K_b/K_s \ll 1$. Hence, the parameters may be approximated by[2]

$$A = K_b + \frac{(1-\Phi)^2}{\Phi}K_f - \frac{2}{3}N, \quad Q = K_f(1-\Phi), \quad R = \Phi K_f. \tag{38}$$

The bulk modulus of the frame is now modeled in line with an isotropic elastic material

$$K_b = \frac{2N(1+\nu)}{3(1-2\nu)}, \tag{39}$$

where ν is Poisson's ratio. The thermal effects are incorporated in the frequency dependent bulk modulus of the fluid according to[2]

$$K_f = \frac{\gamma_h p_{\text{atm}}}{\gamma_h - (\gamma_h - 1)\left(1 + i\frac{\omega_c}{\omega}\left(1 - \frac{i\omega}{2\omega_c}\right)^{1/2}\right)^{-1}},$$
$$\omega_c = \frac{\eta}{\rho_f \kappa_0 \alpha_\infty c^2 \text{ Pr}}, \tag{40}$$

where Pr denotes the Prandtl number, p_{atm} is the atmospheric pressure, γ_h is the ratio of specific heats, and c is a form factor which depends on the shape of the pores. Table I involves those parameters needed for the numerical results.

For this material combination $c_{s,\infty} \approx 76$ m s^{-1}. From Fig. 2(b) it is seen that for the slow compressional wave the limit $k_{p2}h=\pi$ occurs for $h=0.4$ m when $\omega \approx 70$ rad s^{-1} which corresponds to $\Omega \approx 0.91$, and for $h=0.04$ m when $\omega \approx 6000$ rad s^{-1} corresponding to $\Omega \approx 3.13$. For the shear wave $k_s h=\pi$ occurs when $\Omega \approx 3.13$ for both thicknesses. By inspection, all three bulk wave speeds in the porous material are here lower than the speed of sound in the surrounding air. Moreover, the shear wave is now slower than the slow compressional wave for high frequencies.

1. Closed pores

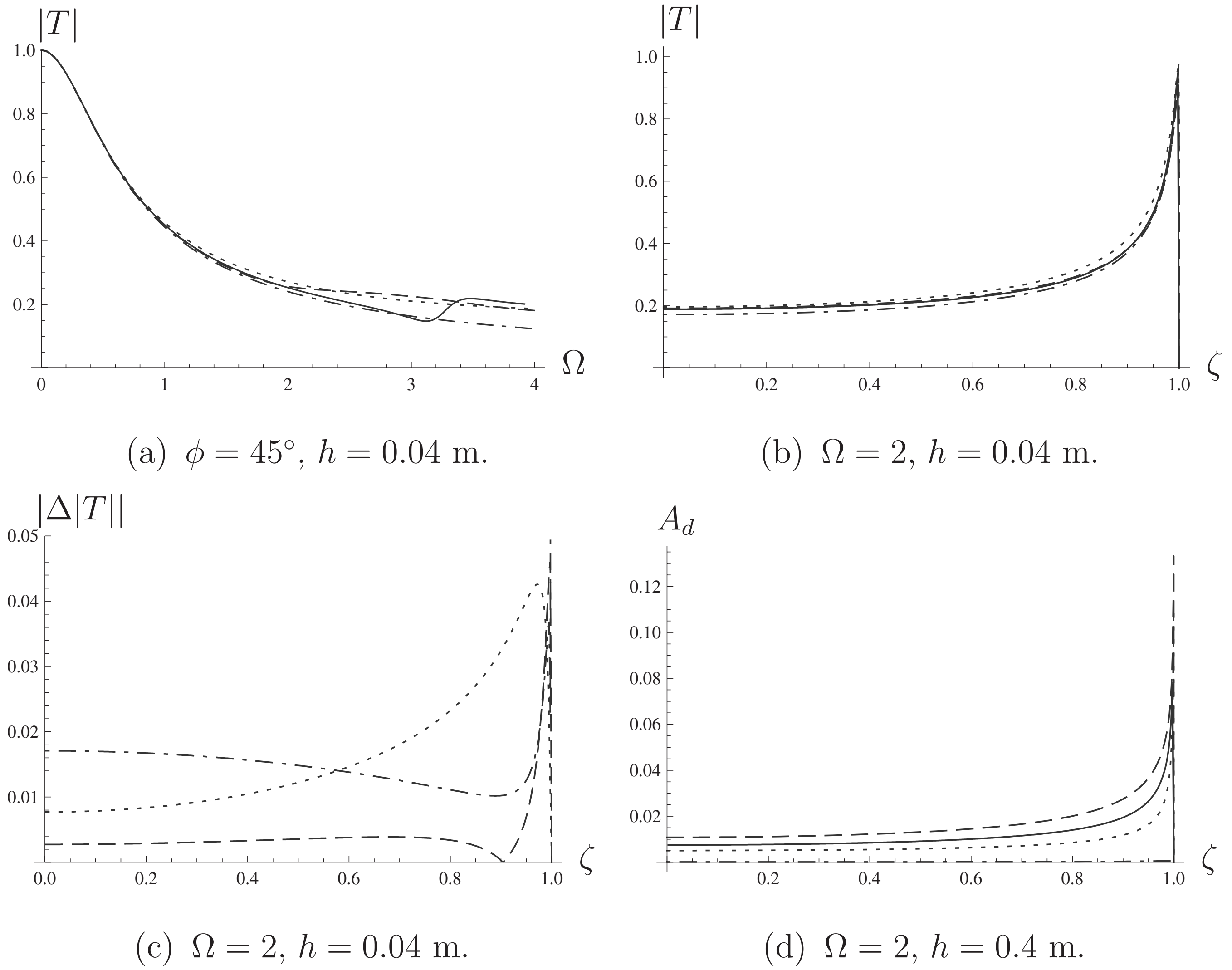
The results for closed pores are presented in Fig. 5. Figure 5(a) shows the modulus of T when the angle of incidence is $\phi=45°$ and $h=0.04$. All approximate theories render good results; the TB theory actually being superior in the interval $\Omega \approx 2$ to $\Omega \approx 3$. It is surprising that the TB theory renders such good results, especially considering the quite high frequencies for which these approximate theories are supposed to approach their limit for applicability. For other angles of incidence, similar results are obtained, even though the h^3 expansion theory is the most accurate for small angles. By inspection, the h^3 expansion theory is in all cases the most accurate in the lower-frequency intervals more suitable for plate theories. Almost identical curves are obtained for the thicker layer $h=0.4$, implying that the influence of the slow compressional wave is limited, see discussion in Sec. IV A 1.

The dependence of $|T|$ on the angle of incidence is presented in Fig. 5(b) for the frequency $\Omega=2$ when $h=0.04$ m. Contrary to the quartz-fiber case there is no rapid transition of the transmission coefficient. The reason for this is that such a coincidence phenomenon with the zeroth symmetrical mode does not correspond to a real-valued angle of incidence, $\zeta \approx 2.45$. Similar plots are obtained for $h=0.4$ m and for other frequencies, albeit the lower frequencies $\Omega < 0.2$ result in pronounced transmission for most angles as expected. When comparing the transmission coefficient presented in Fig. 5(b) to the quartz-fiber case [Fig. 3(c)], it is clear that considerably less is transmitted in the foam-air case. The accuracies of the different approximations are on a similar level, and the differences in the results are displayed in Fig. 5(c), where the absolute values of the error in the transmission coefficient modulus, $|\Delta|T||$, are plotted. Note that the presented plot has been truncated as $|\Delta|T|| \to 1$ for the TB theory as $\zeta \to 1$.

When studying the quite modest absorption coefficient in Fig. 5(d) when $h=0.4$ m for $\Omega=2$, it is clear that much of the wave is reflected considering the magnitude of $|T|$, see Fig. 5(b). Here, the TB curve is hard to distinguish being placed almost on the horizontal coordinate line. Similar plots are obtained for other frequencies, where the magnitude of the absorption coefficient increases with frequency. It is seen in this figure that the simpler h theory is slightly more accurate than the h^3 expansion theory, resulting from the accuracy of R. The reason for this puzzling behavior is that this quite high frequency seems to be on the upper scale when using plate theories for the reflection coefficient. For lower frequencies, the h^3 expansion theory gives the most satisfying results.

2. Open pores

The results for open pores are presented in Fig. 6. Here Fig. 6(a) shows the modulus of T when the angle of incidence is $\phi=45°$ for $h=0.04$ m. Contrary to the closed pore case, the asymptotic expansion theories are more accurate than the TB theory in the interval considered. As for the material combination quartz-fiber and water, the approximate curves deviate from the exact curve at lower frequencies when compared to the closed pore case due to the more pronounced excitation of the slow compressional wave. Therefore a more narrow frequency interval is studied in Fig. 6(a) compared to Fig. 5(a). Similar curves are obtained for other angles and frequencies.



(a) $\phi = 45°$, $h = 0.04$ m.

(b) $\Omega = 2$, $h = 0.04$ m.

(c) $\Omega = 2$, $h = 0.04$ m.

(d) $\Omega = 2$, $h = 0.4$ m.

FIG. 5. Closed pores for foam and air. ——, exact; – – –, h^3; ···, h; and –·–, TB.

The dependence of $|T|$ on the angle of incidence is presented in Fig. 6(b) for the frequency $\Omega=1.5$ when $h=0.04$ m. The accuracies of the different approximations show a similar behavior as in Fig. 6(a). It is noted that slightly more is transmitted when compared to the closed pore case [Fig. 5(b)].

Figures 6(c) and 6(d) present the absorption coefficient against the angle of incidence when $\Omega=1.5$ for $h=0.04$ m and $h=0.4$ m, respectively. When compared to the closed pore case, the absorption is more pronounced here mainly due to the influence of the slow compressional wave. As the absorption coefficient increases with frequency, the thin layer case depicted in Fig. 6(c) shows more absorption than the thick layer case in Fig. 6(d). Note that the TB curves are almost on the horizontal coordinate line.

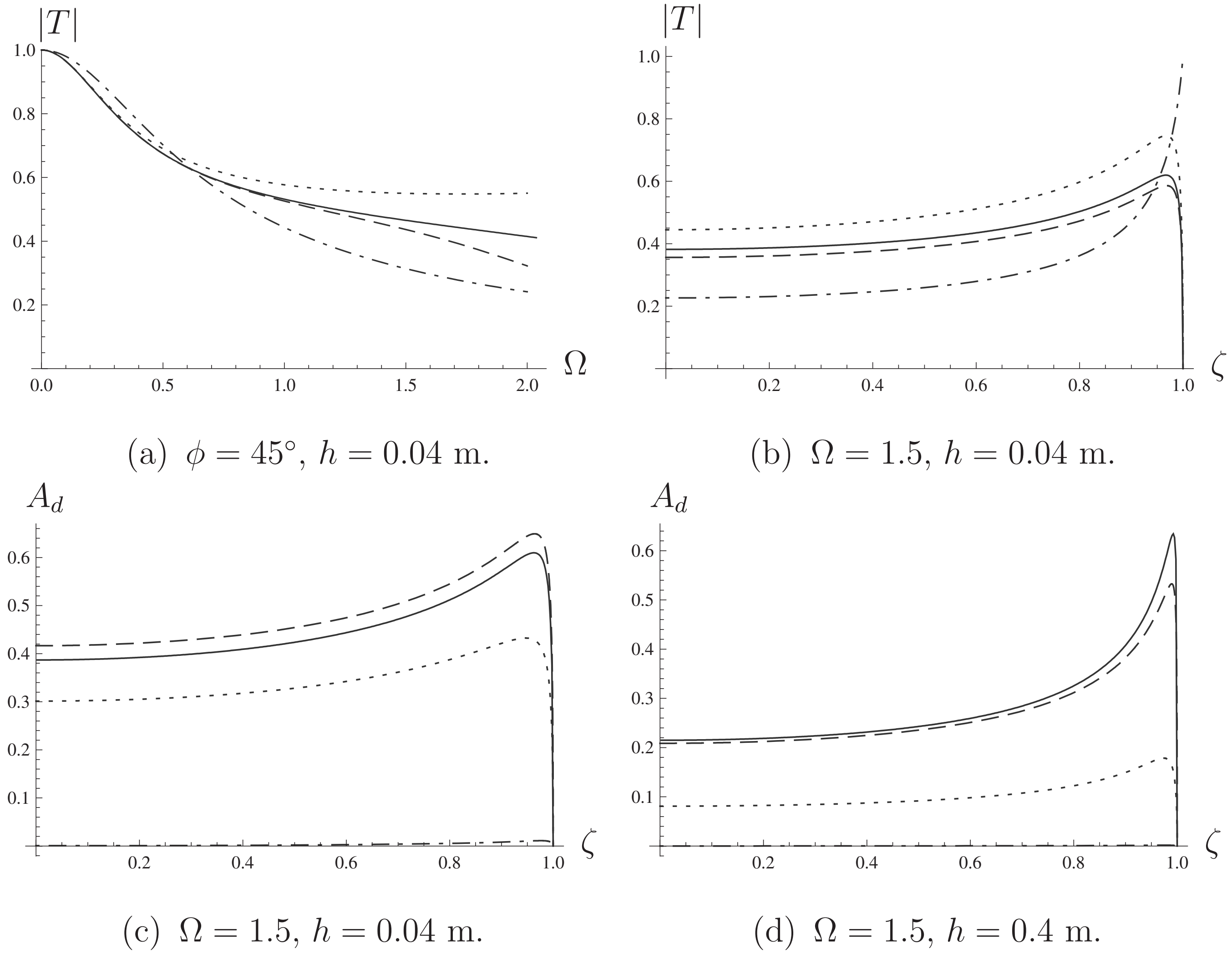
As for the quartz-fiber case, the TB solutions scarcely depend on the boundary conditions (closed or open pores) for the foam-air case.

V. CONCLUSIONS

This paper considers derivation of dynamical equations for fluid-loaded porous plates. Using a series expansion technique in the thickness coordinate, separate symmetric and antisymmetric plate equations are derived that are believed to be asymptotically correct. These equations are expressed in terms of approximate boundary conditions under time-harmonic conditions. Analytical and numerical results are presented for open and closed pores, using asymptotic expansions up to order 3 in thickness. These results are compared to the exact 3D theory and the TB theory based on pure bending assumptions.

The analytical equations generally show a pronounced discrepancy between the presented asymptotic theory and the TB theory. The asymptotic equations have the same structure as the corresponding approximate boundary conditions for an elastic layer. However, sixth order tangential derivatives appear in the symmetric equations. A further similarity with the elastic case is that the differential operator for the symmetric pressure term for closed pores may be factorized.

Numerical results are presented for two material configurations: quartz-fiber saturated with water and plastic foam saturated with air. The performances of the asymptotic approximate boundary conditions are accurate over a quite wide frequency interval. In many cases acceptable approximations are provided even by the much simpler h^1-theory.



(a) $\phi = 45°$, $h = 0.04$ m.

(b) $\Omega = 1.5$, $h = 0.04$ m.

(c) $\Omega = 1.5$, $h = 0.04$ m.

(d) $\Omega = 1.5$, $h = 0.4$ m.

FIG. 6. Open pores for foam and air. ——, exact; – – –, h^3; ···, h; and –·–, TB.

This is probably due to the limited influence from the slow compressional wave, which would otherwise decrease the range of applicability as its wave speed is comparably low. Another explanation is that the influence from flexural plate motions here is less pronounced than for the case of an elastic layer, which in the latter case drastically limited the validity of the h^1-theory.[15]

For the future there are good reasons to believe that series expansion techniques can be successfully applied to similar more complicated problems, e.g., anisotropic porous media or porous structures with curved surfaces. As the present paper shows, there are needs to develop new equations in a rigorous way in order to model advanced porous structures properly. What has not been presented in the current paper is how to implement the derived equations in a bounded case, including edge boundary conditions. Such higher order end boundary conditions can be derived in an equally systematic manner using variational methods, recently adopted for homogeneous rods. Eventually, the new plate equations may be implemented in a finite element environment, where plate elements still are preferable to 3D elements due to reasonable length-to-thickness ratio of the elements.

ACKNOWLEDGMENTS

This work has been supported in part by the Swedish Research Council (VR) and in part by the Volvo-Chalmers Vehicle Research Program.

APPENDIX: DISPLACEMENT BASED PLATE EQUATIONS

In Sec. III the asymptotic plate equations are written as approximate boundary conditions in terms of the pressure in the surrounding fluid: p_0 and $\partial_z p_0$. However, plate equations are traditionally expressed in terms of the plate fields and the external pressure field p_0. In the derivation process, this is accomplished by eliminating the field $\partial_z p_0$ and retaining the plate fields. Here the results are presented for the antisymmetric case only since this is generally the most pronounced vibration mode in engineering situations. Moreover, this situation corresponds to modest modifications of the approximate boundary conditions presented above, while the symmetric case results in virtually new sets of equations.

So, in the antisymmetric case the plate equations are now expressed in terms of Δp_0 and $u_{z,c}$, where $u_{z,c} = u_z(x,y,0)$. Hence, the vertical displacement of the center plane $u_{z,c}$ is used in favor of $\Sigma\partial_z p_0$ at the boundaries. This could be accomplished directly from Eqs. (26) and (35) by using Eqs. (15) and (17) for closed and open pores, respectively. The final results are obtained from adopting series expansion of the boundary plate fields in terms of the center plane fields, together with the boundary conditions and the equations of motion (1). Hereby the right-hand sides of Eqs. (26) and (35) still hold by changing $\Sigma\partial_z p_0$ to $u_{z,c}$, while the left-hand sides now become

$$\frac{1}{2N}\left\{1-\frac{1}{2}\left(\frac{h}{2}\right)^2[2(2-\gamma_1)\nabla_s^2+(k_s^2-k_1^2+k_{p1}^2k_{p2}^2/k_s^2)]\right\}\Delta p_0$$

in the closed pore case and

$$\frac{1-\Phi\beta_3}{2N}\left\{1-\frac{1}{6}\left(\frac{h}{2}\right)^2[(10(1-\gamma_1)+\gamma_4)\nabla_s^2\right.$$
$$\left.+3k_{\mathrm{av}}^2+k_s^2(2+\gamma_4)]\right\}\Delta p_0$$

in the open pore case, where

$$\gamma_4=\frac{N(\Phi Q-(1-\Phi)R)}{a_3(1-\Phi\beta_3)}.$$

[1] 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).

[2] J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier Science, London, 1993).

[3] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[4] A. L. Taber, "A theory for transverse deflection of poroelastic plates," Trans. Am. Soc. Mech. Eng. **59**, 628–634 (1992).

[5] D. D. Theodorakopoulos and D. E. Beskos, "Flexural vibrations of poroelastic plates," Acta Mech. **103**, 191–203 (1994).

[6] P. Leclaire, K. V. Horoshenkov, and A. Cummings, "Transverse vibrations of a thin rectangular porous plate saturated by a fluid," J. Sound Vib. **247**, 1–18 (2001).

[7] P. Leclaire, K. V. Horoshenkov, M. J. Swift, and D. C. Hothersall, "The vibrational response of a clamped rectangular porous plate," J. Sound Vib. **247**, 19–31 (2001).

[8] M. Etchassahar, S. Sahraout, and B. Brouard, "Bending vibrations of a rectangular poroelastic plate," C. R. Acad. Sci., Ser. IIb Mec. **329**, 615–620 (2001).

[9] H. Aygun, A. Attenborough, and A. Cummings, "Predicted effects of fluid loading on the vibration of elastic porous plates," Acta. Acust. Acust. **93**, 284–289 (2007).

[10] L. P. Li, G. Cederbaum, and K. Schulgasser, "Theory of poroelastic plates with in-plane diffusion," Int. J. Solids Struct. **34**, 2515–4530 (1997).

[11] A. Boström, "On wave equations for elastic rods," Z. Angew. Math. Mech. **80**, 245–251 (2000).

[12] N. Losin, "Asymptotics of flexural waves in isotropic elastic plates," ASME J. Appl. Mech. **64**, 336–342 (1997).

[13] N. Losin, "Asymptotics of extensional waves in isotropic elastic plates," ASME J. Appl. Mech. **65**, 1042–1047 (1998).

[14] A. Boström, G. Johansson, and P. Olsson, "On the rational derivation of a hierarchy of dynamic equations for a homogeneous, isotropic, elastic plate," Int. J. Solids Struct. **38**, 2487–2501 (2001).

[15] M. Johansson, P. Folkow, A. Hägglund, and P. Olsson, "Approximate boundary conditions for a fluid-loaded elastic plate," J. Acoust. Soc. Am. **118**, 3436–3446 (2005).

[16] G. Johansson and A. Niklasson, "Approximate dynamic boundary conditions for a thin piezoelectric layer," Int. J. Solids Struct. **40**, 3477–3492 (2003).

[17] K. Mauritsson, A. Boström, and P. Folkow, "Modelling of thin piezoelectric layers on plates," Wave Motion **45**, 616–628 (2008).

[18] N. Losin, "On the equivalence of dispersion relations resulting from Rayleigh-Lamb frequency equation and the operator plate model," J. Vibr. Acoust. **123**, 417–420 (2001).

[19] P. Bövik, "On the modelling of thin interface layers in elastic and acoustic scattering problems," Q. J. Mech. Appl. Math. **47**, 17–42 (1994).

[20] H. Ammari and S. He, "Generalized effective impedance boundary conditions for an inhomogeneous thin layer in electromagnetic scattering," J. Electromagn. Waves Appl. **11**, 1197–1212 (1997).

[21] P. Olsson, S. Datta, and A. Boström, "Elastodynamic scattering from inclusions surrounded by thin interface layers," ASME J. Appl. Mech. **57**, 672–676 (1990).

[22] S. I. Rokhlin and W. Huang, "Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids. II. Second-order asymptotic boundary conditions," J. Acoust. Soc. Am. **94**, 3405–3420 (1993).

[23] 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).

[24] T. Bourbié, O. Coussy, and B. Zinsner, *Acoustics of Porous Media* (Technip, Paris, 1987).

[25] Registered trademark of Wolfram Research Inc.

[26] D. L. Johnson, T. J. Plona, and H. Kojima, "Probing porous media with first and second sound. II. Acoustic properties of water-saturated porous media," J. Appl. Phys. **76**, 115–125 (1994).

[27] J.-F. Allard, C. Depolliér, and W. Lauriks, "Measurement and prediction of surface impedance at oblique incidence of a plastic foam of high flow resistivity," J. Sound Vib. **132**, 51–60 (1989).

[28] H. Belloncle, H. Franklin, F. Luppé, and J. Conoir, "Normal modes of a poroelastic plate and their relation to the reflection and transmission coefficients," Ultrasonics **41**, 207–216 (2003).

[29] P. N. J. Rasolofosaon, "Importance of interface hydraulic condition on the generation of second bulk compressional wave in porous media," Appl. Phys. Lett. **52**, 780–782 (1988).

Optimal simulations of ultrasonic fields produced by large thermal therapy arrays using the angular spectrum approach

Xiaozheng Zeng and Robert J. McGough[a)]

Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824

(Received 20 September 2008; revised 14 February 2009; accepted 17 February 2009)

The angular spectrum approach is evaluated for the simulation of focused ultrasound fields produced by large thermal therapy arrays. For an input pressure or normal particle velocity distribution in a plane, the angular spectrum approach rapidly computes the output pressure field in a three dimensional volume. To determine the optimal combination of simulation parameters for angular spectrum calculations, the effect of the size, location, and the numerical accuracy of the input plane on the computed output pressure is evaluated. Simulation results demonstrate that angular spectrum calculations performed with an input pressure plane are more accurate than calculations with an input velocity plane. Results also indicate that when the input pressure plane is slightly larger than the array aperture and is located approximately one wavelength from the array, angular spectrum simulations have very small numerical errors for two dimensional planar arrays. Furthermore, the root mean squared error from angular spectrum simulations asymptotically approaches a nonzero lower limit as the error in the input plane decreases. Overall, the angular spectrum approach is an accurate and robust method for thermal therapy simulations of large ultrasound phased arrays when the input pressure plane is computed with the fast nearfield method and an optimal combination of input parameters. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097499]

PACS number(s): 43.40.Rj, 43.20.El, 43.20.Rz [TDM] Pages: 2967–2977

I. INTRODUCTION

Pressure fields generated by ultrasound therapy arrays are typically calculated by superposing the fields produced by individual transducer sources. Traditionally, these sources are modeled with point source superposition applied to the Rayleigh–Sommerfeld integral,[1–3] the rectangular radiator method,[4–6] the spatial impulse response method,[7–9] and other analytically equivalent integral approaches. All of these methods calculate the pressure at each grid point; therefore, the simulation time is proportional to the number of array elements multiplied by the size of the computational grid. These simulations are relatively slow due to the large number of calculations involved. In contrast, the angular spectrum approach[10] rapidly computes pressures in parallel planes. This approach decomposes the diffracted wave into plane waves via the two dimensional (2D) Fourier transform, propagates these components in the spatial frequency domain, and recovers the pressure field in planes parallel to the input plane through the 2D inverse Fourier transform.

The numerical accuracy of the angular spectrum approach has been extensively discussed for single planar radiators. For example, Williams and Maynard[11] analyzed the difference between the analytical Fourier transform and the discrete Fourier transform (DFT) in angular spectrum simulations. Williams and Maynard proposed an averaging approach to reduce the aliasing error and the error induced by certain singularities in the spectral propagator. Orofino and Pedersen[12,13] derived a relationship between the angular range for the decomposed plane wave components and the sampling rate of the spectra in the spatial frequency domain. These parameters are correlated to the spatial sampling rate, which in part determines the accuracy of the angular spectrum simulation. Wu *et al.*[14–16] derived the maximum angular range that satisfies the Nyquist sampling criteria for the spectral propagator. Wu *et al.* also used the analytical Fourier transform of a rectangular radiator to eliminate the numerical errors introduced by the DFT of the input normal particle velocity distribution. Zeng and McGough[17] compared the performance of the spatial propagator and the spectral propagator in terms of numerical accuracy and time. After identifying an artifact caused by the truncation of the spatial propagator, Zeng and McGough showed that the spatial propagator yields more accurate simulation results once the region containing the artifact is removed, especially for simulations in non-attenuating media. In attenuating media, the spectral propagator achieves similar accuracy in less time. Zeng and McGough also showed that including attenuation in angular spectrum simulations effectively reduces aliasing errors through a spatial frequency filtering effect and that apodizing the input particle velocity distribution achieves the same result. Overall, the spectral propagator is preferred over the spatial propagator for calculations in attenuating media or with an apodized particle velocity source, whereas simulations in non-attenuating media favor the spatial propagator.[17]

[a)] Author to whom correspondence should be addressed. Electronic mail: mcgough@egr.msu.edu

The angular spectrum approach is also widely used for phased array simulations. For example, the pressure fields generated by concentric ring arrays and sector vortex arrays have been compared with experimental results,[18] and the angular spectrum approach has also been applied to high intensity focused ultrasound simulations.[19] Zemp and Tavakkoli[20] compared the sampling of the spatial and spectral propagators for phased array simulations and then derived the maximum unaliased spectral sampling rate for the spectral propagator. Other important issues that impact the accuracy of thermal therapy simulations remain unsolved; therefore, more thorough evaluations of the angular spectrum approach are needed.

This paper evaluates the angular spectrum approach for calculations of time-harmonic pressure fields generated by large ultrasound phased arrays in an attenuating medium. First, the performance of the angular spectrum approach is compared for input planes that consist of particle velocity distributions and pressure distributions. Second, the effect of the location of the input pressure plane and the size of the window that truncates the input pressure plane are determined. Third, the impact of the numerical accuracy of the 2D input pressure on the three dimensional (3D) output pressure field computed with the angular spectrum approach is evaluated. Finally, temperature fields are computed from the 3D pressure fields obtained with the angular spectrum approach, and the errors are compared. The results show that when the input pressure planes are computed with optimal parameters applied to the fast nearfield method, the angular spectrum approach rapidly and accurately calculates pressures for thermal therapy simulations with large ultrasound phased arrays.

II. THEORY

A. Integral approaches

In simulations of individual transducers and large ultrasound phased arrays, the pressure is often calculated with the Rayleigh–Sommerfeld integral.[10,21] This 2D integral is ordinarily evaluated with the midpoint rule,[22] which is equivalent to subdividing the transducer surface into point sources and superposing all of the contributions.[23] The Rayleigh–Sommerfeld diffraction integral[10] is

$$p(\mathbf{r},t) = j\rho c k e^{j\omega t}\int_{\mathbf{S}'} u(\mathbf{r}')\frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{2\pi|\mathbf{r}-\mathbf{r}'|}d\mathbf{S}', \qquad (1)$$

where ρ and c represent the density and the speed of sound, respectively, ω is the driving frequency, $k=\omega/c=2\pi/\lambda$ is the acoustic wavenumber, u is the distribution of the normal velocity on the radiator with surface area $\mathbf{S}'$, j is $\sqrt{-1}$, and $|\mathbf{r}-\mathbf{r}'|$ is the distance between the source coordinates $\mathbf{r}'=(x',y',z')$ and the observation coordinates $\mathbf{r}=(x,y,z)$.

The spatial impulse response approach is an analytically equivalent method that computes the pressure field with a one dimensional (1D) integral. This integral evaluates the convolution of the impulse response function of a transducer with the time derivative of the excitation function.[7] For a rectangular piston excited with a continuous wave input, the pressure is proportional to the Fourier transform of the spatial impulse response, which is described by[24,25]

$$p(x,y,z,t) = j\rho c u_0 e^{j\omega t}\sum_{i=1}^{2}\sum_{j=1}^{2} \pm I_{s_i,l_j}(x,y,z) \qquad (2)$$

and

$$I_{s,l} = \frac{k}{2\pi}\left[\frac{j\pi}{2k}\left(e^{-jk\sqrt{z^2+s^2+l^2}} - e^{-jkz}\right) - \int_{\sqrt{z^2+s^2}}^{\sqrt{z^2+s^2+l^2}} \cos^{-1}\left(\frac{s}{\sqrt{\sigma^2-z^2}}\right)e^{-jk\sigma}d\sigma - \int_{\sqrt{z^2+l^2}}^{\sqrt{z^2+s^2+l^2}} \cos^{-1}\left(\frac{l}{\sqrt{\sigma^2-z^2}}\right)e^{-jk\sigma}d\sigma\right], \qquad (3)$$

where σ is a distance variable, $s_1=|x-a|$, $s_2=|x+a|$, $l_1=|y-b|$, and $l_2=|y+b|$, and the + or − sign in Eq. (2) is determined by the spatial location of the observation point with respect to the transducer aperture.

The fast nearfield method[24,25] is a 1D integral approach for nearfield pressure calculations that converges much more rapidly than the spatial impulse response.[7] The fast nearfield method for a rectangular piston that is uniformly excited is derived in Ref. 24, and the fast nearfield method for a rectangular piston with an apodized surface velocity distribution is derived in Ref. 25. This method achieves rapid convergence by subtracting singularities, and the computation time is reduced by exploiting repeated calculations. The fast nearfield expression for a uniformly excited rectangular piston[24] is

$$p(x,y,z,t) = -1\rho c u_0 e^{j\omega t}\frac{1}{2\pi}\left(s_1\int_{-l_1}^{l_2}\frac{e^{-jk\sqrt{z^2+\sigma^2+s_1^2}} - e^{-jkz}}{\sigma^2+s_1^2}d\sigma + l_1\int_{-s_1}^{s_2}\frac{e^{-jk\sqrt{z^2+\sigma^2+l_1^2}} - e^{-jkz}}{\sigma^2+l_1^2}d\sigma + s_2\int_{-l_1}^{l_2}\frac{e^{-jk\sqrt{z^2+\sigma^2+s_2^2}} - e^{-jkz}}{\sigma^2+s_2^2}d\sigma + l_2\int_{-s_1}^{s_2}\frac{e^{-jk\sqrt{z^2+\sigma^2+l_2^2}} - e^{-jkz}}{\sigma^2+l_2^2}d\sigma\right), \qquad (4)$$

where the limits of integration are $s_1=a-x$, $l_1=b-y$, $s_2=a+x$, and $l_2=b+y$, and a and b represent the half width and the half height of the rectangular source, respectively. The fast nearfield method achieves high numerical accuracy in a very short time.[24,25]

B. Phased array beamforming

For linear simulations of therapeutic ultrasound, the pressure field generated by a phased array is computed via the superposition of complex pressures produced by the array transducers according to

$$p=\sum_{n=1}^{N} p_n A_n e^{j\Phi_n}. \tag{5}$$

In Eq. (5), N is the number of transducers in the array, and p_n is the pressure generated by each transducer. A_n represents the apodization weight, and Φ_n is the phase shift applied to each transducer element. An ultrasound beam with a single focus is obtained with the phase conjugation method.[26] To focus the array, a complex exponential term that equals the conjugate of the pressure produced by each array element at the focus is applied to that element. Therefore, the pressure waves generated by all of the array elements achieve constructive interference at the focal spot. Through beamforming, the pressure fields generated by ultrasound phased arrays are maximized at selected locations.

C. Angular spectrum approach

The angular spectrum approach calculates the pressure in a sequence of parallel planes by propagating each spatial frequency component of the diffracted wave in the spatial frequency domain.[10] The pressure or normal particle velocity field in an initial plane is defined as the input, and the output from angular spectrum calculations is the pressure evaluated in a series of parallel planes. The pressure field and the angular spectrum in each plane are related through the 2D Fourier transform. In a linear homogeneous medium, the propagation of acoustic waves in the spatial frequency domain is described by[27]

$$P(k_x,k_y,z)=P_0(k_x,k_y,z_0)H_p(k_x,k_y,\Delta z) \tag{6}$$

or

$$P(k_x,k_y,z)=j\rho c U_0(k_x,k_y,z_0)H_u(k_x,k_y,\Delta z), \tag{7}$$

where $\Delta z=z-z_0$, k_x and k_y are the transverse wavenumbers, and $k_x^2+k_y^2+k_z^2=k^2$. $P_0(k_x,k_y,z_0)$ is the angular spectrum of the input pressure field $p_0(x,y,z_0)$; i.e., $P_0(k_x,k_y,z_0)$ is the 2D Fourier transform of $p_0(x,y,z_0)$ with respect to x and y, and $U_0(k_x,k_y,z_0)$ is the 2D Fourier transform of the normal particle velocity on the radiator surface. $P(k_x,k_y,z)$ is the angular spectrum of the pressure in a plane parallel to the source plane. The pressure field in each subsequent plane is then obtained by applying a 2D inverse Fourier transform to $P(k_x,k_y,z)$ with respect to k_x and k_y. The spectral propagator $H_p(k_x,k_y,\Delta z)$ for an input pressure plane is described by[28]

$$H_p(k_x,k_y,\Delta z)=\begin{cases} e^{-j\Delta z\sqrt{k^2-k_x^2-k_y^2}} & \text{for } k_x^2+k_y^2\leqslant k^2 \\ e^{-\Delta z\sqrt{k_x^2+k_y^2-k^2}} & \text{for } k_x^2+k_y^2>k^2, \end{cases} \tag{8}$$

and the spectral propagator $H_u(k_x,k_y,\Delta z)$ for an input particle velocity distribution is represented by

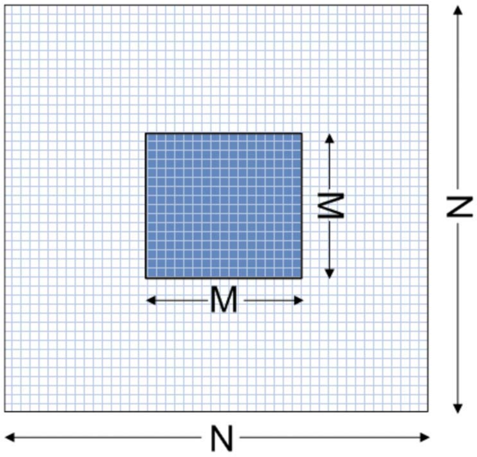


FIG. 1. (Color online) The discretized input plane, where the input pressure or normal particle velocity plane is initially computed in an $M\times M$ grid and then zero-padded to an $N\times N$ grid for angular spectrum calculations. The spectral propagator, which is not zero-padded, is then evaluated in an $N\times N$ grid for angular spectrum calculations.

$$H_u(k_x,k_y,\Delta z) =\begin{cases} \dfrac{k}{j\sqrt{k^2-k_x^2-k_y^2}}e^{-j\Delta z\sqrt{k^2-k_x^2-k_y^2}} & \text{for } k_x^2+k_y^2\leqslant k^2 \\ \dfrac{k}{\sqrt{k_x^2+k_y^2-k^2}}e^{-\Delta z\sqrt{k_x^2+k_y^2-k^2}} & \text{for } k_x^2+k_y^2>k^2, \end{cases} \tag{9}$$

Both $H_p(k_x,k_y,\Delta z)$ and $H_u(k_x,k_y,\Delta z)$ describe propagating waves in the region where $k_x^2+k_y^2\leqslant k^2$ and evanescent waves that decay exponentially where $k_x^2+k_y^2>k^2$. The propagator functions in Eqs. (8) and (9) are multiplied by an exponential term[17] for angular spectrum calculations in attenuating media,

$$S(k_x,k_y,\Delta z)=e^{-\alpha k\Delta z/\sqrt{k^2-k_x^2-k_y^2}}, \tag{10}$$

where α is the attenuation coefficient for a given ultrasound frequency. Multiplying the spatial frequency components by $S(k_x,k_y,\Delta z)$ achieves equivalent attenuation of pressure waveforms in the spatial domain.

To implement the angular spectrum approach, the input pressure or normal particle velocity field is first discretized, where the geometry of the input plane is illustrated in Fig. 1. An $L\times L$ square plane is discretized into a grid containing $M\times M$ points with a spatial sampling interval of δ. This grid is zero-padded to a larger $N\times N$ grid, and the angular spectrum of the input plane is computed with a 2D fast Fourier transform (FFT). The spectral propagator is then evaluated on the larger $N\times N$ grid in the spatial frequency domain. By extending the size of the grid to $N\times N$ ($N>M$), the resolution in the spatial frequency domain is increased and spectral aliasing errors are diminished. The spectral sampling rate is inversely proportional to N via $\Delta k=2\pi/(N\delta)$, and the discretized transverse wavenumbers are

$$k_x=m\Delta k,\quad m=-N/2+1+\phi,\ldots,N/2+\phi,$$

$$k_y=n\Delta k,\quad n=-N/2+1+\phi,\ldots,N/2+\phi, \tag{11}$$

where ϕ is defined by

$$\phi = \begin{cases} -\frac{1}{2} & \text{when } N \text{ is odd} \\ 0 & \text{when } N \text{ is even}. \end{cases} \quad (12)$$

The parameter ϕ compensates for the offset induced by the odd number of grid points so that m and n are integers.

D. Error evaluations

The numerical error in the simulated pressure field is evaluated with a root mean squared error (RMSE) defined by

$$\text{RMSE} = \sqrt{\frac{1}{n_x n_y n_z} \sum_{i,j,k} |p^{i,j,k} - p_{\text{ref}}^{i,j,k}|^2}, \quad (13)$$

where the superscripts (i,j,k) represent discrete field points in the computational grid, n_x, n_y, and n_z describe the number of points in the x, y, and z directions, respectively, p_{ref} is the complex reference pressure field computed with the spatial impulse response method, and p is the complex pressure field computed with the angular spectrum approach, the Rayleigh–Sommerfeld integral, or the fast nearfield method. The RMSE is evaluated either in a 3D volume or in a single transverse plane perpendicular to the array normal.

E. Temperature simulations

As acoustic waves propagate through a lossy medium, mechanical energy dissipates and is converted into heat. The power deposition is approximated by

$$Q_p(x,y,z) = \frac{\alpha}{\rho c} p(\mathbf{r}) p^*(\mathbf{r}), \quad (14)$$

and the localized heat transfer in biological media is modeled by the bio-heat transfer equation (BHTE),[29]

$$K\nabla^2 T - W_b C_b (T - T_a) + Q_p = 0, \quad (15)$$

where $T = T(x,y,z,t)$ is the tissue temperature, T_a is temperature of the arterial blood, K is the thermal conductivity of tissue, and W_b and C_b are the perfusion rate and the specific heat of blood, respectively. Equation (15) is the steady state BHTE, which models the temperature distribution under equilibrium conditions. For numerical calculations, Eq. (15) is evaluated with an iterative finite difference routine.[30]

III. SIMULATION RESULTS

A. Reference pressure field generated by a 32×32 element phased array

To evaluate the numerical performance of the angular spectrum approach, a 32×32 element 2D planar array is simulated and compared to a reference field. The array is comprised of 1.8 mm×1.8 mm square transducers with a 0.5 mm kerf between adjacent elements. The structure of this array is illustrated in Fig. 2. The array is located in the xy plane at $z=0$ cm and centered at the origin of the coordinate system. The z axis is coincident with the normal evaluated at the center of the array aperture. The excitation frequency is 1 MHz, the speed of sound is 1500 m/s, and the attenuation coefficient is $\alpha=1$ dB/cm/MHz. The total extent of the array aperture is 7.31 cm×7.31 cm, which is equal to 48.7λ ×48.7λ for a 1 MHz excitation. The reference field is calculated with the spatial impulse response method.[7] Using 1000 Gauss abscissas to compute the pressure generated by each square element, the spatial impulse response method calculates the total field generated by this array to an accuracy of 11 digits, as determined by a comparison with the fast nearfield method evaluated with the same number of abscissas. All simulations are performed on a 2.4 GHz Pentium 4 PC (1 Gbyte random access memory) running the Windows XP operating system. All routines are written in the C language, compiled by Microsoft VISUAL C/C++ Version 7.0, and called by MATLAB 7.1 as MEX files.

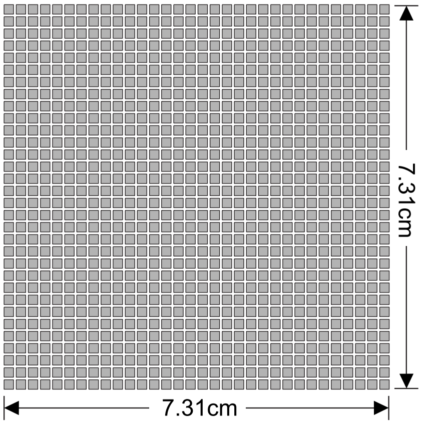


FIG. 2. A planar ultrasound phased array comprised of 32×32 square elements. The size of the array is 7.31 cm×7.31 cm (48.7λ×48.7λ for a 1 MHz driving frequency). The array consists of 1.8 mm×1.8 mm (1.2λ ×1.2λ) square transducers with a 0.5 mm kerf between adjacent elements.

The initial evaluations of the pressure field generated by this 32×32 element phased array are performed in a 20.4 cm×20.4 cm×12 cm (136λ×136λ×80λ) volume with an equal transverse extent in both the x and the y directions. With a sampling interval of $\delta=0.075$ cm ($\delta=\lambda/2$), the computational volume is discretized to a 273×273×161 point grid. Figure 3 shows the reference pressure field generated by this 32×32 element phased array in the $y=0$ plane. The array elements are phased such that a single focus

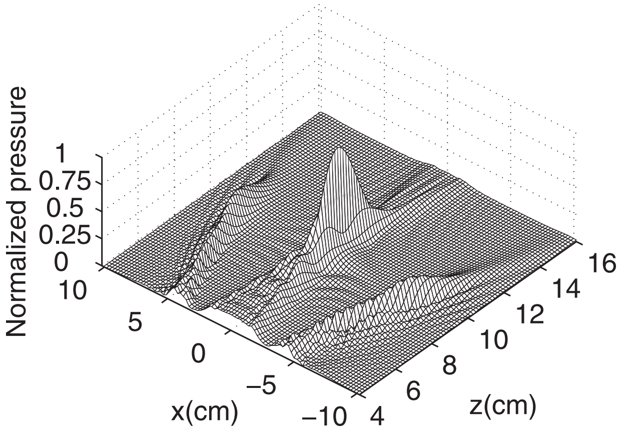


FIG. 3. Reference pressure field generated by the 32×32 element planar array in Fig. 2, where the array is focused at (0,0,10) cm. The pressure, which is normalized by the maximum amplitude, is shown in the xz plane at $y=0$. The excitation frequency for the array is 1 MHz, and the attenuation coefficient is $\alpha=1$ dB/cm/MHz.

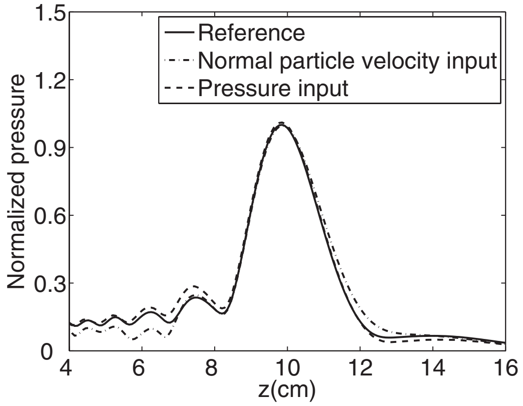


FIG. 4. Simulated axial pressures generated by the 32 × 32 element planar array in Fig. 2. The array is located at $z=0$ cm and electronically focused at $(0,0,10)$ cm. The reference pressure calculated by the spatial impulse response method is indicated by the solid line, the pressure computed with the angular spectrum approach using an input normal particle velocity plane is represented by the dash-dot line, and the pressure computed with the angular spectrum approach using an input pressure plane is represented by the dashed line.

is produced at $(0,0,10)$ cm. The pressure distribution depicted in Fig. 3 is normalized by the overall maximum pressure amplitude in the 3D volume.

B. Evaluation of pressure and normal particle velocity inputs

Although angular spectrum calculations with input pressure planes [Eq. (6)] and input normal particle velocity planes [Eq. (7)] are analytically equivalent, the numerical errors differ. To demonstrate the difference between these two approaches, the reference pressure field generated by the 32 × 32 element phased array in Fig. 2 is simulated with the spatial impulse response method and then compared to the results obtained with the angular spectrum approach using input pressure and normal particle velocity planes. In each simulation, the normal particle velocity is uniform across each element on the array aperture. The input pressure plane ideally extends to infinity in both lateral directions; however, for computer simulations, the pressure field is truncated by a 20.4 cm × 20.4 cm (136λ × 136λ) square window. In these calculations, the input pressure is calculated with the fast nearfield method using 20 abscissas for each integral. The spatial sampling interval is $\delta=0.075$ cm $(\delta=\lambda/2)$, and the value $N=512$ specifies the number of grid points in the x and y directions. Figure 4 shows the reference axial pressure (solid line) and the axial pressures computed with the angular spectrum approach using an input normal particle velocity plane (dash-dot line) and an input pressure plane (dashed line). For these angular spectrum simulations, the input normal particle velocity and the input pressure are both calculated in the plane at $z=0$. The resulting 3D fields are normalized by the maximum amplitude of the reference pressure.

Figure 4 shows that the output pressure fields obtained from the input pressure and the input normal particle velocity match the reference closely near the focus. The amplitude of the output pressure field computed with the input pressure is slightly larger than the reference field before the focus and slightly smaller than the reference field in the region beyond the focus. The amplitude of the output pressure field computed with the input normal particle velocity plane is smaller than the reference before the focus and slightly larger than the reference beyond the focus. Overall, the output axial pressure computed with the input normal particle velocity plane has a much larger error than that obtained with the input pressure plane.

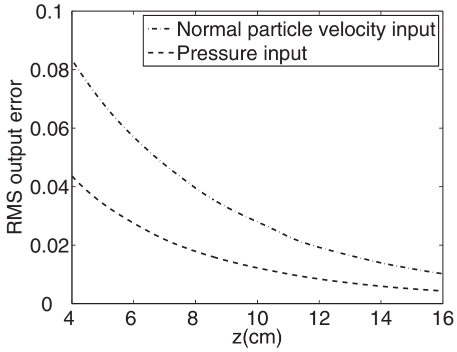


FIG. 5. RMS output errors for the 32 × 32 element array focused at $(0,0,10)$ cm evaluated in transverse planes for z ranging from 4 cm (26.67λ) to 16 cm (106.67λ). The pressure is calculated with the angular spectrum approach using an input normal particle velocity plane (dash-dot line) and an input pressure plane (dashed line). For this result, the input particle velocity and pressure planes are both located at $z_0=0$ cm and truncated with a 20.4 cm × 20.4 cm (136λ × 136λ) square window.

The output RMSE values obtained from angular spectrum calculations are evaluated and plotted in 2D transverse planes along the z direction in Fig. 5. The RMS output errors computed with the input pressure and the input normal particle velocity both decrease monotonically as z increases. Figure 5 also shows that, for the combination of parameters evaluated here, the RMS output error obtained from the input normal particle velocity plane is about twice as large as that obtained with the input pressure plane.

C. Optimal parameters for the input plane

When the spectral propagator $H_p(k_x,k_y,\Delta z)$ for an input pressure plane described in Eq. (8) is used for angular spectrum simulations, the input pressure field is truncated by a rectangular window in the x and y directions. The size of this window and the location of the plane that contains the input pressure field both influence the accuracy of the result obtained with the angular spectrum approach. The optimal size and location of the input pressure plane are determined from parametric simulations of the planar phased array in Fig. 2, where the spatial sampling interval is $\delta=\lambda/2$, and the $N \times N$ grid for the 2D FFT is evaluated with $N=512$.

To demonstrate the impact of the input pressure plane location on angular spectrum calculations, the RMS output errors in a 3D volume are calculated in Fig. 6 as a function of the input pressure plane location z_0. The input pressure planes for these calculations are evaluated for z_0 ranging from 0 to 3.9 cm (26λ) with an interval of 0.15 cm (λ). The input pressure is calculated with the fast nearfield method using 20 abscissas for each integral. In Fig. 6(a), the pressure in the input plane is evaluated within a 7.8 cm × 7.8 cm (52λ × 52λ) square window, which is slightly larger than the

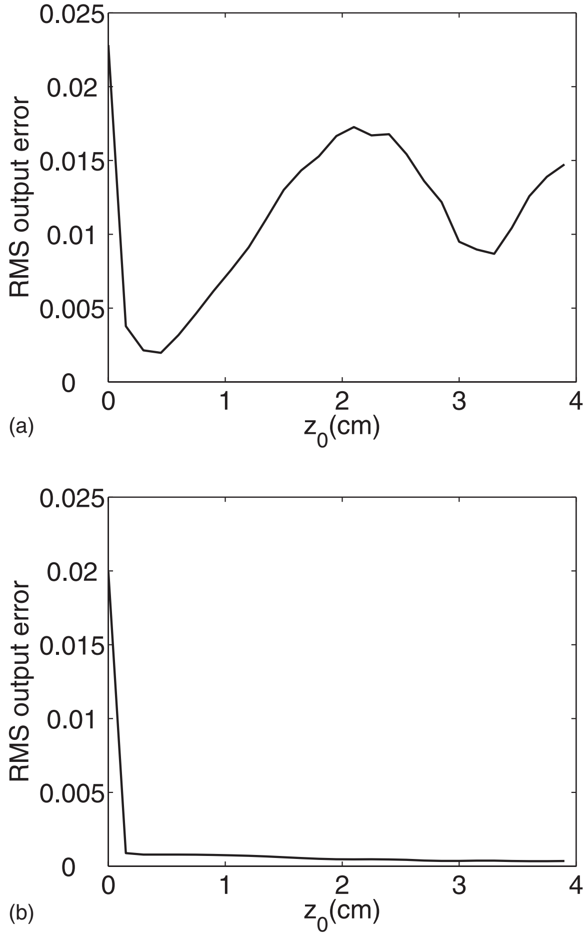


FIG. 6. RMS output errors obtained with the 32 × 32 element array evaluated in 3D volumes as a function of the input pressure plane location z_0. The size of the input pressure plane is (a) 7.8 cm × 7.8 cm (52λ × 52λ) and (b) 20.4 cm × 20.4 cm (136λ × 136λ). The location of the input pressure plane ranges between $z_0=0$ and $z_0=3.9$ cm (26λ) with an increment of 0.15 cm (λ).

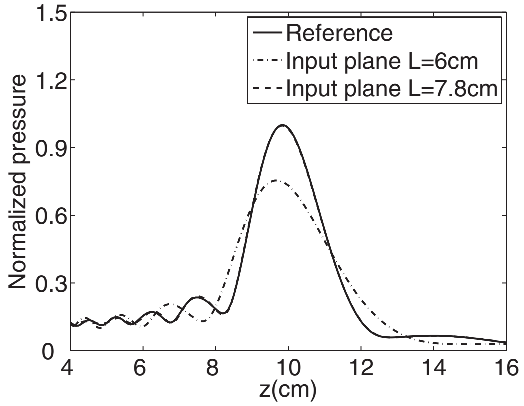


FIG. 7. Axial pressures simulated with the angular spectrum approach using input pressure planes that are truncated by square windows of different sizes. The reference pressure is indicated by the solid line, the output pressure computed with the angular spectrum approach using a 6 cm × 6 cm (40λ × 40λ) input pressure plane is represented by the dash-dot line, and the axial pressure computed with the angular spectrum approach using a 7.8 cm × 7.8 cm (52λ × 52λ) input pressure plane is represented by the dashed line. The solid line and the dashed line are nearly coincident, which indicates that $L=7.8$ cm ($L=52\lambda$) is sufficiently large for the 7.31 cm × 7.31 cm phased array in Fig. 2.

7.31 cm × 7.31 cm (48.7λ × 48.7λ) array aperture. The resulting RMS output error decreases for a short distance and then oscillates between 0 and 0.02 as z_0 increases. In Fig. 6(b), the input pressure is evaluated in a 20.4 cm × 20.4 cm (136λ × 136λ) plane. The 20.4 cm × 20.4 cm input pressure plane also includes the contribution from the grating lobes for each z_0. The RMS output error in Fig. 6(b) drops sharply when z_0 increases from 0 to 0.15 cm (λ), and then the error remains small for $z_0 \geqslant \lambda$. Figure 6 shows that when a smaller window is used, the RMS output error oscillates as z_0 changes, but the output error is relatively flat for a much larger window with $z_0 \geqslant \lambda$. In Fig. 6(a), the minimum error occurs at $z_0=0.45$ cm (3λ); however, the error is also small at $z_0=\lambda$. In Fig. 6(b), the error is very small for all values of $z_0 \geqslant \lambda$. Figure 6 shows that the input pressure plane should be at least one wavelength from the array aperture to avoid sampling problems with evanescent waves near the array aperture. Figure 6(a) also suggests that when the window that truncates the input pressure plane is slightly larger than the array aperture, the resulting error is sufficiently small for z_0 equal to 0.15 cm (λ). Thus, for either a 7.8 cm × 7.8 cm or a 20.4 cm × 20.4 cm window, $z_0=\lambda$ is an appropriate input pressure plane location for the 32 × 32 element array. For smaller input pressure planes, selecting $z_0=\lambda$ successfully prevents the truncation of grating lobes that could otherwise occur with larger values of z_0.

To demonstrate the influence of the size of the input pressure plane window on the RMS output error, the angular spectrum calculations are evaluated for a 6 cm × 6 cm (40λ × 40λ) input pressure plane and a 7.8 cm × 7.8 cm (52λ × 52λ) input pressure plane located at $z_0=\lambda$. The resulting axial pressures are shown in Fig. 7. The input pressure plane extent specified by $L=6$ cm (40λ) is smaller than the 7.31 cm × 7.31 cm array aperture, so the resulting pressure field (dash-dot line) deviates from the reference by a significant amount, especially in the region around the focus. When the extent of the input pressure plane is specified by $L=7.8$ cm (52λ), the axial pressure (dashed line) closely matches the reference pressure (solid line) in Fig. 7.

Figure 8 shows the RMS output errors evaluated in a 3D volume, where the input pressure planes are truncated by

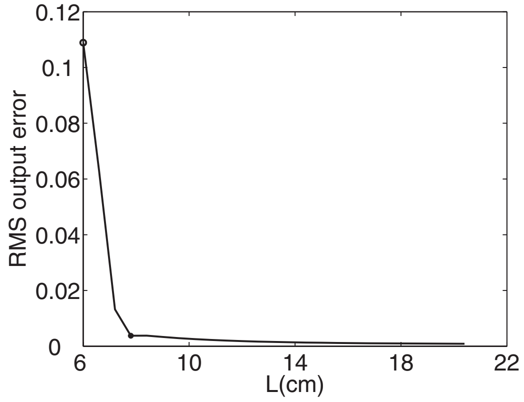


FIG. 8. RMS output errors plotted as a function of the extent L of the input pressure plane. The input pressure plane for angular spectrum simulations is located at $z_0=0.15$ cm ($z_0=\lambda$) and truncated by $L \times L$ square windows, where L ranges from 6 cm (40λ) to 20.4 cm (136λ) with an increment of 0.15 cm (λ). The two markers indicate the values of L shown in Fig. 7.

square windows with sizes ranging from L=6 cm (L=40λ) to L=20.4 cm (L=136λ). For all of the results shown in Fig. 8, the input pressure plane is located at z_0=λ. The two markers in Fig. 8 indicate the values of L evaluated in Fig. 7, where the circle denotes L=6 cm (L=40λ) and the solid mark denotes L=7.8 cm (L=52λ). In Fig. 8, the errors monotonically decrease as L increases. The errors are larger when the window size is smaller than the 7.31 cm × 7.31 cm array aperture, and the errors are smaller when the window size is larger than the array aperture. Figure 8 suggests that only a moderate reduction in the output error is achieved for values of L>7.8 cm. Furthermore, in Fig. 7, the results for L=7.8 cm are nearly coincident with the reference. If L<7.8 cm is used, the truncation of the pressure wavefront causes an increase in the RMS output errors. An input pressure plane with L>7.8 cm consistently produces small errors. However, there is a trade-off between the accuracy and the efficiency of these calculations for larger L. On the one hand, accurate results are achieved when large values of L are used, and larger input pressure planes are often necessary for array simulations that also include grating lobes, as in Fig. 3. On the other hand, L should be as small as possible because the computation time and computer memory required are proportional to the number of grid points, which is determined by the value of L when δ is fixed. Figure 6 indicates that the minimum error is achieved at z_0=0.45 cm (z_0=3λ) and that the error at z_0=0.15 cm (z_0=λ) is acceptably small, and Fig. 8 suggests that L =7.8 cm (L=52λ) is optimal for the 32 × 32 element planar array in Fig. 2.

D. Evaluation of input and output errors

In angular spectrum calculations that use the spectral propagator $H_p(k_x, k_y, \Delta z)$, the input pressure is typically simulated with analytical integral approaches, and the angular spectrum simulations then evaluate the output pressure in a 3D volume. For these simulations, fast and accurate calculations of the input pressure are desirable. This motivates numerical evaluations of the errors associated with the input pressure that impact the error in the computed 3D pressure output.

Using the spatial impulse response method as the reference, two analytical integral methods are compared for simulations of the input pressure: the Rayleigh–Sommerfeld integral and the fast nearfield method. The accuracy of these methods is determined by the number of abscissas used in evaluations of the integrals in Eqs. (1) and (4), respectively. As demonstrated in Ref. 24, the fast nearfield method achieves higher accuracy with fewer abscissas relative to other single integral approaches for simulations of single transducers. This result also holds for phased array simulations. To demonstrate the change in the RMSE in the input pressure plane as the number of abscissas increases, the pressure field is computed in a 7.8 cm × 7.8 cm (52λ × 52λ) plane at a depth of z_0=0.15 cm (z_0=λ), where 7.8 cm (52λ) is the optimal value of L determined in the previous section for the phased array in Fig. 2. With a spatial sampling interval of 0.075 cm (λ/2), the input pressure plane is discretized to 105 points in both the x and the y directions. The RMSE values are evaluated for the input pressure computed with the Rayleigh–Sommerfeld integral using 2 × 2 to 10 × 10 abscissas and with the fast nearfield method using two to ten abscissas. The results are plotted in Fig. 9, where the errors from both methods decrease as the number of abscissas increases. The RMSE for the input pressure obtained with the Rayleigh–Sommerfeld integral approach is 0.216 for 2 × 2 abscissas, and the error decreases to 0.006 for 10 × 10 abscissas. In contrast, the RMSE for the input pressure obtained with the fast nearfield method is 0.076 for two abscissas, and the error quickly decreases to 0.0004 with only four abscissas. In Fig. 9, errors less than or equal to 0.0004 are coincident with the horizontal axis when the values in the range shown are plotted on a linear scale.

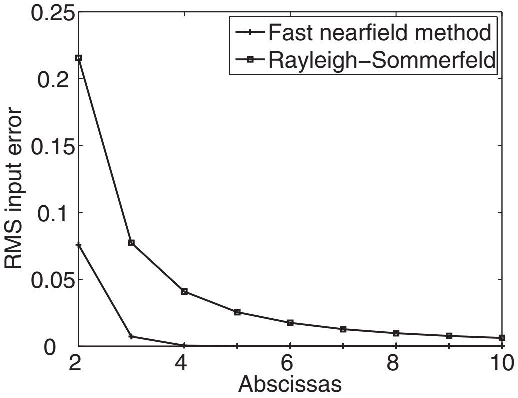


FIG. 9. RMSE values in the input pressure plane plotted as a function of the number of abscissas. The input pressure plane is located at z_0=0.15 cm (z_0=λ) and truncated by a 7.8 cm × 7.8 cm (52λ × 52λ) window. The input pressure is computed with the Rayleigh–Sommerfeld integral using 2 × 2 to 10 × 10 abscissas and with the fast nearfield method using two to ten abscissas.

Figure 10 demonstrates the influence of the number of abscissas used for input pressure calculations on the 3D angular spectrum results. In Fig. 10, the horizontal axis contains the number of abscissas for input pressure calculations with the Rayleigh–Sommerfeld integral and the fast nearfield

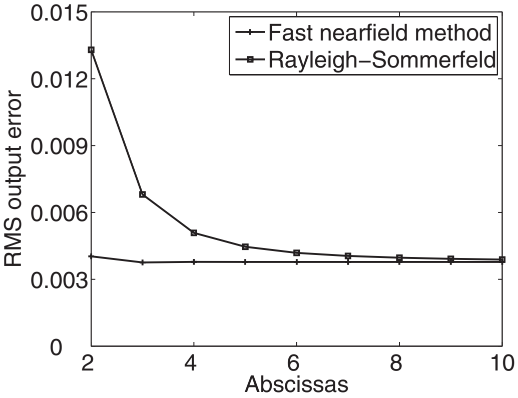


FIG. 10. RMSE values for 3D pressure field outputs plotted as a function of the number of abscissas used for input pressure calculations. The input pressure is calculated with the Rayleigh–Sommerfeld integral using 2 × 2 to 10 × 10 abscissas and with the fast nearfield method using two to ten abscissas. The errors obtained from both methods approach the same limiting value, but the fast nearfield method achieves convergence with far fewer abscissas.

method, and the vertical axis contains the output RMSE values in the 3D pressure field computed with the angular spectrum approach. The output RMSE approaches a limiting value of 0.004 when the fast nearfield method with three or more abscissas calculates the input pressure. When the input pressure is calculated with the Rayleigh–Sommerfeld integral, the output error asymptotically approaches the same value. The saturation of the RMSE in Fig. 10 indicates that there is a lower limit for the output error in angular spectrum calculations that is determined by the grid size, the grid spacing, and the location of the input pressure plane. Moreover, Fig. 10 shows that the fast nearfield method requires far fewer abscissas to achieve the minimum output error. Therefore, the fast nearfield method is a much more efficient approach for calculating the input pressure.

E. Temperature simulations

In thermal therapy simulations, the power deposition is generally modeled by Eq. (14). The resulting power deposition provides the input to the BHTE,[29] which simulates the temperature distribution. To determine the influence of the angular spectrum simulation parameters on the calculated temperature, the bio-heat transfer model in Eq. (15) is evaluated for the 32×32 element planar array in Fig. 2, which generates a single focus at $(0,0,10)$ cm. In these simulations, the temperature field is computed in a 7.8 cm $\times$7.8 cm$\times$12 cm $(52\lambda\times 52\lambda\times 80\lambda)$ volume, where the boundaries of the computational grid are maintained at 37 °C, the blood perfusion is 8 $\mathrm{kg/m^3/s}$, the thermal conductivity is 0.55 W/m/°C, and the specific heat of blood is 4000 J/kg/°C. The goal of each simulation is to elevate the temperature at the focus to 43 °C for hyperthermia cancer therapy[31] or for targeted drug delivery.

The temperature fields are calculated with power depositions as inputs, and the power depositions are obtained from the pressure fields calculated with the angular spectrum approach. The results from three types of inputs are compared for these angular spectrum calculations: (1) an input normal particle velocity plane, (2) an input pressure plane obtained from the Rayleigh–Sommerfeld integral, and (3) an input pressure plane obtained from the fast nearfield method. The reference temperature distribution is computed from the power deposition calculated with the spatial impulse response method. In these simulations, the input particle velocity plane is coincident with the array surface at $z_0=0$, whereas both of the input pressure planes are located at $z_0=0.15$ cm $(z_0=\lambda)$. The extent of each input pressure plane is 7.8 cm$\times$7.8 cm $(52\lambda\times 52\lambda)$, and the computational grid is discretized with a sampling rate of $\delta=0.075$ cm $(\delta=\lambda/2)$. An $N\times N$ grid with $N=512$ is used in all angular spectrum calculations. The power deposition corresponding to the reference pressure is normalized such that the resulting reference temperature field has a maximum value of 43 °C. The power depositions obtained from the angular spectrum simulations are normalized by the same factor.

The resulting axial temperature field evaluated along the array normal is shown in Fig. 11. In this figure, the solid line represents the reference, the dash-dot line is obtained from the angular spectrum approach using an input normal particle velocity plane, the dotted line with "+" markers is obtained from the angular spectrum approach using an input pressure plane computed with two abscissas applied to the fast nearfield method, and the dashed line is obtained from the angular spectrum approach using an input pressure computed with 2×2 abscissas applied to the Rayleigh–Sommerfeld integral. Figure 11 shows that when the angular spectrum calculation is performed with an input normal particle velocity plane, the largest errors are in the focal zone, and the simulated axial temperature field deviates from the reference by as much as 0.40 °C, where the maximum target temperature rise is 6 °C. When the input pressure for angular spectrum calculations is computed with the Rayleigh–Sommerfeld integral using 2×2 abscissas, the largest deviations in the simulated temperature field are again located in the focal zone, and the maximum axial temperature difference is 0.45 °C. When the input pressure is computed with the fast nearfield method using two abscissas, the axial temperature field closely matches the reference, and the maximum axial temperature difference is 0.027 °C, which is more than an order of magnitude smaller than the maximum axial temperature errors computed for the other two methods. Figure 11 shows that the temperature obtained from the reference and the temperature obtained from the results of the angular spectrum approach are almost indistinguishable when the input pressure is computed with the fast nearfield method using two abscissas. However, if the input pressure for the angular spectrum simulation is computed with the Rayleigh–Sommerfeld integral approach using 2×2 abscissas or with

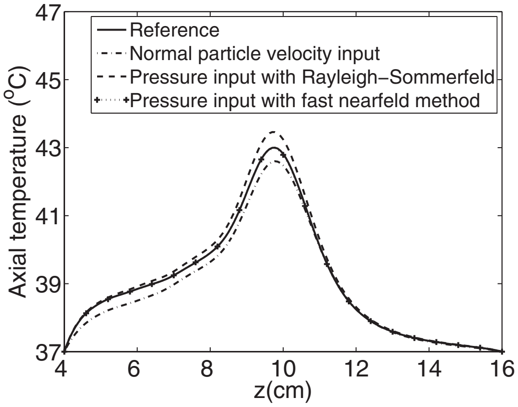


FIG. 11. Axial temperatures computed with the BHTE for power depositions calculated with the angular spectrum approach using different input planes. The pressures are generated by the 32×32 element planar phased array in Fig. 2, which is electronically focused at $(0,0,10)$ cm. The temperature obtained from the reference power deposition is indicated by the solid line, the temperature obtained from the power deposition calculated with the angular spectrum approach using the input normal particle velocity plane is indicated by the dash-dot line, the temperature obtained from the power deposition calculated with the angular spectrum approach when the input pressure is computed with the Rayleigh–Sommerfeld integral using 2×2 abscissas is represented by the dashed line, and the temperature obtained from the power deposition calculated with the angular spectrum approach when the input pressure is computed with the fast nearfield method using two abscissas is represented by the dotted line with "+" markers. The result obtained with the angular spectrum approach where the input pressure is computed with the fast nearfield method is nearly coincident with the reference temperature field, whereas the other simulated temperature fields contain noticeable errors.

the input normal particle velocity, the simulated temperature contains noticeable errors. As shown in Fig. 10, the output errors in the 3D pressure field are the same when the input pressure is computed with the Rayleigh–Sommerfeld integral using 7×7 abscissas or with the fast nearfield method using two abscissas. In addition, the simulated 3D temperature fields demonstrate that the temperature distribution obtained from the angular spectrum approach using the input normal particle velocity plane underestimates the temperature field everywhere and that the temperature distribution obtained from the angular spectrum approach combined with the Rayleigh–Sommerfeld calculation using 2×2 abscissas overestimates the temperature field everywhere. In contrast, the temperature distribution obtained from the angular spectrum approach when the input pressure is computed with the fast nearfield method closely matches the reference temperature throughout the 3D volume. This suggests that the fast nearfield method is the preferred method for computing the input pressure in angular spectrum calculations, especially for thermal therapy simulations with large ultrasound phased arrays.

IV. DISCUSSION

A. Simulations with other arrays

The simulation results from Sec. III were also validated with two other planar arrays driven by a 1 MHz continuous wave source. The first was a 14.45 cm $\times$ 14.45 cm ($96.67\lambda \times 96.67\lambda$) planar array containing 50×50 square elements. The size of each element was 2.4 mm $\times$ 2.4 mm ($1.6\lambda \times 1.6\lambda$), and the kerf between adjacent elements was 0.5 mm. The first array was electronically focused at $(0,0,12)$ cm. The computational grid for this array extended from 6 cm (40λ) to 18 cm (120λ) in the z direction. The second was a 4.55 cm $\times$ 4.55 cm ($30.33\lambda \times 30.33\lambda$) planar array containing 20×20 elements. The size of each element was 1.8 mm $\times$ 1.8 mm ($1.2\lambda \times 1.2\lambda$), and the kerf between adjacent elements was 0.5 mm. The second array was electronically focused at $(0,0,8)$ cm. The computational grid for the second array extended from 3 cm (20λ) to 12 cm (80λ) in the z direction. The sampling rate in the x, y, and z directions was $\delta=\lambda/2$ for both arrays. Angular spectrum simulation results computed for these arrays consistently demonstrated that smaller errors were achieved with the input pressure plane than with the input normal particle velocity plane. Results also showed that the smallest errors were obtained when the input pressure plane was located a short distance from the array and that $z_0=0.15$ cm ($z_0=\lambda$) produced acceptably small errors. The largest errors were consistently obtained when the input pressure plane was coincident with the array aperture ($z_0=0$). For the 20×20 element array, the optimal value of L was 4.8 cm (32λ) or 1.056 times the lateral extent of the array aperture, and for the 50×50 element array, the optimal value of L was 14.7 cm (98λ) or 1.017 times the lateral extent of the array aperture. In general, the optimal input plane size is between 1 and 1.1 times the lateral extent of the array aperture when the input pressure plane is located at $z_0=\lambda$.

Other focal patterns were simulated with the 32×32 element planar array in Fig. 2, including a pressure field with a steered focus at $(0,3,10)$ cm and a pressure field with a mode[32] consisting of two symmetric foci. These pressure patterns are of interest for thermal therapy applications because most tumors are much larger than the size of a single focal spot. The location of the input pressure plane and the size of the truncating window had similar influence on both the off-axis focal pattern and the multiple-focus pattern. The temperature fields for the 32×32 element planar array with a steered focus and multiple focal spots were also simulated with the BHTE after the pressure was computed with the angular spectrum approach. The error in the temperature distributions was again very small when the input pressure was calculated by the fast nearfield method in a plane located one wavelength from the array aperture, where the truncating window was slightly larger than the array aperture.

A 38×38 element spherical-section array[33] with square elements was also simulated with the angular spectrum approach. The opening angles in both lateral directions were equal to 60°. Each element in this array was 0.24 cm high and 0.24 cm wide. The array was geometrically focused at 12 cm. The pressure generated by this spherical-section array was also obtained with the angular spectrum approach. As for each of the results presented in previous sections, accurate results were obtained when the input pressure was calculated by the fast nearfield method in a plane truncated by a window slightly larger than the array aperture and located one wavelength from the nearest point on the array aperture.

B. The size and location of the input pressure plane

The location z_0 and the extent L of the input pressure plane determine the accuracy of the calculated pressure field. As demonstrated in Fig. 6(b), when a sufficiently large L is used, consistently small errors are achieved for any $z_0 \geqslant \lambda$. In contrast, Fig. 6(a) shows that when an intermediate value of L is used, the error oscillates as z_0 varies. In this case, the input pressure plane should be closer to $z_0=\lambda$. When $z_0=\lambda$ and $L=7.8$ cm ($L=52\lambda$) for the array in Fig. 2, the computed axial pressures are coincident with the reference pressure. Figures 7 and 8 show that an input pressure plane with $L=7.8$ cm is sufficient for this array when $z_0=\lambda$. A larger input pressure plane is required to capture the wave energy when z_0 is larger because the pressure wavefront in the corresponding input plane is broader. However, larger input pressure planes require more computer memory and computation time.

C. The spatial sampling rate

The spatial sampling interval δ is an important parameter in angular spectrum simulations. Undersampling in the spatial domain leads to aliased spectra in the spatial frequency domain. In Fig. 5, the large errors produced by the input normal particle velocity plane or the input pressure plane located at $z_0=0$ are largely due to aliasing. In the phased array model evaluated here, the normal particle velocity distribution on each transducer element is represented by a 2D rect function. The discontinuities at the edge of each element and at the edge of the array aperture introduce high spatial frequency components that are inherently aliased. Thus, the normal particle velocity field encounters some sampling difficulties that cause aliasing. In contrast, the changes in the input pressure distribution are less abrupt. In fact, the angular spectra for the input pressure plane sampled by $\delta = \lambda/2$ and $\lambda/4$ are similar, while the angular spectra for the input particle velocity plane sampled by $\delta = \lambda/2$ and $\lambda/4$ differ by a significant amount due to aliasing.

For either the normal particle velocity or the input pressure, the output error is reduced by decreasing the spatial sampling interval. By decreasing δ from $\lambda/2$ to $\lambda/4$, the maximum RMSE for the result obtained from the input particle velocity plane is reduced from 0.084 in Fig. 5 to 0.035, and the maximum RMSE for the result obtained from the input pressure is reduced from 0.044 to 0.005. The RMSE values in Fig. 10 are also reduced by decreasing δ. However, smaller values of δ increase the computation time and the amount of computer memory required, and as shown in Figs. 7 and 11, $\lambda/2$ sampling is sufficient for angular spectrum simulations of the 32×32 element planar array when the input pressure is computed with the fast nearfield method using appropriate values for L and z_0.

D. The spectral sampling rate

The spectral sampling interval Δk describes the resolution of the angular spectrum. If the angular spectrum is undersampled, wrap-around errors will appear in the reconstructed spatial field. The spectral sampling interval is determined by the relationship $\Delta k = 2\pi/(N\delta)$, where increasing N enhances the angular resolution of the angular spectrum and reduces wrap-around errors for a fixed value of δ. In the simulations presented here, the spectral propagator is evaluated within a 512×512 grid. The input pressure plane is discretized to 113 points in the x and the y directions and then zero-padded on a 512×512 grid before the 2D FFT is performed. If no zero-padding is used, the RMSE for the simulated pressure in the $113 \times 113 \times 161$ grid is about 37 times higher than the result obtained from the $512 \times 512 \times 161$ grid. However, as indicated earlier, $N = 512$ is sufficiently accurate for the results presented here. Further increases in N will result in an unnecessary increase in the computation time and the amount of computer memory.

The numerical error can be especially large for simulations in non-attenuating media, as demonstrated in Ref. 17. An angular restriction technique[15] that applies a lowpass filter to the spectral propagator can eliminate some of the high spatial frequency components in the angular spectrum that contribute to the numerical error. However, the numerical simulations shown here are evaluated in an attenuating medium, so the excessive high spatial frequency spectra are filtered out by the attenuation term in Eq. (10). The angular spectrum simulations in this paper produce accurate results without angular restriction, which is not needed for simulations in attenuating media.[17]

E. Computation time

The angular spectrum approach computes pressures in parallel planes by propagating fields in the spatial frequency domain. This method reduces the computation time significantly compared to conventional integral approaches, which compute the pressure at individual field points and superpose the results. The pressure field in a 7.8 cm × 7.8 cm × 12 cm volume is discretized to a $105 \times 105 \times 161$ grid when the spatial sampling rate is $\delta = 0.075$ cm. To compute the pressure field generated by the 32×32 element planar array in Fig. 2 in this 3D grid, the Rayleigh–Sommerfeld integral calculation in Eq. (1) is completed in 46.15 min if each integral is evaluated with 2×2 abscissas. In contrast, the angular spectrum approach with $N = 512$ only uses 2.12 min to compute the pressure in the same grid, where 0.52 min of this time is spent computing the 105×105 point input pressure plane with the fast nearfield method using two abscissas. In comparison, the Rayleigh–Sommerfeld integral with 7×7 abscissas computes the input pressure plane in 2.84 min and achieves similar accuracy, so the total calculation time for a 3D pressure field with the angular spectrum approach is 4.44 min. Based on this analysis and the results presented in previous sections, the fast nearfield method is preferred for calculations of the input pressure plane in angular spectrum simulations.

V. CONCLUSION

The angular spectrum approach is a computationally efficient method for simulating 3D pressure fields generated by large ultrasound phased arrays comprised of hundreds or thousands of elements. The results show that the input pressure plane produces more accurate simulation results than the input normal particle velocity plane in angular spectrum computations. In addition, for angular spectrum calculations performed with an input pressure plane, the largest errors are obtained when this plane is coincident with the array aperture, and much smaller errors are obtained when this plane is located one wavelength away from the array aperture. Furthermore, the error in the simulated pressure field decreases as the extent L of the input pressure plane increases. When L reaches a sufficiently large value, the error in the simulated pressure field becomes very small. The optimal value of L is between 1 and 1.1 times the lateral extent of the array aperture. Results also show that the output errors from angular spectrum computations asymptotically approach a limiting value as the number of abscissas used for input pressure plane calculations increases. To achieve this error limit in the computed 3D pressure field, fewer abscissas are required by the fast nearfield method than the Rayleigh–Sommerfeld integral approach for input pressure plane calculations due to the rapid convergence of the fast nearfield method. Evaluations of angular spectrum results in bio-heat transfer simulations demonstrate that the angular spectrum approach combined with the fast nearfield method achieves much smaller errors than the angular spectrum approach combined with the Rayleigh–Sommerfeld integral approach or angular spectrum calculations with an input normal particle velocity plane. Thus, the angular spectrum approach is an accurate and robust method for thermal therapy simulations with large ultrasound phased arrays when the input pressure is computed with the fast nearfield method in a plane located one wavelength away from the array and truncated by a window slightly larger than the array aperture.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grant No. R21CA121235 and NSF Theoretical Foundations Grant No. 0634786.

[1] E. S. Ebbini and C. A. Cain, “Multiple-focus ultrasound phased-array pattern synthesis: Optimal driving signal distributions for hyperthermia,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 540–548 (1989).

[2] E. G. Moros, X. Fan, and W. L. Straube, “An investigation of penetration depth control using parallel opposed ultrasound arrays and a scanning reflector,” J. Acoust. Soc. Am. **101**, 1734–1741 (1997).

[3] K. Y. Saleh and N. B. Smith, “Two-dimensional ultrasound phased array design for tissue ablation for treatment of benign prostatic hyperplasia,” Int. J. Hyperthermia **20**, 7–31 (2004).

[4] K. B. Ocheltree and L. A. Frizzell, “Sound field calculation for rectangular sources,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 242–248 (1989).

[5] J. S. Tan, L. A. Frizzell, N. Sanghvi, S.-J. Wu, R. Seip, and J. T. Kouzmanoff, “Ultrasound phased arrays for prostate treatment,” J. Acoust. Soc. Am. **109**, 3055–3064 (2001).

[6] R. J. McGough, D. Cindric, and T. V. Samulski, “Shape calibration of a conformal ultrasound therapy array,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 494–505 (2001).

[7] 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).

[8] B. Piwakowski and K. Sbai, “A new approach to calculate the field radiated from arbitrarily structured transducer arrays,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 422–440 (1999).

[9] B. Piwakowski and K. Sbai, “A new calculation procedure for spatial impulse responses in ultrasound,” J. Acoust. Soc. Am. **105**, 3266–3274 (1999).

[10] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).

[11] E. G. Williams and J. D. Maynard, “Numerical evaluation of the Rayleigh integral for planar radiators using the FFT,” J. Acoust. Soc. Am. **72**, 2020–2030 (1982).

[12] D. P. Orofino and P. C. Pedersen, “Efficient angular spectrum decomposition of acoustic sources. 1. Theory,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 238–249 (1993).

[13] D. P. Orofino and P. C. Pedersen, “Efficient angular spectrum decomposition of acoustic sources. 2. Results,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 250–257 (1993).

[14] P. Wu, R. Kazys, and T. Stepinski, “Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I. Errors due to the discrete Fourier transform and discretization,” J. Acoust. Soc. Am. **99**, 1339–1348 (1996).

[15] P. Wu, R. Kazys, and T. Stepinski, “Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part II. Characteristics as a function of angular range,” J. Acoust. Soc. Am. **99**, 1349–1359 (1996).

[16] P. Wu, R. Kazys, and T. Stepinski, “Optimal selection of parameters for the angular spectrum approach to numerically evaluate acoustic fields,” J. Acoust. Soc. Am. **101**, 125–134 (1997).

[17] X. Zeng and R. J. McGough, “Evaluation of the angular spectrum approach for simulations of nearfield pressures,” J. Acoust. Soc. Am. **123**, 68–76 (2008).

[18] G. T. Clement and K. Hynynen, “Field characterization of therapeutic ultrasound phased arrays through forward and backward planar projection,” J. Acoust. Soc. Am. **108**, 441–446 (2000).

[19] P. Godden, G. ter Haar, and I. Rivens, “Numerical modelling of high intensity focused ultrasound phased arrays by an angular spectrum decomposition using highly-oscillatory quadrature,” in International Symposium on Therapeutic Ultrasound (2007), pp. 36–42.

[20] R. J. Zemp and J. T. Tavakkoli, “Modeling of nonlinear ultrasound propagation in tissue from array transducers,” J. Acoust. Soc. Am. **113**, 139–152 (2003).

[21] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).

[22] G. R. Linfield and J. Penny, *Numerical Methods Using MATLAB* (Prentice-Hall, Upper Saddle River, NJ, 1999).

[23] J. Zemanek, “Beam behavior within the nearfield of a vibrating piston,” J. Acoust. Soc. Am. **49**, 181–191 (1971).

[24] R. J. McGough, “Rapid calculations of time-harmonic nearfield pressures produced by rectangular piston,” J. Acoust. Soc. Am. **115**, 1934–1941 (2004).

[25] D. Chen and R. J. McGough, “A 2D fast nearfield method for calculating nearfield pressures generated by apodized rectangular pistons,” J. Acoust. Soc. Am. **124**, 1526–1537 (2008).

[26] M. S. Ibbini and C. A. Cain, “A field conjugation method for direct synthesis of hyperthermia phased-array heating patterns,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 3–9 (1989).

[27] P. R. Stepanishen and K. C. Benjamin, “Forward and backward projection of acoustic fields using FFT methods,” J. Acoust. Soc. Am. **71**, 803–812 (1982).

[28] D. Liu and R. C. Wagg, “Propagation and backpropagation for ultrasonic wavefront design,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1–13 (1997).

[29] H. H. Pennes, “Analysis of tissue and arterial blood temperatures in the resting human forearm,” J. Appl. Physiol. **1**, 93–122 (1948).

[30] K. B. Ocheltree and L. A. Frizzell,“Determination of power deposition patterns for localized hyperthermia: A steady state analysis,” Int. J. Hyperthermia **3**, 269–279 (1987).

[31] P. R. Stauffer, “Evolving technology for thermal therapy of cancer,” Int. J. Hyperthermia **21**, 731C744 (2005).

[32] R. J. McGough, H. Wang, E. S. Ebbini, and C. A. Cain, “Mode scanning: Heating pattern synthesis with ultrasound phased arrays,” Int. J. Hyperthermia **10**, 433–442 (1994).

[33] E. S. Ebbini and C. A. Cain, “A spherical-section ultrasound phased array applicator for deep localized hyperthermia,” IEEE Trans. Biomed. Eng. **38**, 634–643 (1991).

Decentralized harmonic control of sound radiation and transmission by a plate using a virtual impedance approach

Nicolas Quaegebeur,[a] Philippe Micheau, and Alain Berry
Department of Mechanical Engineering, GAUS, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, Québec J1K 2R1, Canada

(Received 4 December 2008; accepted 24 February 2009)

The problem under study in this article is the active control of sound transmission and radiation of a panel under a periodic excitation. The control strategy investigated uses independent control loops between an individual polyvinylidene fluoride (PVDF) sensor and an individual lead zirconate titanate (PZT) actuator. The specific approach employed here uses the concept of virtual impedance. The aim is to determine for each frequency the optimal impedance between each PVDF sensor and the corresponding PZT actuator in order to reduce the sound power radiated by the plate. Theoretical predictions are compared to measurements of the sound radiated and transmission loss of a panel mounted with eight PZT-PVDF units. Reductions of up to 20 dB of the acoustic power can be achieved around mechanical resonances of the system, while the control strategy has little effect for off-resonance excitations. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106124]

PACS number(s): 43.40.Vn, 43.50.Ki, 43.50.Gf [ADP] Pages: 2978–2986

I. INTRODUCTION

Active control is an efficient approach for attenuating low frequency vibrations and sound radiation of structures. Classically, the vibrational response is sensed at a number of locations on the structure and modified by a number of local actuators using a centralized controller. The sensors and the actuators must be located in order to sense and control structural modes of interest. Different technologies of actuator/sensor couples have been tested in previous studies.[1,2] Piezoelectric materials appear to be good candidates for active vibration control of plates because under pure bending assumption they can form collocated, dual actuator-sensor pairs.[3]

The use of piezoelectric materials for active control of vibration and sound transmission has been investigated in the past using centralized algorithms and state space estimation[4] or optimal solution derived from theoretical observations.[2] However, a centralized control strategy involves modeling a large number of secondary transfer functions, requires cumbersome wiring, and is prone to instability due to plant uncertainty or individual actuator or sensor failure.

The problem under study in this article is the active control of bending vibrations and sound radiation of a panel and the control strategy investigated is the use of independent control loops between an individual polyvinylidene fluoride (PVDF) sensor and an individual lead zirconate titanate (PZT) actuator instead of a centralized controller. The main advantage of such a decentralized control strategy is its reduced complexity, reduced processing requirement, ease of implementation, and robustness to individual control unit failure. However, performance and stability of decentralized control are difficult to predict *a priori*. Decentralized control approaches were applied to the active control of free-field sound radiation using loudspeaker-microphone pairs,[5] the active control of sound radiation and transmission using velocity feedback control (active damping)[6,7] where predictions were compared to measurements for a plate excited by a loudspeaker in a box, or the harmonic control of local strain in a plate.[8]

Under the assumption of collocated and dual actuator-sensor pairs, decentralized control has the very attractive property that each local feedback loop (with the other feedback loops being active) is stable regardless of the local feedback gains applied, leading to a globally stable and robust implementation. When applied to globally reducing the vibration response of panels, or sound transmission through panels, decentralized control leads to control performance very similar to a fully centralized control structure.[7] However, piezoelectric strain actuators (PZT) and strain sensors (PVDF) cannot, strictly speaking, be collocated and dual because of coupling through extensional excitation.[6] As a result, instability may occur at low frequency.

Most previous work on decentralized control used velocity feedback loops (active damping) to control the sound radiation or transmission by a panel. More recently, mixed active/passive strategies have been developed using active damping combined to added virtual mass/stiffness. Theoretical studies on beams and plates[9] show the potential of this approach and experimental results were recently published on control of sound transmission using re-active passive devices.[10]

When using velocity feedback control (active damping) or active mass/damping units, one parameter (real controller gain for active damping or complex gain for active mass/damping) is set for all frequencies and all units. The specific situation investigated here is the active control of sound ra-

[a] Author to whom correspondence should be addressed. Electronic mail: nicolas.quaegebeur@gmail.com

TABLE I. Panel data.

Parameter	Value
Thickness	$h=3.18\times 10^{-3}$ m
Length	$Lx=48\times 10^{-2}$ m
Width	$Ly=42\times 10^{-2}$ m
Young's modulus	$E=68.5\times 10^{9}$ Pa
Poisson's coefficient	$\nu=0.33$
Density	$\rho=2700$ kg m^{-3}
Loss factor	$\eta=5.0\times 10^{-3}$

TABLE II. Piezoelectric data.

Parameter	PZT actuator	PVDF sensor
Thickness	$h_A=1.02\times 10^{-3}$ m	$h_S=28\times 10^{-6}$ m
Length	$Lx_A=2.54\times 10^{-2}$ m	$Lx_S=1.5\times 10^{-2}$ m
Width	$Ly_A=2.54\times 10^{-2}$ m	$Ly_S=1.5\times 10^{-2}$ m
Young's modulus	$E=61.1\times 10^{9}$ Pa	
Poisson's coefficient	$\nu=0.29$	
Piezoelectric coefficient	$d_{31}=-1.9\times 10^{-10}$ C/N	
Piezoelectric coefficient		$e_{31}=46\times 10^{-3}$ C/m^2
Capacitance		$C_S=5\times 10^{-10}$ F

diation or transmission of a panel under a periodic excitation using a virtual impedance approach. The aim is to determine for each frequency the optimal impedance between each sensor (PVDF film) and the corresponding actuator (PZT patch) in order to reduce the sound power radiated (SPR) by the plate. The strategy is based on an estimation of the diagonal terms of the plant matrix and the implementation of the harmonic controller is achieved using a centralized or a decentralized algorithm, so that the present strategy can be fully decentralized.

Section II introduces the problem and details the plant modeling. The harmonic control strategy using an impedance approach and its practical implementation are presented in Secs. III and IV. Finally, experimental results on sound transmission and sound radiation are presented in Sec. V and compared to theoretical predictions.

II. MECHANICAL MODEL

A. Plate equation

We consider a rectangular, simply-supported panel whose characteristics are given in Table I equipped with eight surface-mounted, identical actuator-sensor pairs, as presented in Fig. 1. Each pair consists of a rectangular piezoceramic (PZT) actuator and a rectangular PVDF sensor, which are here collocated on the panel and whose characteristics are given in Table II. The location, size, and number of units are chosen so that all bending modes of the plate are controllable below 1 kHz.

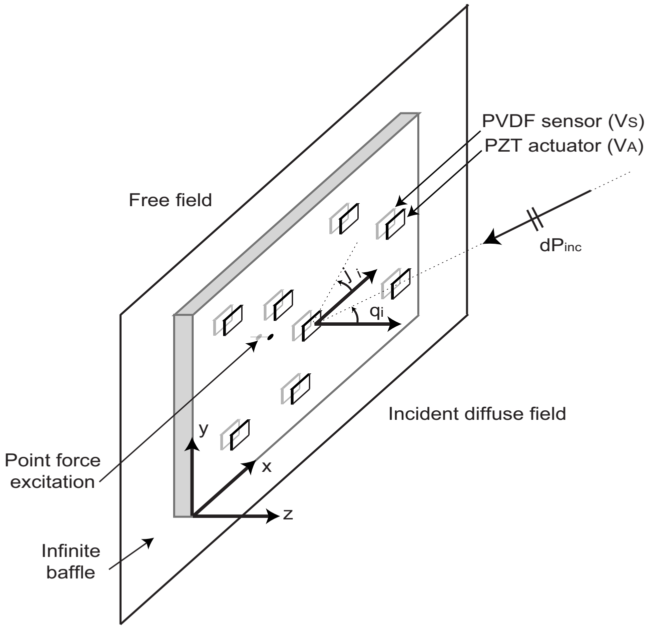


FIG. 1. Transmission Loss Setup with smart panel having eight collocated PZT/PVDF units.

In the following analysis, pure bending response is assumed; therefore, the effect of extensional deformation of the panel on the actuator-sensor transfer functions is not considered. Previous studies mention that accounting for both bending and extensional deformations leads to non-duality of the system, especially in low frequencies.[11] The transverse displacement $w(x,y,t)$ is solution of

$$D\Delta\Delta w(x,y,t)+M\ddot{w}(x,y,t)+C\dot{w}(x,y,t) = F^{\mathrm{pri}}(x,y,t)+F^{\mathrm{sec}}(x,y,t), \quad (1)$$

where Δ is the Laplacian, $D=Eh^3/12(1-\nu^2)$ represents the stiffness of the plate, $M=\rho h$ is the mass per unit area, C denotes the viscous damping, F^{pri} is the primary sources (acoustical or mechanical excitation), and F^{sec} is the secondary sources (sum of the forces applied by each piezoelectric actuator). The solution is classically expanded on the linear modes of the structure considering a harmonic regime of angular frequency ω:

$$w(x,y,t)=\exp(j\omega t)\sum_{p=0}^{\infty}\Phi_p(x,y)q_p, \quad (2)$$

where the modal shapes are computed assuming simply-supported boundary conditions:

$$\Phi_p(x,y)=\sin\left(\frac{m\pi x}{Lx}\right)\sin\left(\frac{n\pi y}{Ly}\right) \quad \text{where } p=(m,n). \quad (3)$$

After projection of Eq. (1) on mode p, one obtains for the complex phasor q_p associated with mode p:

$$q_p=\frac{F_p^{\mathrm{pri}}+F_p^{\mathrm{sec}}}{\omega_p^2-\omega^2+j\eta\omega}, \quad (4)$$

where $\eta=C/M$ denotes the modal damping and ω_p denotes the natural angular frequency of mode p:

$$\omega_p=\sqrt{\frac{D}{M}}\left(\left(\frac{m\pi}{Lx}\right)^2+\left(\frac{n\pi}{Ly}\right)^2\right), \quad (5)$$

and all the phasors of the external forces are expanded over the mode p:

$$F_p^\star = \int\int_S F^\star(x,y)\Phi_p(x,y)dS, \tag{6}$$

where $(.)^\star$ denotes "pri" or "sec" defined in Eq. (4).

B. Diffuse field excitation

A diffuse acoustic field is characterized by an infinite number of uncorrelated incident plane-waves. The acoustic transmission under plane-wave incidence has been treated by Roussos[12] and then developed in the case of diffuse field by Nelisse,[13] who showed that the transmission loss (TL) factor has to be calculated by integrating (numerically with a Gaussian quadrature scheme) the contributions of all plane-waves of modulus $|P_{\mathrm{inc}}|$ under incidence angles (θ_i, ϕ_i) defined in Fig. 1. For an incident plane-wave of orientation (θ_i, ϕ_i), one obtains the following formulation for primary excitation under harmonic excitation with angular frequency ω:

$$F_p^{\mathrm{pri}} = \frac{8}{2\pi}|P_{\mathrm{inc}}|I_m I_n, \tag{7}$$

with

$$I_m = \frac{-i}{2}\mathrm{sgn}(\sin\theta_i\cos\phi_i) \quad \text{if} \quad \left(\frac{m\pi}{kLx}\right)^2 = (\sin\theta_i\cos\phi_i)^2,$$

$$I_m = \frac{m\pi(1-(-1)^m\exp(-i\sin\theta_i\cos\phi_i kLx))}{(m\pi)^2 - (\sin\theta_i\cos\phi_i kLx)^2} \quad \text{otherwise} \tag{8}$$

and

$$I_n = \frac{-i}{2}\mathrm{sgn}(\sin\theta_i\sin\phi_i) \quad \text{if} \quad \left(\frac{n\pi}{kLy}\right)^2 = (\sin\theta_i\sin\phi_i)^2,$$

$$I_n = \frac{n\pi(1-(-1)^n\exp(-i\sin\theta_i\sin\phi_i kLy))}{(n\pi)^2 - (\sin\theta_i\sin\phi_i kLy)^2} \quad \text{otherwise}, \tag{9}$$

where $k=\omega/c$ is the acoustic wavenumber. For each orientation (θ_i, ϕ_i) of plane-waves incident on the plate, the mechanical excitation and the radiation of the system must be calculated and then integrated in order to compute the radiation under diffuse field excitation. The expressions I_m and I_n are also used in the calculation of the far field radiation of the plate.

In the case of a mechanical excitation by a point force located in (x_i, y_i), the modal primary force is expressed as follows:

$$F_p^{\mathrm{pri}} = F^{\mathrm{pri}}\Phi_p(x_i, y_i), \tag{10}$$

where F^{pri} denotes the phasor of the transverse force applied.

C. Piezoelectric coupling

Neglecting the effects of longitudinal waves in the plate, the force applied on mode p by the piezoelectric actuators is given by the sum of contributions of each unit j:[6,14]

$$F_p^{\mathrm{sec}} = -\left(\frac{C_0 d_{31}}{4\rho h h_A Lx Ly}\right)\sum_j V_j^A \Gamma_p^j \Lambda_p^j, \tag{11}$$

where C_0 is a coefficient determined from the piezoelectric constants[8] of the actuators, V_j^A represents the phasor of the tension applied to unit j, and all piezoelectric constants are given in Table II:

$$\Gamma_p^j = \left(\frac{\gamma_m^2+\gamma_n^2}{\gamma_m\gamma_n}\right), \quad \gamma_m = \frac{m\pi}{Lx}, \quad \gamma_n = \frac{n\pi}{Ly}, \tag{12}$$

and

$$\Lambda_p^j = (\cos\gamma_m x_1^j - \cos\gamma_m x_2^j)(\cos\gamma_n y_1^j - \cos\gamma_n y_2^j), \tag{13}$$

where x_1^j, x_2^j, y_1^j, and y_2^j represent the positions of the jth actuator boundaries in the coordinate system of Fig. 1. In the same way, for collocated PVDF sensor, the local bending strains in the plate induce a voltage, defined after modal reconstruction by:[8]

$$V_i^S = -\frac{e_{31}}{C_S}\left(\frac{h+h_S}{2}\right)\sum_p q_p \Gamma_p^i \Lambda_p^i. \tag{14}$$

This equation assumes that the PVDF sensor and the PZT actuators have identical dimensions and are exactly collocated on both sides of the plate.

D. Matrix formulation

In order to solve the elastic and electric variables of the problem numerically, a matrix formulation of the mechanical equations [Eq. (4)] coupled with the piezoelectric colocalized units [Eqs. (11) and (14)] is used. A large enough number N of modes are taken into account in the modal expansion [Eq. (2)] such that $\omega_N \gg \omega_0$. where ω_0 represents the natural angular frequencies in the considered bandwidth. In the simulations, a modal truncation of $N=100$ modes is considered in order to observe the local effect of piezoelectric units.

$$\mathbf{A}X = \mathbf{B}V^A + P,$$

$$V^S = \mathbf{C}X, \tag{15}$$

where the state vector, observation vector, secondary voltage, and primary force vectors are defined as follows:

$$X = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \quad V^S = \begin{pmatrix} V_1^S \\ \vdots \\ V_J^S \end{pmatrix}, \quad V^A = \begin{pmatrix} V_1^A \\ \vdots \\ V_J^A \end{pmatrix}, \quad P = \begin{pmatrix} F_1^{\mathrm{pri}} \\ \vdots \\ F_N^{\mathrm{pri}} \end{pmatrix},$$

and the state matrix are defined by

$$\mathbf{A} = \begin{pmatrix} \omega_1^2 - \omega^2 + j\eta\omega & & 0 \\ & \ddots & \\ 0 & & \omega_N^2 - \omega^2 + j\eta\omega \end{pmatrix},$$

$$\mathbf{B} = C_B\begin{pmatrix} \Gamma_1^1\Lambda_1^1 & \cdots & \Gamma_1^J\Lambda_1^J \\ \vdots & \ddots & \vdots \\ \Gamma_N^1\Lambda_N^1 & \cdots & \Gamma_N^J\Lambda_N^J \end{pmatrix},$$

$$\mathbf{C} = C_C\begin{pmatrix} \Gamma_1^1\Lambda_1^1 & \cdots & \Gamma_N^1\Lambda_N^1 \\ \vdots & \ddots & \vdots \\ \Gamma_1^J\Lambda_1^1 & \cdots & \Gamma_N^J\Lambda_N^J \end{pmatrix},$$

where $C_B=-(C_o d_{31})/(4\rho h h_A LxLy)$ and $C_C=-e_{31}(h+h_S)/(2C_S)$ represent mechanical coefficients related to the PZT actuators or PVDF sensors, respectively. Applying an oscillatory voltage of angular frequency ω on an individual PZT actuator generates forced bending vibrations of the panel which are sensed by all PVDF films. The relation between actuator response V^A, sensor response V^S, and primary disturbance P is defined for each frequency by

$$V^S = \mathbf{H}V^A + \mathbf{G}P, \tag{16}$$

where $\mathbf{H}=C\mathbf{A}^{-1}B$ is the matrix of transfer functions between individual actuators and sensors and $\mathbf{G}=C\mathbf{A}^{-1}$. In the case of an isolated, collocated PZT/PVDF pair, and under the assumption of pure bending, it can be easily shown that the product of the actuator input voltage V_j^A and the sensor output voltage rate $j\omega V_j^S$ is proportional to the power supplied to the panel: such a collocated PZT/PVDF pair is therefore dual and defined for each frequency in matrix form by

$$Y = \mathbf{M}U + D, \tag{17}$$

where $Y=j\omega V^S$, $U=V^A$, $D=j\omega V^S|_{V^A=0}$, and $\mathbf{M}=j\omega\mathbf{H}$ is the mobility matrix of the system at angular frequency ω. The passivity of the matrix $\mathbf{M}$ is defined by

$$\mathrm{Re}(\lambda_i(\mathbf{M})) \geq 0, \tag{18}$$

where $\lambda_i(\mathbf{M})$ denotes the eigenvalues of matrix $\mathbf{M}$. The passivity of the system is theoretically guaranteed if and only if Eq. (18) is guaranteed for every frequency and the system should therefore be dual and collocated.

E. Radiated sound power

The formulation given by Roussos[12] is used to compute the sound radiation of the plate in the far field and the acoustic power is obtained by integration of the radiated sound field over a hemisphere surrounding the plate:

$$\Pi_{\mathrm{rad}} = X^H \ \mathbf{Rad} \ X, \quad \mathbf{Rad} = \begin{pmatrix} R_1^2 & \cdots & R_1R_N \\ \vdots & \ddots & \vdots \\ R_NR_1 & \cdots & R_N^2 \end{pmatrix}, \tag{19}$$

where

$$R_p = -\omega^2 \int_{\theta=0}^{\theta=\pi/2} \int_{\phi=0}^{\phi=2\pi} I_m I_n \sin\theta d\phi d\theta \tag{20}$$

and the integrations are performed numerically using a Gaussian quadrature scheme. The Transmission Loss (TL) factor is computed as follows:

$$\mathrm{TL} = 10\log_{10}\left(\frac{\Pi_{\mathrm{inc}}}{\Pi_{\mathrm{rad}}}\right), \tag{21}$$

where $\Pi_{\mathrm{inc}}=|P_{\mathrm{inc}}^2|L_xL_y\cos\theta_i/\rho_f c$ represents the incident power in the diffuse field and ρ_f and c represent the density and sound speed.

III. CONTROL: STRATEGY AND IMPLEMENTATION

A. Impedance control

When using velocity feedback control (active damping) or active mass/damping units, one parameter (real control gain for active damping or complex gain for active mass/damping) is needed for all frequencies and all units. We propose here to study the harmonic control of sound transmission using a virtual impedance approach. The aim is to determine for each frequency the optimal impedance between dual variables: PVDF sensor voltage rate ($j\omega V^S$) and PZT actuator voltage (V^A) in order to reduce the SPR by the plate. The virtual impedance formulation consists in imposing the following matrix relation between the phasors of the two dual variables:

$$U = \mathbf{Z_D}Y, \tag{22}$$

where $\mathbf{Z_D}$ is a complex diagonal matrix, so that the control strategy is decentralized and defined for each unit independently by $V^A=j\omega\mathbf{Z_D}V^S$, where $\mathbf{Z_D}$ is a complex variable. From a theoretical point of view, matching impedances are obtained when $\mathbf{Z_D}=\mathbf{M}_d^{-1}$, where $\mathbf{M}_d=\mathrm{diag}(\mathbf{M})$. Hence, in order to conduct an exhaustive study of impedance control, we propose to introduce two real parameters $(\alpha_{\mathrm{Re}},\alpha_{\mathrm{Im}})$ that scale the nondimensional virtual impedance:

$$\mathbf{Z_D} = \mathbf{Z_R} + j\mathbf{Z_I} = \alpha_{\mathrm{Re}}\Re(\mathbf{M}_d^{-1}) + j\alpha_{\mathrm{Im}}\Im(\mathbf{M}_d^{-1}). \tag{23}$$

With nondual variables (V^A,V^S), Eqs. (22) and (23) become

$$V^A = \mathbf{Z_{ND}}V^S,$$

$$\mathbf{Z_{ND}} = \alpha_{\mathrm{Im}}\Re(\mathbf{H}_d^{-1}) + j\alpha_{\mathrm{Re}}\Im(\mathbf{H}_d^{-1}), \tag{24}$$

where $\mathbf{H}_d$ and $\mathbf{M}_d$ represent the diagonal terms of the matrices $\mathbf{H}$ and $\mathbf{M}$ defined in Eqs. (16) and (17). Thus, according to Eqs. (23) and (24), the virtual impedances imposed by the different units are different, but they are related through the two parameters α_{Re} and α_{Im}. The existence of a solution is ensured if and only if the matrix $(\mathbf{Id}-\mathbf{Z_{ND}}\mathbf{H})$ is invertible. Since $\mathbf{Z_{ND}}$ is diagonal, only the marginal cases where $\mathbf{H}$ is diagonal and $\mathbf{Z_{ND}}=\mathbf{H}^{-1}$ [equivalent to $\alpha_{\mathrm{Re}}=\alpha_{\mathrm{Re}}=1$ in Eq. (24)] lead to the non-existence of a solution of Eq. (22). The aim is to estimate for each frequency the sound power reduction for each value of nondimensionalized terms $(\alpha_{\mathrm{Re}},\alpha_{\mathrm{Im}})$ between -5 and 5 and then to choose the optimal couple that represents, respectively, the virtual active damping and active mass/stiffness added to the plate at the considered frequency.

The equivalent electrical circuit for the ith unit using dual description is represented in Fig. 2. In this equivalent electrical circuit, the primary disturbance generates a current D_i, the ith PVDF sensor provides the total voltage rate output Y_i which is applied to the virtual impedance to generate the actuator voltage U_i.

B. Integral MIMO controller

From a theoretical point of view, the electrical equivalent circuit is stable for any added dissipative impedance (for $Z_R \geq 0$). From a practical point of view, the actuator voltage controlled by the sensor voltage can lead to unstable feed-

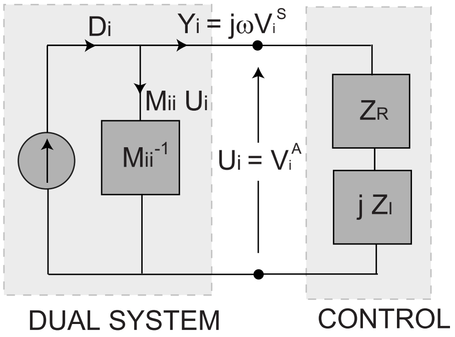


FIG. 2. Impedance control: electrical equivalent circuit for the ith unit.

back loops. In order to study active added impedance ($Z_R < 0$) and to ensure stable feedback loop, the impedance approach presented in Eq. (24) is not directly implemented in a controller. We propose an iterative integral Multi Input Multi Output (MIMO) controller to reach the actuator voltage that ensures the control objective defined in Eq. (24) through the minimization of $|V^A - \mathbf{Z}_{\mathbf{ND}}V^S|$. At step k, the relation between actuator and sensor phasors $V^A[k]$ and $V^S[k]$ is expressed under its matrix form by

$$V^A[k+1] = V^A[k] - \mu\mathbf{K}(V^A[k] - \mathbf{Z}_{\mathbf{ND}}V^S[k]), \qquad (25)$$

where μ is the loop-gain and $\mathbf{K}$ is a compensation matrix which is invertible, diagonal for decentralized algorithm, and nondiagonal for centralized algorithm. The stability of the algorithm described by Eq. (25) is ensured if and only if

$$|1 - \mu\lambda_i(\mathbf{K} - \mathbf{K}\mathbf{Z}_{\mathbf{ND}}\mathbf{H})| < 1, \quad \forall\, i = 1, \ldots, J, \qquad (26)$$

and in this case, the algorithm is stopped when the sensor phasors V^A are closed to the solution defined in Eq. (24), i.e., when $|V^A - \mathbf{Z}_{\mathbf{ND}}V^S| < \epsilon$, where ϵ is a small value determined by the user. Three different algorithms (and therefore three different compensation matrices $\mathbf{K}$) are proposed in order to implement

(1) a centralized algorithm: $\mathbf{K_1} = (\mathbf{Id} - \mathbf{Z}_{\mathbf{ND}}\mathbf{H})^H$,
(2) a decentralized algorithm: $\mathbf{K_2} = (\mathbf{Id} - \mathbf{Z}_{\mathbf{ND}}\mathbf{H}_d)^H$, and
(3) a diagonal compensation matrix defined as a rotation: $\mathbf{K_3} = \exp(j\phi)\mathbf{Id}$: This case is a general case of the rotation algorithm[15] when the loop-gain is not set to 1.

In the frequency and virtual impedance range considered in the present study, the centralized algorithm 1 appears to be always stable. For the decentralized algorithm 2, instability appears for certain added impedances because the stability condition defined in Eq. (26) is no longer respected. Only the algorithm 3 can guarantee the stability of the decentralized algorithm for all frequencies and added virtual impedances if the two parameters μ and ϕ are correctly set. Figure 3 illustrates a typical case where a region of stability for the decentralized controller is defined as a function of rotation angle ϕ and loop-gain μ. A necessary condition of stability of algorithm 3 derived from Eq. (26) is that all the eigenvalues of matrix ($\mathbf{Z}_{\mathbf{ND}}\mathbf{H}$) must lie in a cone of apex angle lower than π.[15] This allows to choose in all cases a diagonal control matrix $\mathbf{K}$ and thus to keep a decentralized controller. Algorithms 1, 2, and 3 were tested in simulations; however, only algorithm 1 was implemented experimentally.

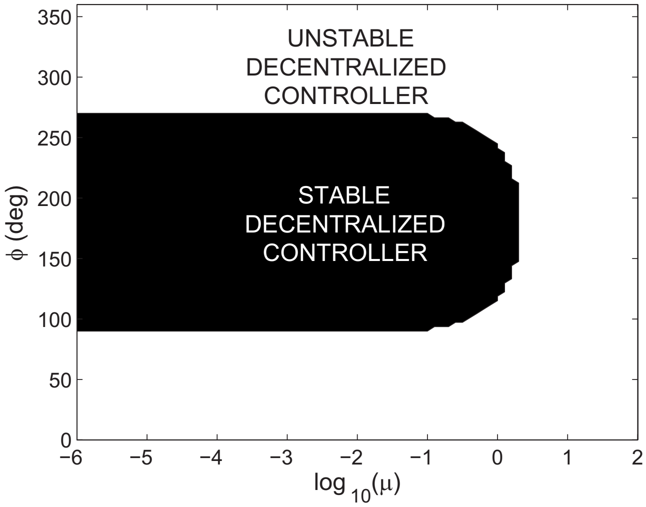


FIG. 3. Zones of stability of the decentralized controller using algorithm 3 as a function of loop-gain μ and rotation angle ϕ [defined in Eq. (26)] for a given frequency and a given virtual impedance.

IV. EXPERIMENTAL SETUP AND PROCEDURE

A. Implementation of the controller

The controller has been implemented in a rapid prototyping dSpace system whose sampling frequency was set to 4 kHz. The retained algorithm corresponds to case 1 (centralized algorithm) which is always stable in the bandwidth of interest [0:1 kHz]. A total of eight PZT-PVDF units are used as described in Fig. 1 and the block diagram of the control loop is shown in Fig. 4.

In practice, the implementation of the frequency-domain control uses demodulation and modulation blocks[8] that extract the phasor of the error $V^S(t)$ and generate the oscillatory control input $V^A(t)$, respectively. The modulation and demodulation blocks make use of low-pass Finite Impulse Re-

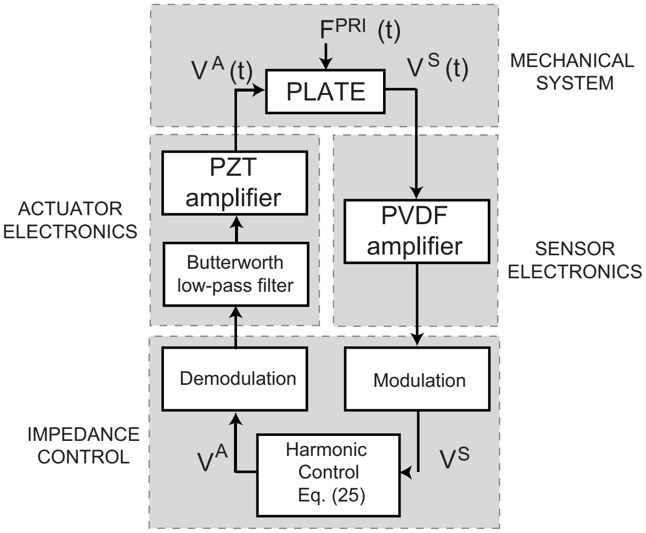


FIG. 4. Block diagram of the experimental setup.

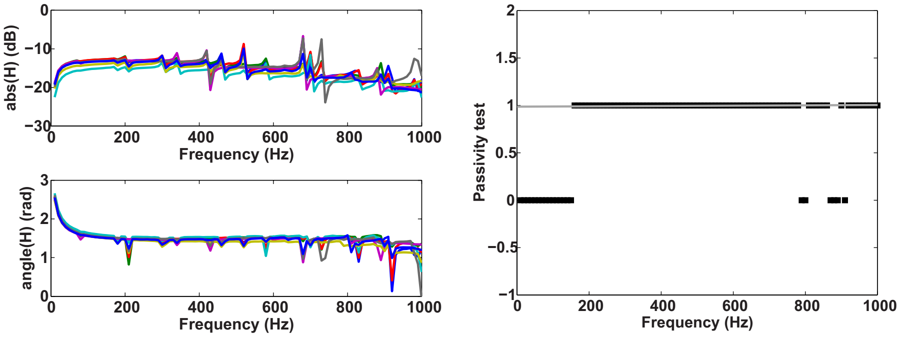


FIG. 5. (Color online) Left: Diagonal terms of the experimental transfer matrix **M** between dual variables (U,Y). The delays due to the low-pass anti-aliasing filters and the acquisition system have been compensated. Right: Passivity test (1 if passive, 0 otherwise) of the system measured (black) and estimated (gray).

sponse (FIR) filters to extract the phasor input V^S and generate the phasor output V^A which are implemented with four Butterworth second-order filters in order to ensure an attenuation of 40 dB outside ±7.5 Hz. The harmonic controller on bottom of Fig. 5 operates with a sampling frequency of 25 Hz on the complex envelopes.

A low-pass eighth order Butterworth filter is added at the output of the dSpace controller in order to prevent aliasing. Its cut-off frequency is set to 1500 Hz. The electronics for the PZT actuator consists of a tension amplifier and the PVDF amplification system is ensured by a current amplifier.

All the electronics (low-pass filter and amplification of piezoelectric sensor and actuator) and the controller itself are responsible for delays (phase shifts) that have to be compensated in order to recover dual signals (see Fig. 5).

B. Mechanical system identification

Figure 5 represents the diagonal terms of experimental dual transfer matrix **M** and the passivity test of the system (measured and estimated). When the system is passive, a simple velocity loop between each collocated pair would provide an unconditionally stable feedback control. However, the coupling of both the PZT and PVDF with extensional deformation of the panel makes the analysis more complicated. When considering extensional deformation, it turns out that a collocated PZT/PVDF pair is no longer dual in the general case.[3] This can be observed in Fig. 5 below 200 Hz and for certain frequencies (around 800 Hz).

In the experimental part of this paper, the results are thus limited to the frequency range where the system is passive and little effect can be obtained below 200 Hz due to longitudinal wave response and limited effectiveness of piezoelectric transducers in low frequency.

C. Experimental setup

In order to measure the TL factor of the plate and sound radiation under mechanical excitation before and after control, the plate instrumented with eight PZT-PVDF units is mounted between a reverberant chamber (whose volume is approximately 100 m^3) and a hemi-anechoic room. Figure 6 shows the setup viewed from the hemi-anechoic room. The acoustic excitation in the reverberant room is ensured by a unique loudspeaker so that the average incident sound pressure level measured by a rotating microphone is approximatively 90 dB. Alternatively, an electrodynamic shaker located in the reverberant room is used to create a mechanical excitation of the plate. The separation between the reverberant and anechoic rooms around the plate is ensured a double wall with 2 cm thick gypsum boards, filled with acoustic foam. The plate is mounted on flexible supports along its edges, which prevent any transverse displacement but allow for free rotation, thus approaching simply-supported boundary conditions.

In order to measure the sound field radiated by the plate in the anechoic room, the technique of near-field acoustic holography (NAH) is employed using an array of 36 microphones (square array of 6×6 microphones with inter-microphone spacing of 15 cm) located at 10 cm of the plate, as proposed by Williams.[16] The sound pressure, the normal velocity, and the sound intensity in the plane of the plate are computed using reconstruction algorithms proposed by Veronesi and Maynard[17] and the radiated sound power is then estimated by integrating the real part of the surface intensity for each added virtual impedance and each frequency.

D. Experimental procedure

The objective of the procedure is to search for the optimal virtual impedance, which minimizes the SPR by the plate at a given frequency. The procedure is automatically executed for frequencies between 100 and 1000 Hz by increments of 10 Hz in order to perform an exhaustive research.

(1) Measure the transfer matrix **H** for the selected frequency.
(2) For $(\alpha_{\mathrm{Re}}, \alpha_{\mathrm{Im}})$ between −5 and 5 by steps of 0.5.
(a) Compute $\mathbf{Z}_D$ using Eq. (23).
(b) Compute the loop-gain μ using Eq. (26).

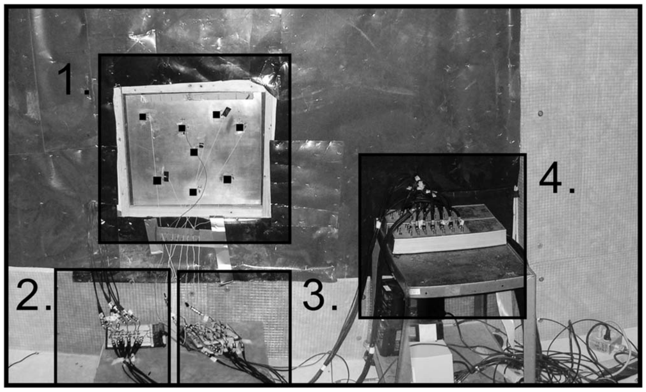


FIG. 6. Photography of the experimental TL setup viewed from the hemi-anechoic room. The plate (1) with eight surface-mounted piezoelectric patches (black squares), the PVDF sensor electronics (2), the PZT actuator electronics (3), and the dSpace acquisition board (4).

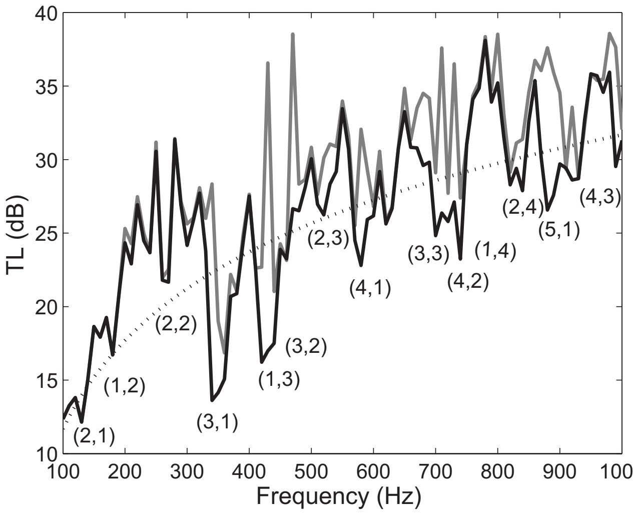


FIG. 7. TL with (gray) and without control (black). The mass law is indicated by a dotted line and plate modes p are identified using the convention $p=(m,n)$ according to Eq. (3).

(c) Search the command U with Eq. (25).
(d) Measure the SPR using NAH.

(3) Search for the optimal values $(\alpha_{\text{Re}},\alpha_{\text{Im}})$ that minimize the SPR.

V. EXPERIMENTAL RESULTS OF ACTIVE CONTROL

A. TL control

1. TL measurements

The TL measured by step of 10 Hz without control and with optimal impedance control is presented in Fig. 7 and the maximal attenuation obtained for each frequency is represented in Fig. 8. In Fig. 7, the mass law indicated by dotted line represents the overall behavior of the TL and is defined by $\text{TL}_M(f)=-42.4+20\log(mf)$, where m is the mass of the plate and f is the frequency. Around mechanical resonances, the structure is more excited and radiates more than for off-resonance excitations, so that a decrease in TL is observed.

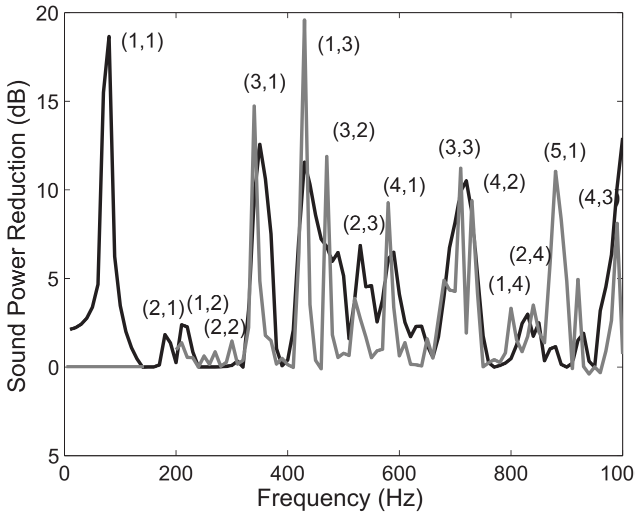


FIG. 8. Comparison of the measured (gray) and predicted (black) attenuation of TL after virtual impedance control.

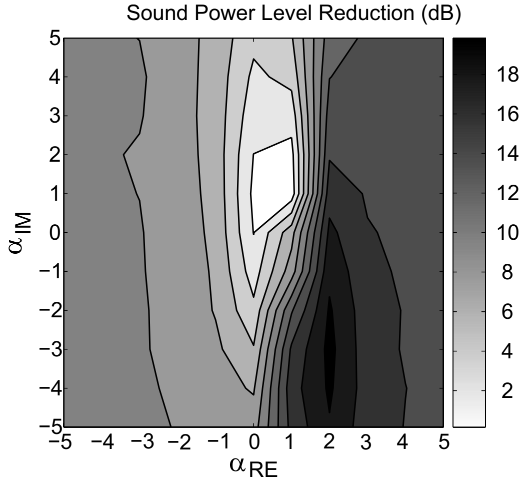


FIG. 9. Measured reductions (in dB) of the SPR by the plate at 430 Hz in the case of acoustic diffuse field excitation as a function of the variables α_{Re} and α_{Im} defined in Eq. (23).

The most radiative plate modes are odd-odd, symmetric modes (monopole type sound radiation) such as (3,1) or (1,3) modes.

Below 200 Hz, the control performance is negligible due to the coupling of both the PZT and PVDF with extensional deformation of the panel which is not taken into account in our model. This explains why the attenuations below 200 Hz predicted by the model are not observed in practice. The strategy begins to be effective above 200 Hz and the (1,1) mode at 77 Hz can thus not be controlled experimentally.

The average attenuation over the whole bandwidth [10 Hz: 1000 Hz] is about 3.5 dB. Fortunately, the control strategy becomes effective at those frequencies, and maximal attenuations are obtained [up to 18 dB at the resonance of the (1,3) mode at 430 Hz]. For off-resonance excitation, the structure radiates less, and the observed attenuations are below 3 dB and are fairly well predicted by the model. Certain frequencies cannot be controlled due to the placement of the units. Indeed the observability and controllability of the system are determined by the location and size of the units[14] and therefore the frequency range where the control strategy is effective could be increased by using a large number of units, or larger units.

2. Analysis around the resonance of the (1,3) mode

In order to illustrate the action of the controller, the results obtained at 430 Hz [near the resonance of the (1,3) mode] are analyzed. Figure 9 represents the measured reductions (in dB) of the SPR by the plate at 430 Hz as a function of the variables α_{Re} and α_{Im} defined in Eq. (23). Around $(\alpha_{\text{Re}},\alpha_{\text{Im}})=(0,0)$ which is equivalent to turning the control off, the reductions are below 2 dB. Tuning $\alpha_{\text{Re}}\geq 2$ allows 10 dB of reduction of SPR and the optimal value is unique in the domain $(\alpha_{\text{Re}},\alpha_{\text{Im}})=[-5;+5]\times[-5;+5]$ and perfectly located at $(+2,-3)$.

Figure 10 presents the normal active intensity reconstruction obtained by the NAH algorithm at 430 Hz without control and with optimal added virtual impedance control $(\alpha_{\text{Re}},\alpha_{\text{Im}})=(+2,-3)$. The maximal value of intensity without

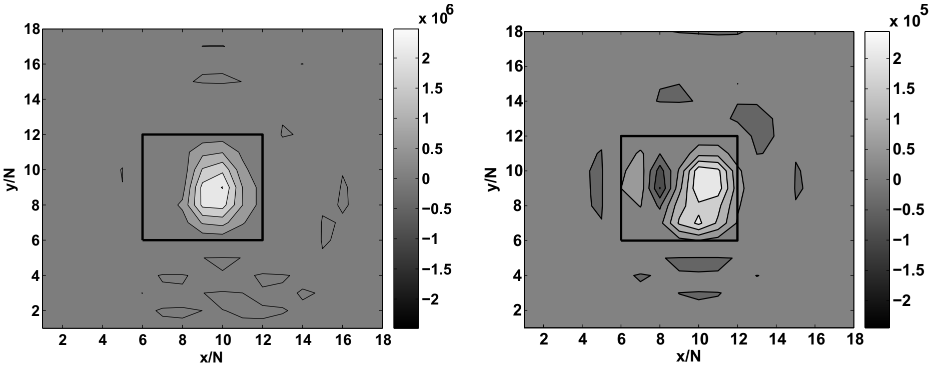


FIG. 10. Real part of the sound intensity around the resonance of the (3,1) mode at 430 Hz without (left) and with optimal control (right) in the case of acoustic diffuse field excitation. The plate location is indicated by a rectangle. For this frequency, a 18 dB reduction in sound power in closed loop is obtained.

control is about ten times larger than with control. This is due to the active damping component ($\alpha_{\mathrm{Re}} > 0$) of the controller at that frequency. A modification of vibration pattern is also observed and recirculation zones appear with control (negative parts of the intensity) which are responsible for a dipolar radiation pattern. This modification of vibrations is not present in the case of active damping and can thus be attributed to the imaginary part (α_{Im}) of the virtual added impedance. Both effects result in a decrease of 18 dB of the SPR by the plate at that frequency.

B. Mechanical excitation

For this experiment, the primary source is provided by a shaker located at $(x_i, y_i) = (0.103\ \mathrm{m}, 0.3\ \mathrm{m})$ and placed in the reverberant room. As previously, the SPR by the plate without and with control is estimated using the NAH technique and the results are represented in Fig. 11. The most radiative modes are again odd-odd, symmetric modes such as the (1,1), (3,1), or (3,3) and are fairly well controlled [except for the (1,1) mode at 77 Hz] such that the maximal value of SPR below 1 kHz decreases of 8 dB with control. The maximal attenuation obtained experimentally for each frequency is compared to theoretical results in Fig. 12 using a point force excitation, as presented in Eq. (10). A fairly good agreement between estimated and measured attenuations in closed loop is achieved, except below 200 Hz where both PZT and PVDF are coupled with extensional deformation of the panel (not taken into account in the theoretical model).

The average attenuation over the whole bandwidth [10 Hz: 1000 Hz] is about 4 dB and attenuations up to 20 dB of acoustic power are obtained experimentally around mechanical resonances of the plate. Only the reductions obtained experimentally around the resonances of the (4,1) and (2,3) modes cannot be explained theoretically. If we compare Figs. 8 and 12, it appears that the same frequencies cannot be controlled (for example, around 390 Hz). This is due to a misplacement of the units which influences the observability and controllability of the system at those frequencies, as mentioned previously.

C. Discussion: Optimal virtual impedance

The originality of the present approach is that the optimal virtual impedance induced by the piezoelectric units is set for each frequency and not for the whole bandwidth as it is the case of active damping or mass/damping approaches. Indeed, the active damping case corresponds to $\mathbf{Z_D}(\omega) = R$ with $R > 0$ constant for all frequencies and the mass/damping approach to the case $\mathbf{Z_D}(\omega) = R + j\omega M$ with $R > 0$. For each frequency, we analyzed the optimal impedance $\mathbf{Z_D}$ added to the system in dual formulation (valid above 200 Hz). For this purpose, all delays introduced by the electronic and control stages have to be compensated and the results are presented

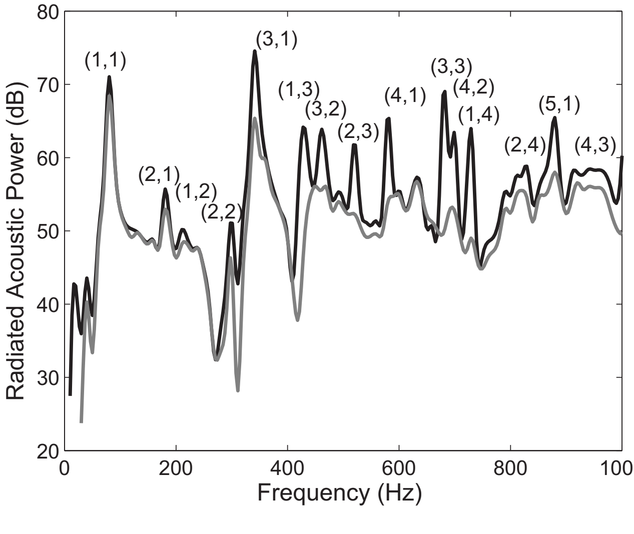


FIG. 11. SPR by the plate before (black) and after control (gray). Plate modes p are identified using the convention $p = (m, n)$ according to Eq. (3).

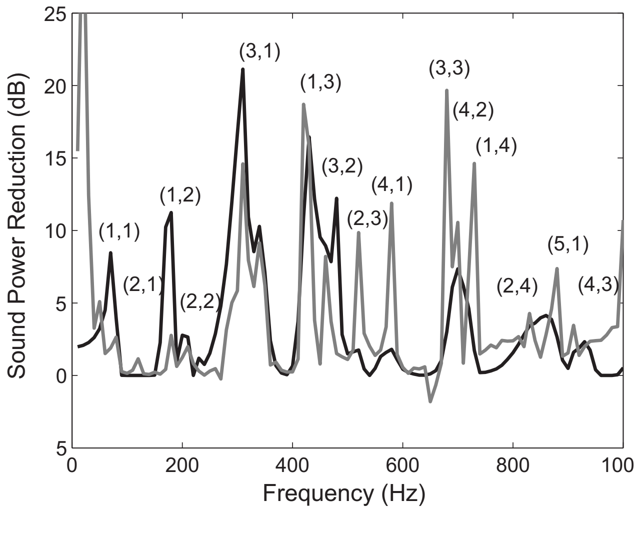


FIG. 12. Comparison of the measured (gray) and predicted (black) attenuation of radiated acoustic power.

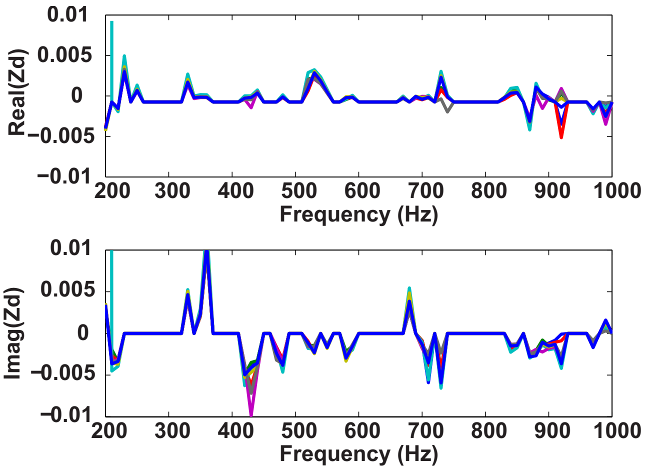


FIG. 13. (Color online) Optimal virtual added impedance matrix $\mathbf{Z_D}$. The impedance defined in dual variables has a positive real part for added damping and a negative imaginary part for added stiffness.

in Fig. 13 in the case of acoustical diffuse field excitation. When the control is off, the virtual impedance is set to 0 (no control). For each frequency, a positive real part of $\mathbf{Z_D}$ indicates that at this frequency, the controller acts as a virtual damper and a positive imaginary part indicates that the controller acts as a virtual added mass.

In Fig. 13, it appears that when the controller is effective, the added virtual impedance at the considered frequency is not only real (as in active damping) but has an imaginary part (that can be either positive or negative). This allows to obtain higher attenuations compared to active damping techniques.[7] In fact, when the excitation frequency is close to a resonance frequency, the controller acts so that the normal sound intensity field is modified and therefore sound radiation decreases. Since the impedance parameters are adjusted for each frequency, it is important to note that the conventional shift of resonance frequencies due to added mass or stiffness[9] cannot occur with the present method.

VI. CONCLUSION

In this paper, a decentralized strategy of control using virtual impedance approach is developed to reduce sound transmission and sound radiation of a panel equipped with piezoelectric units. The aim is to determine for each frequency the optimal impedance between PVDF sensor and PZT actuator in order to reduce the SPR by the plate. Theoretical predictions are compared to experimental measurements of the TL under incident diffuse field and with mechanical excitation ensured by a shaker. Looking at the optimal added virtual impedance, it appears that the active damping approach (velocity feedback controller) is not the optimal strategy but can be improved by choosing a complex impedance dependent on the frequency which allows to modify the vibration pattern of the panel. The strategy employed here allows high reductions up to 20 dB of the SPR by the plate when excited close to a resonance. However, when excited off resonance, the control strategy has little effect. Future work should be done to evaluate off-resonance control (by, for example, acting on the number, the size, and the placement of the units).

[1] B. Petitjean, I. Legrain, F. Simon, and S. Pauzin, “Active control experiments for acoustic radiation reduction of a sandwich panel: Feedback and feedforward investigations,” J. Sound Vib. **252**, 19–36 (2002).

[2] B.-T. Wang, C. R. Fuller, and E. K. Dimitriadis, “Active control of noise transmission through rectangular plates using multiple piezoelectric or point force actuators,” J. Acoust. Soc. Am. **90**, 2820–2831 (1991).

[3] Q. Sun, “Some observations on physical duality and collocation of structural control sensors and actuators,” J. Sound Vib. **194**, 765–770 (1996).

[4] J. S. Vipperman and R. L. Clark, “Multivariable feedback active structural acoustic control using adaptive piezoelectric sensoriactuators,” J. Acoust. Soc. Am. **105**, 219–226 (1999).

[5] E. Leboucher, P. Micheau, A. Berry, and A. L’Esperance, “A stability analysis of a decentralized adaptive feedback active control system of sinusoidal sound in free space,” J. Acoust. Soc. Am. **111**, 189–199 (2002).

[6] S. J. Elliott, P. Gardonio, T. C. Sors, and M. J. Brennan, “Active vibroacoustic control with multiple local feedback loops,” J. Acoust. Soc. Am. **111**, 908–915 (2002).

[7] P. Gardonio, E. Bianchi, and S. Elliott, “Smart panel with multiple decentralized units for the control of sound transmission. Parts I-II-III,” J. Sound Vib. **274**, 162–232 (2004).

[8] M. Baudry, P. Micheau, and A. Berrya, “Decentralized harmonic active vibration control of a flexible plate using piezoelectric actuator-sensor pairs,” J. Acoust. Soc. Am. **119**, 262–278 (2006).

[9] V. Lhuillier, L. Gaudiller, C. Pezerat, and S. Chesne, “Improvement of transmission loss using active control with virtual modal mass,” Advances in Acoustics and Vibration **2008**, 1–9 (2008).

[10] J. P. Carneal, M. Giovanardi, C. R. Fuller, and D. Palumbo, “Re-active passive devices for control of noise transmission through a panel,” J. Sound Vib. **309**, 495–506 (2008).

[11] Y. Brunet, “Unités actionneurs-capteurs colocalisées-duales pour le contrôle actif vibratoire décentralisé (Colocalized and dual actuator-sensor units for decantralized harmonic control of plate vibration),” MS thesis, GAUS, Université de Sherbrooke, Sherbrooke, QC (2008).

[12] L. Roussos, “Noise transmission loss of a rectangular plate in an infinite baffle,” NASA Technical Report No. TP2398, NASA, Langley Research Center, Hampton, VA, 1985.

[13] H. Nelisse, O. Beslin, and J. Nicolas, “Fluid-structure coupling for an unbaffled elastic panel immersed in a diffuse field,” J. Sound Vib. **198**, 485–506 (1996).

[14] D. Halim and S. R. Moheimani, “An optimization approach to optimal placement of collocated piezoelectric actuators and sensors on a thin plate,” Mechatronics **13**, 27–47 (2003).

[15] P. Micheau, R. Louviot, and A. Berry, “Decentralized resonant controller for vibroacoustic active control,” in 15th Mediterranean Conference on Control and Automation (2007).

[16] E. G. Williams, J. D. Maynard, and E. Skudrzyk, “Sound source reconstructions using a microphone array,” J. Acoust. Soc. Am. **68**, 340–344 (1980).

[17] W. Veronesi and J. D. Maynar, “Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer implementation,” J. Acoust. Soc. Am. **81**, 1307–1322 (1987).

Subjective evaluation of heavy-weight floor impact sounds in relation to spatial characteristics

Jin Yong Jeon,[a] Pyoung Jik Lee, Jae Ho Kim, and Seung Yup Yoo
Department of Architectural Engineering, Hanyang University, Seoul 133-791, Korea

(Received 22 October 2008; revised 10 January 2009; accepted 18 January 2009)

This study investigated the effect of a spatial factor, the magnitude of interaural cross-correlation (IACC) function, on subjective responses to heavy-weight floor impact sounds. Heavy-weight impact sounds were generated by a heavy/soft impact source (impact ball) in real apartments, so that impact sound pressure levels (SPLs) ($L_{A\max}$) and IACC could be analyzed. Just noticeable differences (JNDs) of impact SPL and IACC were investigated through the use of impact ball sounds. JNDs were determined by the criteria of 75% correct answers by participants, and it was found that JNDs of impact SPL and IACC were around 1.5 dB and 0.12–0.13, respectively. In addition, the annoyance caused by an impact ball was evaluated by changes in these two parameters. The results show that annoyance increased with increasing impact SPL and with decreasing IACC; the contributions of the two parameters to the scale value of annoyance were 79.3% and 20.4%, respectively. This indicates that the effects of IACC should be considered for the evaluation of annoyance, and the subjective response to impact ball sounds can be improved by controlling IACC, as well as impact SPL. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081390]

PACS number(s): 43.50.Ba, 43.50.Jh, 43.50.Pn [NX] Pages: 2987–2994

I. INTRODUCTION

Floor impact sounds are considered one of the most annoying noises in apartment buildings. Specifically, complaints by residents about heavy-weight impact sounds, caused by children jumping or running, have increased steadily in Korea. For evaluation of heavy-weight impact sounds, standard impact sources such as a bang machine and impact ball (heavy/soft impact source) have been used.[1–3] However, recent studies have shown that the physical characteristics of an impact ball are more similar to those of a real impact source, and the impact sound level of an impact ball is also more similar to a human-made impact sound than the bang machine.[4,5] Thus, only the impact ball was applied as a standard heavy-weight impact source for ISO 140–11; ISO 10140–3 also contains only the impact ball.[3,6]

Heavy-weight impact sounds are evaluated according to sound pressure level (SPL) in the frequency range of 63–500 Hz, and single number rating values such as *L -number* and $L_{i,F\max,AW}$ (inverse *A*-weighted impact SPL) are calculated on the basis of the maximum SPL.[7,8] In addition, noise indices, which showed a good correlation with annoyance caused by heavy-weight impact sounds, were reported in a previous study.[4] These noise indices are determined by only the SPL, which is categorized into temporal factors, and do not include any of the spatial attributes of the sound field.

Perception of loudness or annoyance caused by heavy-weight impact sounds were also mainly evaluated by temporal features such as SPL and pitch. The relationship between subjective response to floor impact sounds and noise indices was investigated, and it was reported that percentile loudness, N_{10}, and Zwicker's loudness were highly correlated with the subjective response to impact ball sounds.[4] Owaki *et al.*[9] recommended L_{Aeq} as a descriptor for bang machine sounds, because L_{Aeq} was more highly correlated with subjective responses than *L*-number and maximum sound level indices such as $L_{\max}$ and $L_{A\max}$. Jeon and Sato[10] used the autocorrelation function (ACF) parameters and sound quality (SQ) metrics to investigate the annoyance of heavy-weight impact sounds. Their results showed that annoyance can be explained by $\Phi(0)$, variances of $\Phi(0)$ and ϕ_1 in terms of ACF parameters, and loudness and fluctuation strengths in terms of SQ metrics. Lee and Jeon[11] classified impact ball sounds into groups A–C according to frequency characteristics, then investigated the subjective response to each group using SQ metrics.

However, even though the SPL, ACF parameters, and SQ metrics are effective in predicting human perception of heavy-weight impact sounds, spatial parameters extracted from the interaural cross-correlation function (IACF) may also affect the perception of a noise source, because information on location or direction of the sound source, subjective diffuseness, and apparent source width (ASW) can be expressed by the factors extracted from the IACF.[12] Therefore, many researchers have applied spatial parameters for the subjective evaluation of various noises. Sato *et al.*[13] investigated the effects of interaural cross-correlation (IACC) and interaural delay time τ_{IACC} on annoyance of noise stimuli in reference to the SPL. Aircraft noise during landing and takeoff conditions was characterized with the physical factors calculated from ACF and IACF, while railway noise were also analyzed using ACF and IACF parameters.[14,15] In addition, IACF parameters have been utilized to describe the acoustical properties of floor impact sounds. ACF and IACF parameters of floor impact sounds generated by standard impact sources were analyzed, and it was found that the IACF of binaural signals differentiates the early perception of loud-

[a] Author to whom correspondence should be addressed. Electronic mail: jyjeon@hanyang.ac.kr

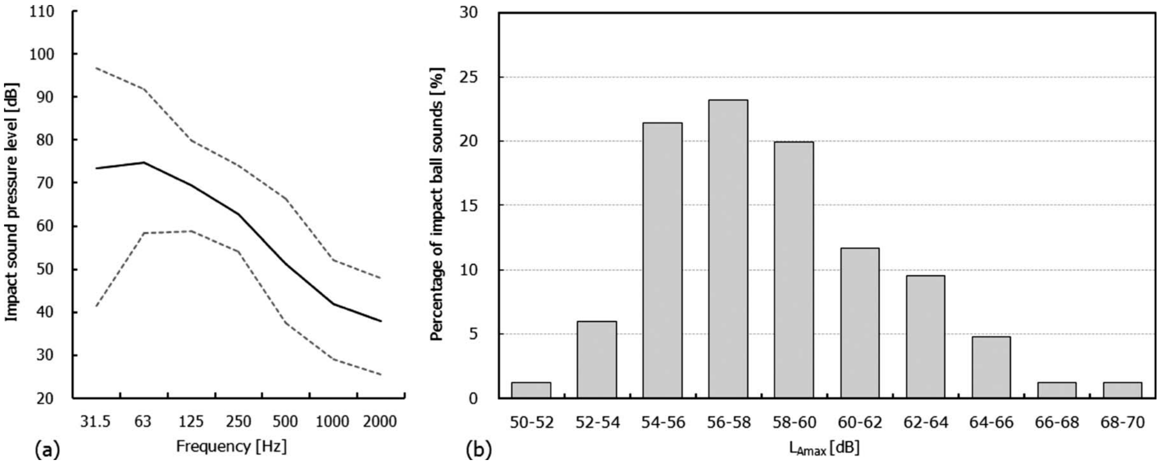


FIG. 1. Measurement result of heavy-weight impact sounds: (a) frequency characteristics [(—) average; (…) minimum and maximum] and (b) distribution of $L_{A\max}$.

ness and noisiness of heavy-weight impact sounds.[16] ACF and IACF parameters were also used to evaluate the similarity between human-made impact sound and floor impact sounds generated by standard impact sources.[17] Subsequently, the relationships between the perception of loudness and these parameters were examined. The results showed that the $\Phi(0)$ and the magnitude of IACC function had a high correlation with loudness of heavy-weight impact sounds. This indicates that loudness of heavy-weight impact sounds can be realized both with the SPL and spatial information of the noise. Sato *et al.*[18] conducted a subjective judgment to evaluate the annoyance of heavy-weight impact sounds in relation to all of temporal factors extracted from the ACF and spatial factors extracted from the IACF. The relationship between annoyance and ACF and IACF parameters showed that important parameters for evaluating annoyance were $\Phi(0)$ and variances of $\Phi(0)$, τ_e, and IACC. However, the effect of IACC on the subjective response to heavy-weight impact sounds needs more accurate investigation.

Several studies have been carried out to determine the just noticeable difference (JND) of IACC. Studies on JND of IACC mostly concern pure tone or narrow band noise in the psychophysics field.[19,20] Gabriel and Colburn[19] found that the JND of IACC in the sound field with such a low IACC is larger than that with a higher IACC. JNDs of IACC were all less than 0.04 at a reference correlation of 1, and JNDs ranged from 0.35 to 0.7 at a reference correlation of 0 in their study. In a room acoustics field, actual sounds (music motif) were applied to determine the JND of IACC.[21,22] Morimoto and Iida[21] reported that the JND values of IACC ranged from 0.03 to 0.12 with changes in the reference sound field. Okano[22] also determined the JND of IACC in terms of 1-IACC$_{E3}$, and found that the JND of 1-IACC$_{E3}$ is measured as 0.065 ± 0.015 with variations in sound field structures. However, no previous study has reported on the JND of IACC for heavy-weight impact noise.

In the present study, JNDs of SPL and IACC for heavy-weight impact sound were investigated based on the measurements of impact ball sounds in real apartments. Standard stimulus and comparison stimuli were randomly presented in pairs to the participants through a stereo dipole system with a subwoofer. JNDs of SPL and IACC were determined when a correct answer was obtained for 75% of responses.[23] Then, annoyance caused by an impact ball was evaluated with changes in SPL and IACC to investigate the effects of IACC on annoyance in reference to the SPL.

The paired comparison method was applied by asking participants which of the two sounds was more annoying. From the results, the contributions of SPL and IACC on the annoyance of impact sounds were obtained.

II. SPL AND IACC OF HEAVY-WEIGHT IMPACT SOUNDS

A. Analysis of SPL for impact ball sounds

It was found that the physical properties of an impact ball (heavy/soft impact source), such as the mechanical impedance and impact force, were more similar to real impact sources than a bang machine (JIS A 1418–2, KS F 2810–2) in a previous study.[4] Therefore, only the impact ball (ISO 140–11, JIS A 1418–2) was used as a standard heavy-weight impact source in this study.

Heavy-weight impact sounds were measured in box-frame, reinforced concrete apartments in Korea. The concrete slab thickness of the apartments ranged from 150 to 180 mm, and the area of most apartments was approximately 100–120 m^2. Eighty-seven heavy-weight impact sounds were recorded binaurally through a head and torso simulator (Brüel & Kjær 4100) when an impact ball was dropped at the center of the source room. The head and torso simulator was also positioned at the center of the receiving room and faced the balcony windows.

Frequency characteristics of heavy-weight impact sounds were calculated by maximum SPL ($L_{\max}$) as shown in Fig. 1(a). The solid line corresponds to the average SPL, and the dashed lines correspond to the range of SPLs. The single number rating value for heavy-weight impact sounds were evaluated in terms of A-weighted maximum SPL ($L_{A\max}$).[7,8] The inverse A-weighted impact SPL ($L_{i,F\max,AW}$) was not considered in this study, because a previous study recommended $L_{A\max}$ as a practical descriptor on the basis of the measurement and calculation procedure. Moreover, $L_{A\max}$

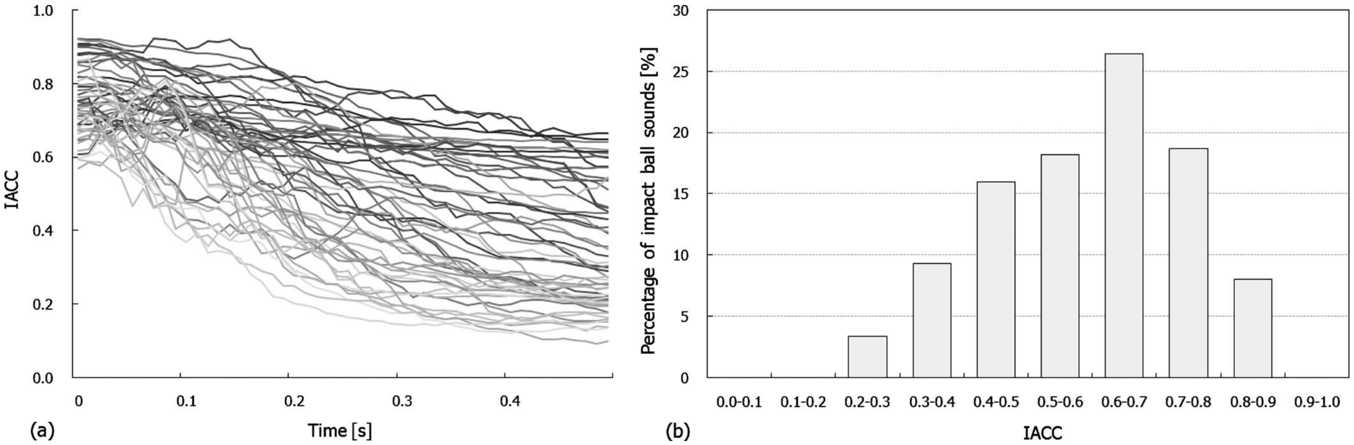


FIG. 2. Analysis results of IACC for heavy-weight impact sounds: (a) the running IACC and (b) distribution of averaged IACC for 0.5 s.

showed a higher correlation with annoyance caused by impact ball sounds than $L_{i,F\max,AW}$.[11] This result is summarized in a bar chart [Fig. 1(b)]. As can be seen, the $L_{A\max}$ of impact ball sounds was normally distributed over the range 50–70 dB.

B. Analysis of IACC for impact ball sounds

The IACF between two sound signals at both ears $f_l(t)$ and $f_r(t)$ is defined by

$$\Phi_{lr}(\tau) = \frac{1}{2T}\int_{-T}^{+T} f'_l(t) f'_r(t+\tau)\,dt, \quad |\tau| \le 1.0 \text{ ms}, \qquad (1)$$

where $f'_l(t)$ and $f'_r(t)$ are obtained after passing through the A-weighted network, which corresponds approximately to the sensitivity of the ear in the low SPL range around 40 dB, $s(t)$, so that $f'_{lr}(t) = f_{lr}(t) * s(t)$.

The normalized IACF is defined by

$$\Phi_{lr}(\tau) = \frac{\Phi_{lr}(\tau)}{\sqrt{\Phi_{ll}(0)\Phi_{rr}(0)}}, \qquad (2)$$

where $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ are the ACFs at $\tau = 0$ for the left and right ears, respectively. The magnitude of the IACC function is defined by

$$\text{IACC} = |\Phi_{lr}(\tau)|_{\max} \qquad (3)$$

for a possible maximum interaural time delay of $|\tau| < 1.0$ ms. The IACC is a significant factor in determining the ASW, degree of subjective diffuseness, and subjective preference for the sound field.[26]

In the calculation of running IACF for heavy-weight impact sounds, the integration interval $2T$ was 0.5 s and the running step was 0.01 s as used in a previous study.[10] The measured values of running IACC for heavy-weight impact sounds are shown in Fig. 2(a), and the average values obtained for the signals in the range 0–5 s were investigated. The IACC peaked initially, and then temporarily decreased with the rapid decrease leveling SPL from the impact ball. The average IACC for the range 0–0.5 s is summarized in a bar chart [Fig. 2(b)]. The average IACC of impact ball sounds normally ranged from 0.2 to 0.9.

III. EXPERIMENT I: JND OF SPL AND IACC

A. Manipulation of SPL and IACC

The sounds for the JND estimations of SPL and IACC were obtained by manipulating the recorded heavy-weight impact sounds produced by an impact ball in real apartments. In the case of SPL, an impact ball sound, which has a dominant SPL at 125 Hz, was used for the experiments because it represents the frequency characteristics of most impact ball sounds;[9] this was also applied to the IACC. These impact ball sounds were considered standard stimuli, and the manipulated sounds for SPL and IACC variations were considered comparison stimuli in this study.

For the JND estimation of SPL, the SPLs of the standard stimuli were manipulated in terms of $L_{A\max}$. The difference of SPL between each of the neighboring comparison stimuli was 0.5 dB in terms of $L_{A\max}$. The maximum SPL difference between the standard and comparison stimuli was 4 dB, whereas the average IACC value for 0.5 s was constant at 0.6, so that the subjective responses were not affected by IACC. Frequency characteristics of standard and comparison stimuli are shown in Fig. 3; the thick line indicates the standard stimulus and the relatively thin lines represent comparison stimuli.

In order to manipulate the IACC in comparison to stimuli, the left channel of the standard stimulus was mixed with the right channel at an interval of 10% using a sound plug-in program (Mix Paste) in ADOBE AUDITION 2.0. The difference of IACC in stimuli comparison was around 0.02–0.03, and the maximum range of the IACC was about 0.43 when the SPL of stimuli was constant at 54 dB in terms of $L_{A\max}$. The variations of other IACF parameters, τ_{IACC} and W_{IACC}, were 0.21 and 0.09 ms so that the effect of IACC alone could be investigated. As shown in Fig. 4, the differences of IACC between stimuli in the initial part are slightly larger than those in the tail part. However, the average values of comparison stimuli were manipulated linearly.

B. Procedure

An auditory experiment was performed to discriminate the difference between standard and comparison stimuli in terms of SPL and IACC. A two-alternative force choice was

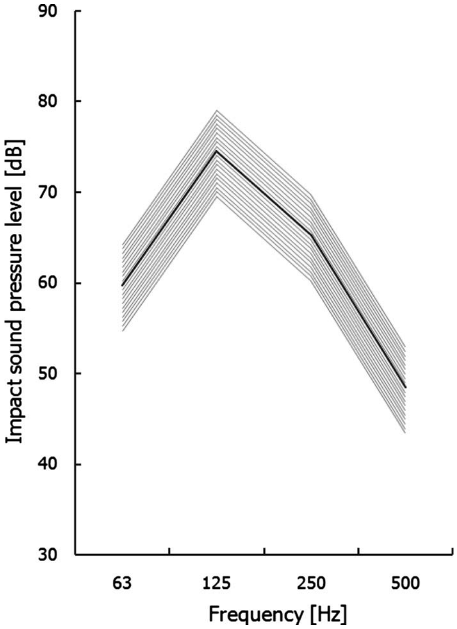


FIG. 3. SPL variations of standard and comparison stimuli for JND tests.

applied for these evaluations.[24] Each of the stimuli pairs consisted of a standard stimulus and a comparison stimulus, which were randomized. The duration of the stimuli, which consisted of two repeated sounds, was about 6 s, and the inter-stimulus interval was set at 1 s. Each pair of stimuli was presented in a random order, separated by an interval of 3 s. Each auditory experiment consisted of two sessions; SPL and IACC values of comparison stimuli were smaller than standard stimulus in the first session, and those were larger than standard stimulus in the second session. The participants were asked to choose the stronger stimulus of each pair, and were asked to mark an X on their response sheet when they were not able to discriminate the difference. The JND value was determined when 75% of the correct answer was obtained from participants.[23]

Twenty participants between the ages of 24 and 35 participated in the experiment. Before the experiment, all participants tested their hearing threshold level with the use of an audiometer (Rion AA-77). The results showed that all participants had normal hearing.

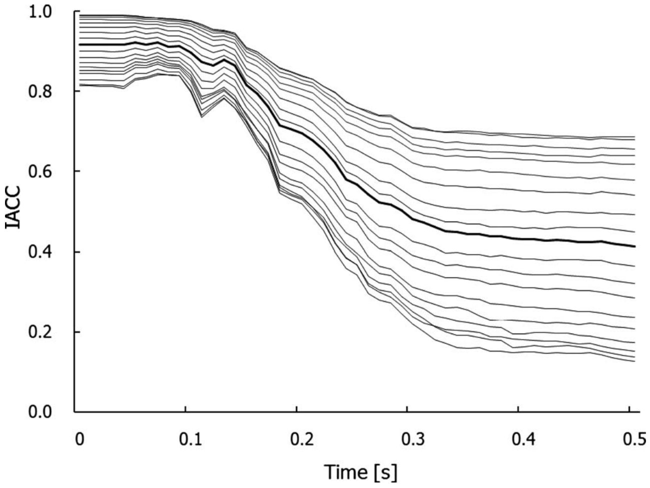


FIG. 4. IACC variations of standard and comparison stimuli for JND tests.

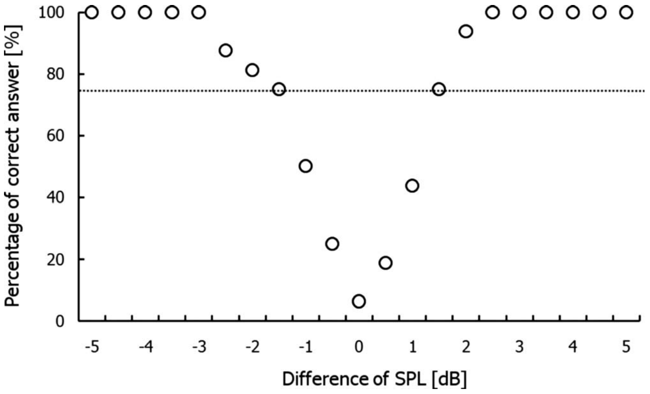


FIG. 5. JND results of SPL.

In this study, stimuli were presented binaurally to participants via a stereo dipole system (Dynaudio BM-10) with a subwoofer (Velodyne DD-10) in order to represent spatial impressions and to reproduce sounds sufficiently at very low frequencies below 63 Hz. In general, two loudspeakers of the stereo dipole system were located in front of the participant to simulate the sound from the stage in room acoustics study.[25] However, loudspeakers were located above the participants to simulate the sound from an upper floor in this study. The SPL of reproduced sound through loudspeakers dropped below 63 Hz compared to the sounds recorded in apartments, whereas the subwoofer reproduced the low frequency sounds well from 15 to 150 Hz. Therefore, a low-pass filter with a cutoff frequency of 63 Hz in the octave band was applied to sounds reproduced by the subwoofer in the same manner as a previous study.[11] ACF/IACF parameters of reproduced stimulus at two-ear entrances of a subject in the listening condition through the stereo dipole system were compared to those of measure stimulus. Differences between ACF/IACF parameters of reproduced and measured stimuli were less than 5% in terms of averaged values for 0.5 s except for τ_e and τ_{IACC}. But the differences of τ_e and τ_{IACC} were 45 and 0.11 ms so that the differences were not distinguishable. The experiments were conducted in a testing booth with approximately 25 dBA of background noise.

C. Results

Figure 5 indicates the JND results of SPL and IACC. The abscissa represents the SPL and IACC difference between the standard and comparison stimuli. The ordinate indicates the percentage of correct answers, P(C) [%], in discriminating the stronger stimulus. In the case of SPL, more than 75% of the participants correctly determined the SPL difference when it was approximately 1.5 dB. The response distributions between the left half (SPL of the comparison stimulus is smaller than that of the standard stimulus) and the right half (SPL of the comparison stimulus is larger than that of the standard stimulus) showed a similar tendency. Therefore, the JND of SPL can be estimated as 1.5 dB in terms of $L_{A\max}$, and this result corresponds to approximately 2 dB in

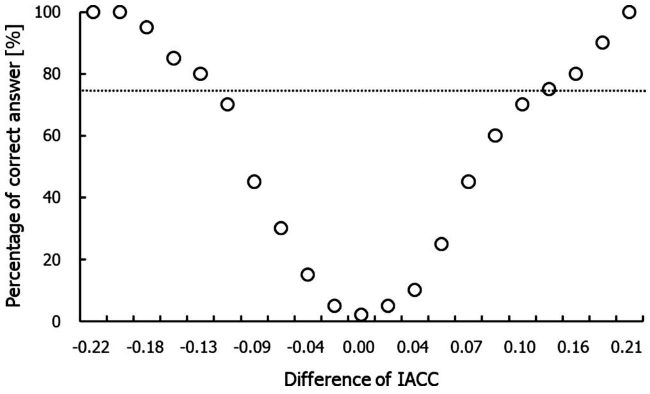


FIG. 6. JND results of IACC.

terms of $L_{iF\max,AW}$. This result is in accord with a previous study, which was conducted in a real apartment and concluded that the JND of the sound generated by the impact ball was about 2 dB ($L_{iF\max,AW}$).[4]

In case of IACC, the response distributions between the left half and the right half were somewhat different, especially asymmetry increased with increased differences between the comparison stimulus and standard stimulus. However, the Analysis of variance (ANOVA) result showed that the response difference between the left half and the right half was not statistically significant. As shown in Fig. 6, the JND of IACC was determined as 0.13 in the left half and 0.12 in the right half with the same procedure. The JND of IACC for the impact ball sound obtained in this study is larger than that of the IACC for the music motif, but smaller than the JNDs of IACC for Gaussian noise at a reference correlation of 0.[19,21]

IV. EXPERIMENT II: ANNOYANCE IN RELATION TO BOTH SPL AND IACC

A. Procedure

The effects of IACC and SPL of heavy-weight impact sounds on annoyance were evaluated by the paired comparison test in this experiment. As shown in Fig. 7, the SPLs were set at 54, 56, and 58 dB ($L_{A\max}$), and measured values of running IACC for stimuli were set at 0.4, 0.6, and 0.8, respectively. The stimuli were also presented binaurally to a participant through the stereo dipole system with a subwoofer.

The paired comparison tests were performed for nine stimuli with changes in SPL and IACC. Twenty participants who participated in the JND test also attended this experiment. The duration of the stimuli, which consisted of three repeated noises, was about 7 s, and the silent interval between the stimuli was 1 s. Each pair of stimuli was presented in random order, separated by an interval of 3 s. The participants were seated in a testing booth and were asked to respond to the following question: "Which stimulus is more annoying if you were exposed to it in the living room?" After the presentation of each stimulus, participants checked the evaluation result on the questionnaire.

B. Results

The scale values of annoyance were calculated from the auditory experiment by applying the law of comparative judgment (Thurstone's case V). The 18 participants who passed the consistency test showed consistent judgments within a 95% confidence interval. This test also indicated that there was significant ($p < 0.05$) agreement among participants. As shown in Fig. 8, louder impact ball sounds were perceived as more annoying, while the IACCs were kept constant. Interestingly, the stimuli with higher IACC was less annoying when the SPL was kept constant. This was because a clear direction from the upper floor could be detected when the stimuli has a higher IACC value. Moreover, clear direction of impact ball sounds was less annoying than diffused sound with a low IACC value. This result is contrary to result in room acoustics studies, which requires subjective diffuseness with low IACC.[26]

The differences of scale values between stimuli with IACC values of 0.4 and 0.6 are larger than those of the scale values between stimuli with IACC values of 0.6 and 0.8. This indicates that the participants are more sensitive to IACC changes when the IACC value is small compared to the stationary sounds.[19,21]

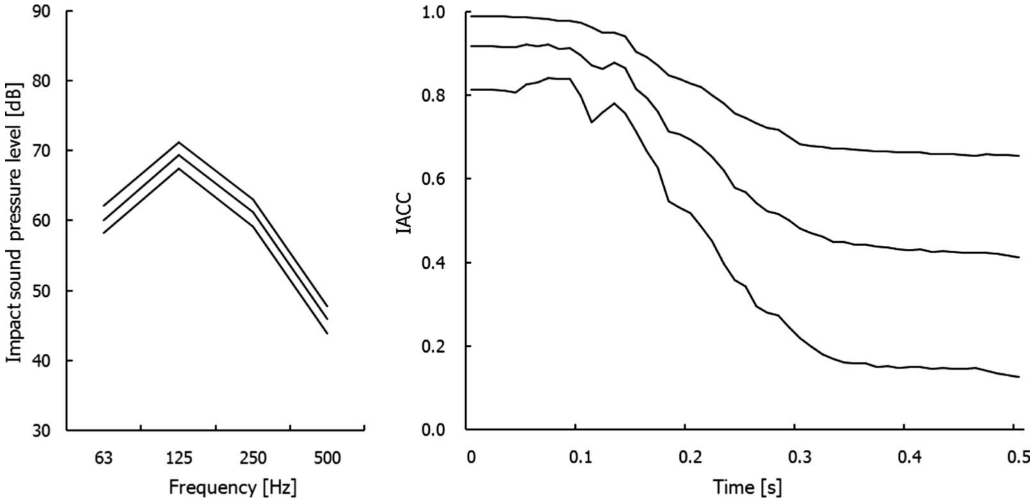


FIG. 7. SPL (left) and IACC (right) of stimuli for annoyance judgment.

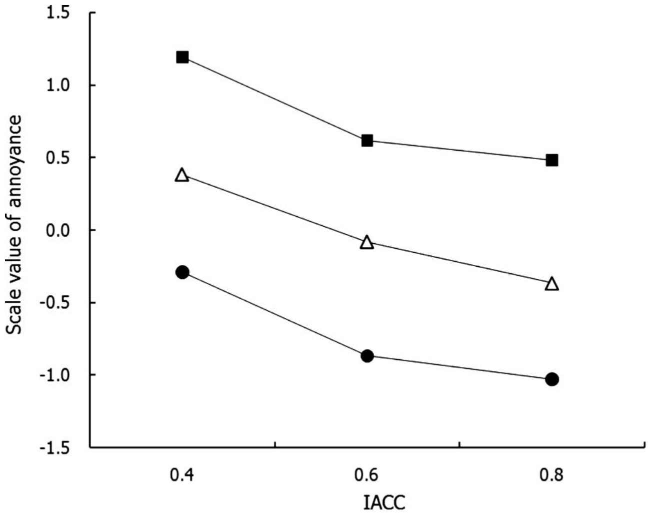


FIG. 8. Average scale values of annoyance as a function of IACC, and as a parameter of SPL: (●) SPL=54 dB, (△) SPL=56 dB, and (■) SPL =58 dB.

The two-way ANOVA for scale values of annoyance was conducted, and the result is listed in Table I. It was found that SPL and IACC are statistically significant ($p < 0.01$), and the effects of the interaction between them were not significant. Thus, SPL and IACC contributed to the scale value of annoyance independently, so the scale value of annoyance may be given by

$$\mathrm{SV}_{\mathrm{annoyance}} \approx f(\mathrm{IACC}) + f(\mathrm{SPL}) \approx a(\mathrm{IACC}) + b(\mathrm{SPL}). \quad (4)$$

The standardized partial regression coefficients of SPL and IACC in Eq. (4) were −0.34 and 0.95, respectively, and these coefficients were statistically significant ($p < 0.01$ for a and b). Using these values, the obtained total coefficient 0.78 was significant ($p < 0.01$). As listed in Table II, the contributions of SPL and IACC to the scale value of annoyance were 79.3% and 20.4%, respectively, and were statistically significant ($p < 0.01$). Although the contribution of SPL to the annoyance of impact ball sounds is larger than that of IACC, the effects of IACC should be considered for the evaluation of annoyance.

In experiment II, it was assumed that other spatial factors extracted from IACF, such as τ_{IACC} and W_{IACC}, were constant to investigate the effect of IACC on annoyance. However, the interaural delay time, τ_{IACC}, of the stimulus was not accurately maintained with the changes in IACC (−0.02, −0.03, and 0.18). Nevertheless, the effect of τ_{IACC} on annoyance was not considered in experiment II because the range of τ_{IACC} was not sufficient to perceive the difference of horizontal sound localization.

TABLE I. Results of two-way ANOVA for scale values of annoyance with the factors SPL and IACC.

Factor	Degree of freedom	Sum of square	Mean square	F-test	p value
SPL	2	16.7	8.3	56.7	<0.01
IACC	2	4.6	2.3	15.5	<0.01
Residual	4	0.03	0.009		

TABLE II. Contributions of IACC and SPL to the scale value of annoyance, and regression correlation coefficients of IACC and SPL.

	Contribution (%)			Regression coefficient	
Subject	IACC	SPL	Total	a	b
Total	20.4	79.3	99.7	−0.34	0.95

V. DISCUSSION

A. Repeatability of IACC

An investigation was made into whether the measurement procedure could ensure the repeatability of IACC for impact ball sounds. In general, repeatability is defined as the variation in the measured sound insulation when measured by the same operator within a short space of time. Hence, measurements of IACC for impact ball sounds were carried out ten times in the same living room of a real apartment by the same operator. Average and 95% confidence intervals are shown in Fig. 9. The 95% confidence intervals of IACC were 0.01 in the initial part, but tended to increase with running time. Moreover, the range of averaged IACC for 0.5 s was 0.06, which was lower than the JND obtained from experiment I. This result shows that the repeatability of IACC for impact ball sounds can be provided in real apartments.

B. Enhancement of satisfaction with change in IACC

The result of experiment II shows that a decrease of 0.3 in IACC is equivalent to an increase of 1.5 dB ($L_{A\max}$) in the SPL. This can be applied to the previous result regarding the subjective response with changes in SPL for heavy-weight impact sounds.[27] In the previous result, the degree of satisfaction with $L_{iF\max,AW}$ was obtained from auditory experiments, and it can be expressed by

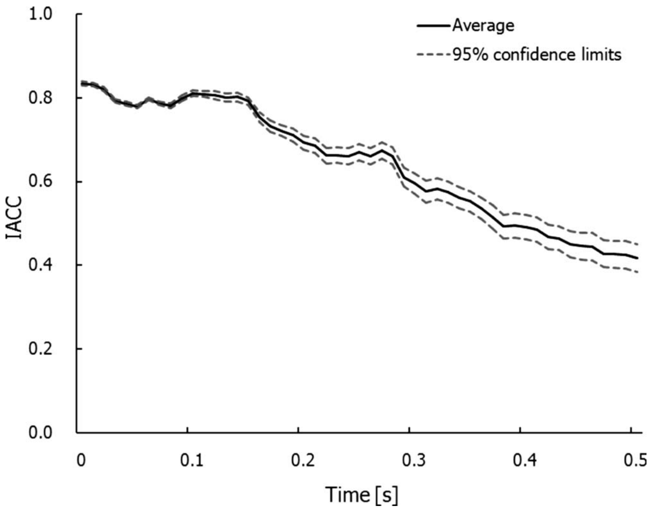


FIG. 9. Average and 95% confidence interval of IACC values.

$$\text{degree of satisfaction}(\%) \approx 6.66\ \text{SPL} - 274.36. \tag{5}$$

Equation (5) indicates that a decrease of 1.5 dB in terms of $L_{iF\max,AW}$ and 1.5 dB in terms of $L_{A\max}$ correspond to an increase of around 10.0% in the degree of satisfaction. Thus an increase of 0.3 in IACC causes an increase of 10.0% in the degree of satisfaction. This indicates that the degree of satisfaction can be controlled by the IACC as well as the SPL for impact ball sounds.

C. Effect of temporal variation of IACC on annoyance

As shown in Fig. 2(a), the temporal variations of IACC are numerous. IACC values of some impact ball sounds rapidly decrease, and others show smooth decay curves. Thus, the effect of the temporal variation of IACC for impact ball sounds on annoyance judgments was investigated through an additional auditory experiment. The temporal variation of IACC (T.var_IACC) for 0.5 s was defined by Eq. (6). Temporal IACC variation of 87 impact ball sounds from real apartments ranged from 0.05 to 1.50.

$$\text{T.var\_IACC} = \frac{\Sigma_{i=i}^{86}(\text{IACC}_{(i)} - \text{IACC}_{(i+1)})/0.01}{86}. \tag{6}$$

The SPL of the stimuli was set at 54, 56, and 58 dB ($L_{A\max}$), and the measured values of running IACC for stimuli were fixed to 0.67. Also, the temporal variation of IACC was set at 0.20, 0.75, and 1.29. The sound reproduction system, participants, and evaluation method were the same as those in experiment II. The scale values of annoyance were obtained, while consistency and agreement tests indicated that all participants showed consistent judgments ($p < 0.05$) and significant ($p < 0.05$) agreement. It was found that SPL and temporal variation of IACC contributed to the scale value of annoyance independently. The contributions of temporal IACC and SPL variation to the scale value of annoyance were 2.7% and 94.2%, respectively, and were statistically significant ($p < 0.01$). The contribution of the temporal variation of IACC to the annoyance is much smaller than that of IACC; thus the temporal variation of IACC does not need to be considered for the evaluation of annoyance.

D. Design factor for IACC of impact ball sounds

1. Sound insulation treatments

Floor impact sounds were measured in box-frame, reinforced concrete apartments with identical floor plans, but different sound insulation treatments. The resilient isolators generally have been inserted between concrete slab and the upper layer of the floor because of their effectiveness in controlling structure-borne and airborne noise. In this study, resilient isolators, which are made of polyethylene foam, were installed in different locations; floor structure, sidewalls, and both sidewalls and ceiling. 20-mm-thick resilient isolator was used in the floor structure, and 7-mm-thick resilient isolator was installed with 7- and 20-mm-thick absorbers in sidewalls and ceiling, respectively. Measurement was also conducted in the plain floor without any sound insulation treatment. As shown in Fig. 10, the IACC of floors with sound insulation treatments was higher than that of plain floor. The average IACC values of floors with the resilient isolator, sidewalls, and both sidewalls and ceiling were 0.69, 0.86, and 0.75, respectively. The plain floor showed a 0.65 averaged IACC value. This indicates that sound insulation treatments, especially in sidewalls, are effective in obtaining higher IACC values because strong reflections from sidewalls were absorbed through sound isolators. Therefore, a clear direction of impact ball sounds can be perceived with sound insulation treatments.

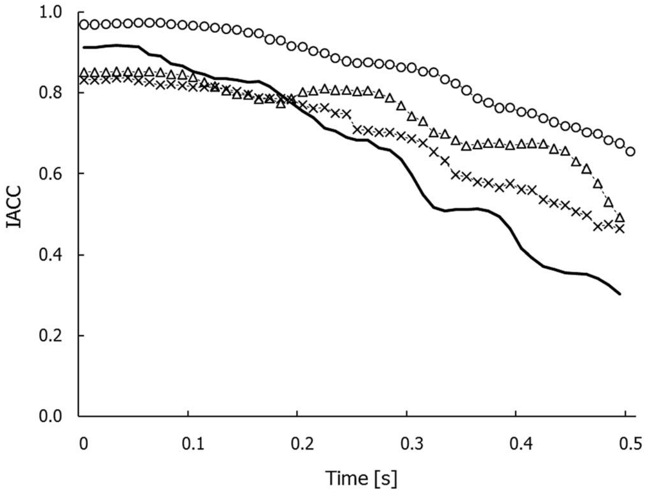


FIG. 10. IACC values for floors with different sound insulation treatments: plain floor (—), sidewalls (○) floor structure (X), and both sidewalls and ceiling (△).

2. Floor structure

Measurements were conducted in order to investigate the effect of different floor structures on the IACC of floor impact sounds in another box-frame, reinforced concrete apartment. A plain floor, a floating floor which contains the resilient isolator between the concrete slab and the upper layer of the floor, and two floors with a constrained damping layer were tested. The damping layers, which contain viscoelastic damping materials, were 3 and 15 mm thick. Figure 11 shows the IACC values of the tested floors. With respect to IACC, the values for the floors with damping material are greater than those with the resilient isolator.

The average IACC in the range 0–0.5 s of the floating floor was 0.32, while that of floors with damping material 3 and 15 mm thick were 0.40 and 0.41, respectively. In contrast, the averaged IACC value of plain floor was the lowest, at 0.24. The result of the IACC analysis indicates that the floors with damping material produced impact sounds with longer IACC values than the floor with a resilient isolator and the plain floor. The reason for this is that concrete slabs and upper layers connected by a damping material act as a single body.[10] Therefore, the impact energies were absorbed in the damping layer, and impact energy transmitted to sidewalls also much decreased.

VI. CONCLUSIONS

Floor impact sounds produced by an impact ball (heavy/soft impact source) were measured in real apartments, and

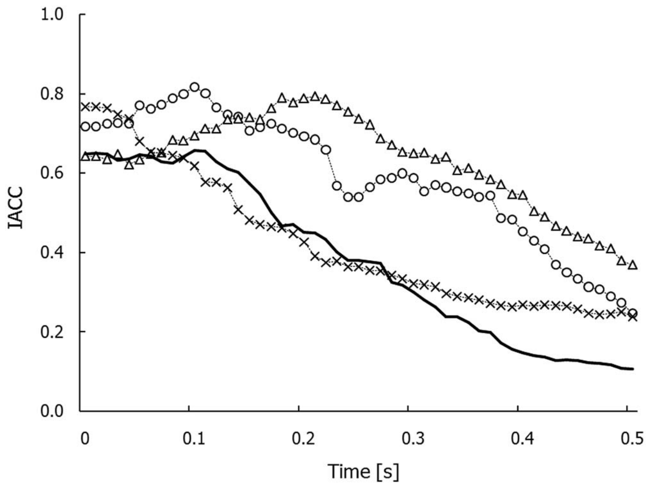


FIG. 11. IACC values for floors with different floor structures: plain floor (—), with resilient isolator (X), with 3-mm-thick damping materials (○), and with 15-mm-thick damping materials (△).

the SPL ($L_{A\max}$) and IACC were analyzed. The auditory experiment for JND estimation of SPL and IACC was performed. The JND value was determined by the criteria of 75% correct answers among all the subjective responses. Then, an auditory experiment was performed in order to investigate the annoyance of impact ball sounds with changes in SPL and IACC.

The conclusions drawn on the basis of the results are shown as follows.

- The JND of SPL was 1.5 dB in terms of $L_{A\max}$, and the JND IACC result was around 0.12–0.13.
- The annoyance increased with decreasing IACC and increasing SPL values.
- IACC and SPL independently contribute to the scale value of annoyance.

These results indicate that the subjective response to the heavy-weight impact sounds can also be improved through the IACC changes. Therefore, spatial features such as IACC should be considered for the subjective evaluation of heavy-weight impact sounds with the temporal features.

ACKNOWLEDGMENTS

This work was supported by Grant No. 10023489 from the Standardized Technology Development Project of the Ministry of Commerce, Industry, and Energy.

[1] JIS A 1418: Acoustics: Measurement of floor impact sound insulation of buildings. Part 2: Method using standard heavy impact sources, Tokyo, Japan, 2000.

[2] KS F 2810: Method for field measurement of floor impact sound insulation. Part 2: Method using standard heavy impact sources, Seoul, Korea, 2001.

[3] ISO 140: Acoustics: Measurement of sound insulation in buildings and of building elements. Part 11: Laboratory measurements of the reduction of transmitted impact sound by floor coverings on lightweight reference floors (International Organization for Standardization, Geneva, 2005).

[4] J. Y. Jeon, J. K. Ryu, J. H. Jeong, and H. Tachibana, "Review of the impact ball in evaluating floor impact sound," Acta. Acust. Acust. **92**, 777–786 (2006).

[5] S. Sato and J. Y. Jeon, "Similarity tests of floor impact sounds in relation to the factors of autocorrelation and interaural cross-correlation functions," Proceedings of the Ninth Western Pacific Acoustics Conference, Seoul, Korea (2006).

[6] ISO Draft 10140–3: Acoustics–Laboratory measurement of sound insulation of building elements. Part 3: Measurement of impact sound insulation (International Standard Organization, Geneva, 2007).

[7] KS F 2863: Rating of floor impact sound insulation for impact source in buildings and of building elements. Part 1: Floor impact sound insulation against standard light impact source (Korean Standards Assn., Seoul, Korea, 2002).

[8] JIS A 1419: Acoustics: Rating of sound insulation in buildings and of building elements. Part 2: Floor impact sound insulation (Japanese Standards Assn., Tokyo, Japan, 2000).

[9] M. Owaki, Y. Yamashita, T. Nagase, and T. Zaima, "A study for the correspondence between the noisiness in the sense of hearing and the light weight and heavy weight floor impact sound with an A-weighted sound pressure level," J. Archit. Plann. Environ. Eng. **537**, 7–12 (2000).

[10] J. Y. Jeon and S. Sato, "Annoyance caused by heavyweight floor impact sounds in relation to the autocorrelation function and sound quality metrics," J. Sound Vib. **311**, 767–785 (2008).

[11] P. J. Lee and J. Y. Jeon, "Psychoacoustical characteristics of impact ball sounds in concrete floors," Proceedings of Noise-Con 2008, Dearborn, MI (2008).

[12] Y. Ando, *Architectural Acoustics-Blending Sound Sources, Sound Fields, and Listeners* (AIP, New York/Springer-Verlag, New York, 1998).

[13] S. Sato, T. Kitamura, and Y. Ando, "Annoyance of noise stimuli in relation to the spatial factors extracted from the interaural cross-correlation function," J. Sound Vib. **277**, 511–521 (2004).

[14] H. Sakai, S. Sato, N. Prodi, and R. Pompoli, "Measurement of regional environmental noise by use of a PC-based system. An application to the noise near the airport "G. Marconi" in Bologna," J. Sound Vib. **241**, 57–68 (2001).

[15] H. Sakai, T. Hotehama, Y. Ando, N. Prodi, and R. Pompoli, "Diagnostic system based on the human auditory brain model for measuring environment noise—An application to railway noise," J. Sound Vib. **250**, 9–21 (2002).

[16] J. Y. Jeon, "Subjective evaluation of floor impact noise based on the model of ACF/IACF," J. Sound Vib. **241**, 147–155 (2001).

[17] J. Y. Jeon, J. H. Jeong, and Y. Ando, "Objective and subjective evaluation of floor impact noise," Journal of Temporal Design in Architectural and the Environment **2**, 20–28 (2002); www.jtdweb.org (Last viewed October, 2008).

[18] S. Sato, J. K. Ryu, and J. Y. Jeon, "Annoyance of floor impact noise in relation to the factors extracted from the autocorrelation and the interaural cross-correlation functions," Proceedings of Forum Acusticum 2005, Budapest, Hungary (2005).

[19] K. J. Gabriel and H. S. Colburn, "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401 (1981).

[20] L. R. Bernstein and C. Trahiotis, "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–315 (1992).

[21] M. Morimoto, and K. Iida, "A practical evaluation method of auditory source width in concert halls," J. Acoust. Soc. Jpn. (E) **16**, 59–69 (1995).

[22] T. Okano, "Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections," J. Acoust. Soc. Am. **111**, 217–229 (2002).

[23] H. Levitt, "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477 (1971).

[24] J. You and J. Y. Jeon, "Just noticeable difference of sound quality metrics of refrigerator noise," Noise Control Eng. J. **56**, 414–424 (2008).

[25] N. Prodi and S. Velecka, "The evaluation of binaural playback systems for virtual sound fields," Appl. Acoust. **64**, 147–161 (2003).

[26] Y. Ando, "Subjective preference in relation to objective parameters of music sound fields with a single echo," J. Acoust. Soc. Am. **62**, 1436–1441 (1977).

[27] J. K. Ryu, P. J. Lee, and J. Y. Jeon, "A combined rating system and criteria for multiple noises in residential buildings," Proceedings of the 34th International Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brazil (2005).

On the level-dependent attenuation of a perforated device

Lan Chen,[a)] Jinqiu Sang, and Xiaodong Li
Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190-2712, China

(Received 7 December 2007; revised 27 January 2009; accepted 17 February 2009)

To investigate the physical principle governing the level-dependent attenuation of a perforated earplug, a mathematical model is first established with the transfer-matrix method to calculate the noise reduction through a simplified device, one perforated panel with back cavity, mounted in an impedance tube. The model prediction is compared with the measured noise reduction through two series of large-scale devices and one device with the dimensions of the ear canal under continuous noise and sinusoidal excitations. The model helps to improve significantly the level-dependent attenuation of the large-scale device. It also illustrates that the attenuation is not solely determined by the resistance of the orifice, which has been a well accepted design concept, but resulted from an incorporated effect of the acoustic filter comprised of the acoustic impedance of the orifice and other elements in the earplug-ear-canal system. This mechanism can interpret a resonance at low incident levels on improper design and reveal approaches to eliminate it. Finally, the model's potential contributions to the design of a perforated earplug are discussed, along with the threshold of level-dependent attenuation supported with experimental evidence.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097494]

PACS number(s): 43.50.Hg, 43.66.Vt [KA] Pages: 2995–3005

I. INTRODUCTION

Impulsive noise generated by weapons can be hazardous to the human ear. Research has shown that permanent threshold shifts, a measure evaluating hearing loss, begins at about 140 dB for rifle impulses and 150–155 dB for canon impulses (Price, 1983). On the other hand, speech communication and the ability to locate and identify other low-level acoustic sources are crucial in the military environment. In the 1970s, the first passive level-dependent earplug, marketed as "Gunfender," was developed by Forrest and Coles (Mosko and Fletcher, 1971; Allen and Berger, 1990; Dancer *et al.*, 1999), which allows the passage of low-level acoustic waves but attenuates high-level sound.

The level-dependent element in such a device is provided by the nonlinear resistance of a sharp-edged orifice, but its quantitative relationship to the earplug's level-dependent attenuation is lacking, so the dimensions of the orifice have been determined mostly by experimentation, and the cost and turn-around time for a new design is unmanageable. At the more preliminary stages of development, the sound transmission characteristics of the orifice are evaluated in a simplified acoustic test fixture with an anechoic termination; in the final analysis, the insertion loss is measured on specialized artificial test fixtures with characteristics of the ear canal (Allen and Berger, 1990). Subjective evaluation tests to acquire the real-ear attenuation at threshold on human subjects are only done under sound pressure levels below the threshold of level-dependent attenuation.

In hope of bringing efficiency for such a design process by providing a convenient computational tool for new designs in the future, this paper investigates a wave-propagation model for a simplified perforated-panel device. Such a device has the essential physical elements of a nonlinear perforated earplug being mounted in the ear canal, with the perforated panel representing the orifice in the earplug and the back cavity representing the space between the earplug and the eardrum. This device is placed in an impedance tube, and the whole system can be taken as a simplified acoustic test fixture. This system is convenient for modeling, manufacturing, and manipulating.

The mathematical model calculates the noise reduction (NR) through the device, which is defined as the difference between the sound pressure level on the front side of the perforated panel and that received at the bottom of the cavity behind the panel. The transfer-matrix method used in developing such a model is also convenient for complex configurations with multiple layers and branches.

Although the majority of the devices studied experimentally have dimensions larger than the average human ear canal, the normalized formulas established by the model are valid in both configurations as long as plane-wave propagation holds. Once the model is proved on the large-scale device, the formulas are applicable to a real earplug within the frequency range investigated [100 Hz to 2.5 kHz; the upper limit is restricted by the fixed microphone locations along the impedance tube used in this study, and plane-wave propagation can be also assumed in the center of the ear canal over this frequency range (Rabbitt and Holmes, 1988)]. As evidence, measured NR on one device with the dimensions of the ear canal is compared with the value predicted by the model.

The model is developed under linear acoustics, and it is also tested in the nonlinear range at high sound pressure levels up to 150 dB. The sound source assumed in the model and applied in the experiments is in the form of either continuous noise or sine signals, instead of impulsive noise, as

[a)] Author to whom correspondence should be addressed. Electronic mail: lan.chen.ls@gmail.com

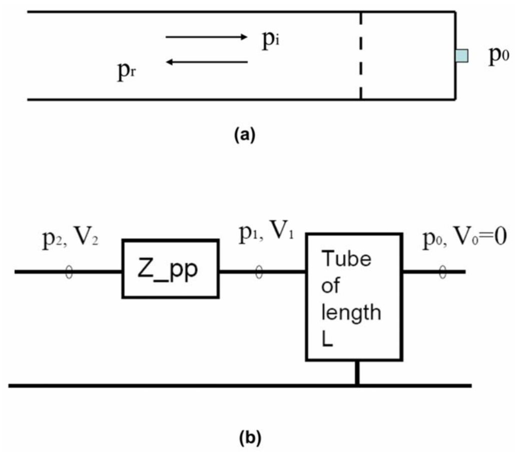


FIG. 1. (Color online) Configuration of the perforated-panel device for modeling and test with its equivalent circuit model of wave propagation in it. (a) Diagram of the perforated panel with back cavity in an impedance tube. NR is defined as the level difference between the incident sound pressure, p_i, and the sound pressure at the "eardrum," p_0. (b) The equivalent circuit model. Sound pressures and acoustic particle velocities for the implementation of the transfer-matrix method are labeled. Z_{pp} represents the acoustic impedance of the perforated panel. The back cavity is a tube with length L (=43 mm), which is a distributed acoustic element in the frequency range under consideration.

this paper addresses the properties of the physical system for the design process, instead of the attenuated signal under protection for evaluation purposes.

II. MODEL DESCRIPTION

Figure 1(a) shows the diagram of the simplified acoustic test fixture, a perforated panel with a back cavity of depth L being mounted in an impedance tube. Plane-wave propagation can be assumed in front of the panel, and the sound pressure is decomposed into an incident part, p_i, and a reflected part, p_r. Figure 1(b) shows the equivalent circuit model for plane-wave propagation in such a configuration, in which the acoustic impedance of the perforated panel, Z_{pp}, is defined as the ratio of the sound pressure on the panel surface to the volume velocity passing through it.

Transfer matrix (Munjal, 1987, pp. 75–83) is then used to relate the sound pressure and the volume velocity through the faces before and after an acoustic element. The sound pressure and the volume velocity on the front face of the perforated panel are represented as (p_2, V_2), and those on the back surface as (p_1, V_1). The sound pressure at the bottom of the cavity is p_0, and the volume velocity V_0 there is zero.

The transfer matrix across the perforated panel is

$$\begin{bmatrix} p_2 \\ V_2 \end{bmatrix} = \begin{bmatrix} 1 & Z_{\mathrm{pp}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ V_1 \end{bmatrix}, \tag{1}$$

and the transfer matrix across the back cavity, which is a tube of length L, is

$$\begin{bmatrix} p_1 \\ V_1 \end{bmatrix} = \begin{bmatrix} \cos(kL) & j\left(\dfrac{\rho c_0}{A}\right)\sin(kL) \\ j\left(\dfrac{A}{\rho c_0}\right)\sin(kL) & \cos(kL) \end{bmatrix} \begin{bmatrix} p_0 \\ 0 \end{bmatrix}, \tag{2}$$

where ρ is the density of air, c_0 is the speed of sound, k is the wave number, and A is the cross-sectional area of the tube.

The origin of the x-axis is defined at the location of the perforated panel, and the sound pressure in the impedance tube is

$$p_2 = p_i e^{jkx} + p_r^{-jkx}. \tag{3}$$

The sound pressure and the acoustic volume velocity on the left side of the panel (at $x=0$) is derived as

$$p_2 = p_i + p_r, \tag{4}$$

$$V_2 = \frac{p_i}{-\rho c_0/A} + \frac{p_r}{\rho c_0/A}. \tag{5}$$

Combining Eqs. (1)–(5), NR, which is defined as the difference in sound pressure level between the left surface of the panel and the bottom of the back cavity, is derived as

$$\mathrm{NR} \equiv 20\log_{10}\left(\left|\frac{p_2}{p_0}\right|\right) = 20\log_{10}(|\cos(kL) + jz_{\mathrm{pp}\_N}\sin(kL)|), \tag{6}$$

with $z_{\mathrm{pp}\_N} = Z_{\mathrm{pp}}A/\rho c_0$, and z_{pp_N} as the normalized (relative to the characteristic impedance of the air, ρc_0) specific acoustic impedance of the perforated panel. The specific impedance is the ratio of the sound pressure to the sound particle velocity.

If the cross-sectional area of the back cavity A_e is different from that of the impedance tube A, then Eq. (2) becomes

$$\begin{bmatrix} p_1 \\ V_1 \end{bmatrix} = \begin{bmatrix} \cos(kL) & j\left(\dfrac{\rho c_0}{A}\right)\sin(kL) \\ j\left(\dfrac{A_e}{\rho c_0}\right)\sin(kL) & \cos(kL) \end{bmatrix} \begin{bmatrix} p_0 \\ 0 \end{bmatrix}, \tag{2$'$}$$

and NR becomes

$$\mathrm{NR} \equiv 20\log_{10}\left(\left|\frac{p_2}{p_0}\right|\right) = 20\log_{10}\left(\left|\cos(kL) + jz_{\mathrm{pp}\_N}\sin(kL)\frac{A_e}{A}\right|\right). \tag{6$'$}$$

In the case of an ear canal, the bottom of the back cavity is not rigid and the impedance Z_m of the termination should be considered. Formula (2) then becomes

$$\begin{bmatrix} p_1 \\ V_1 \end{bmatrix} = \begin{bmatrix} \cos(kL) & j\left(\dfrac{\rho c_0}{A}\right)\sin(kL) \\ j\left(\dfrac{A_e}{\rho c_0}\right)\sin(kL) & \cos(kL) \end{bmatrix} \begin{bmatrix} p_0 \\ V_0 \end{bmatrix}, \tag{2$''$}$$

and p_0 and V_0 are related by $z_m = p_0 A_e/\rho c_0 V_0$, where z_m is the normalized specific impedance of the bottom surface. NR now becomes

TABLE I. Geometric details, including thickness, diameter of the orifices, and number of the orifices, of the two sets of perforated panels tested in the experiments. Their material is listed in the column under Specifications.

Panel series	Panel No.	Thickness (mm)	Diameter of orifices (mm)	Number of orifices	Specifications
I	1	8.5	1	0	Aluminum
	2	8.7	1	1	Aluminum
	3	8.6	1	9	Aluminum
	4	8.5	1	16	Aluminum
	5	5	1	0	Aluminum
	6	5	1	1	Aluminum
	7	5	1	9	Aluminum
	8	5	1	16	Aluminum
II	9	0.5	0.4	9	Copper
	10	0.5		0	Copper
	11	0.1	0.4	9	Copper
	12	0.1		0	Copper
	13	0.05	0.4	9	Copper
	14	0.05	0.4	0	Copper
		0.05	0.4	9	Copper
	15	1	5	9	Supporting panel aluminum
		0.05	0.4	9	Copper
	16	0.5	5	9	Supporting panel aluminum

$$\mathrm{NR} \equiv 20\log_{10}\left(\left|\frac{p_2}{p_0}\right|\right) = 20\log_{10}\left(\left|\cos(kL)\left(1+\frac{z_{\mathrm{pp}}}{z_m}\right) + j\left(z_{\mathrm{pp}\_N}+\frac{1}{z_m}\right)\sin(kL)\frac{A_e}{A}\right|\right). \qquad (6'')$$

Moreover, the transfer-matrix method is convenient for more complex configurations. For example, the transfer matrix for a double-layered perforated-panel device is

$$\begin{bmatrix} p_4 \\ v_4 \end{bmatrix} = \begin{bmatrix} 1 & Z_{\mathrm{pp}1} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \cos(kL_1) & j(\rho c)\sin(kL_1) \\ j(\rho c)\sin(kL_1) & \cos(kL_1) \end{bmatrix} \times \begin{bmatrix} 1 & Z_{\mathrm{pp}2} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \cos(kL_2) & j(\rho c)\sin(kL_2) \\ j(\rho c)\sin(kL_2) & \cos(kL_2) \end{bmatrix} \times \begin{bmatrix} p_0 \\ 0 \end{bmatrix}, \qquad (7)$$

and NR can be derived accordingly (v_4 is the particle velocity in front of the first panel).

As the sound pressure on the left hand side of Eq. (6) can be also taken as the input and output of the perforated-panel device respectively, their ratio gives the frequency response of the physical system.

All the above derivation is done under linear acoustics; when the sound pressure level increases to certain threshold, nonlinearity shows up both in the wave propagation in the impedance tube and in the impedance of the perforated panel. Fortunately, as the sound pressure level being tested is below 150 dB, nonlinearity is insignificant in the wave propagation. The model is tested by comparing the NR calculated with the measured acoustic impedance at high sound pressure levels to the measured NR (details are introduced in Sec. IV).

III. EXPERIMENTAL SETUP

Experiments are arranged to measure the acoustic impedance of two sets of perforated panels and the NR through the perforated-panel devices comprised of these panels and with the same back cavity, whose cross-sectional area is 40 ×40 mm^2 and whose depth is 43 mm. The perforated-panel devices are fixed alternatively at one end of an impedance tube with a cross-sectional area of 40×40 mm^2, in which the cut-off frequency for plane-wave propagation is 4.3 kHz. Therefore, plane-wave propagation can be assumed in the frequency range from 100 Hz to 2.5 kHz under consideration. As stated before, the upper limit of the frequency range is restricted by the distance between microphones 1 and 2 introduced next.

Perforated-panel series I has alumina panels with orifices of 1-mm diameter and of various thickness and perforation ratios (perforation ratio is defined as the ratio of the total area of the orifices to the cross-sectional area of the tube). Series II has copper panels whose dimensions are improved by the model, and they have orifices of 0.4-mm diameter. Panels without orifices are also tested for comparison. Geometry and material details of the panels are listed in Table I.

Diagrams in Fig. 2 show the experimental systems adopted in this study: (a) the impedance-tube setup to measure both the NR and acoustic impedance and (b) the branch-tube setup to measure the acoustic impedance, in which the incident sound pressure has more uniform spatial and frequency distributions [also used by Zhang *et al.* (2006)]. The length of the main tube in the impedance-tube setup is 1.45 m and is 2.70 m in the branch-tube setup.

In the impedance-tube setup shown in Fig. 2(a), three 1/4-in. 40BD G.R.A.S. microphones are used: Microphone 2 is placed 21.87 cm away from the panel; microphone 1 is 4.8

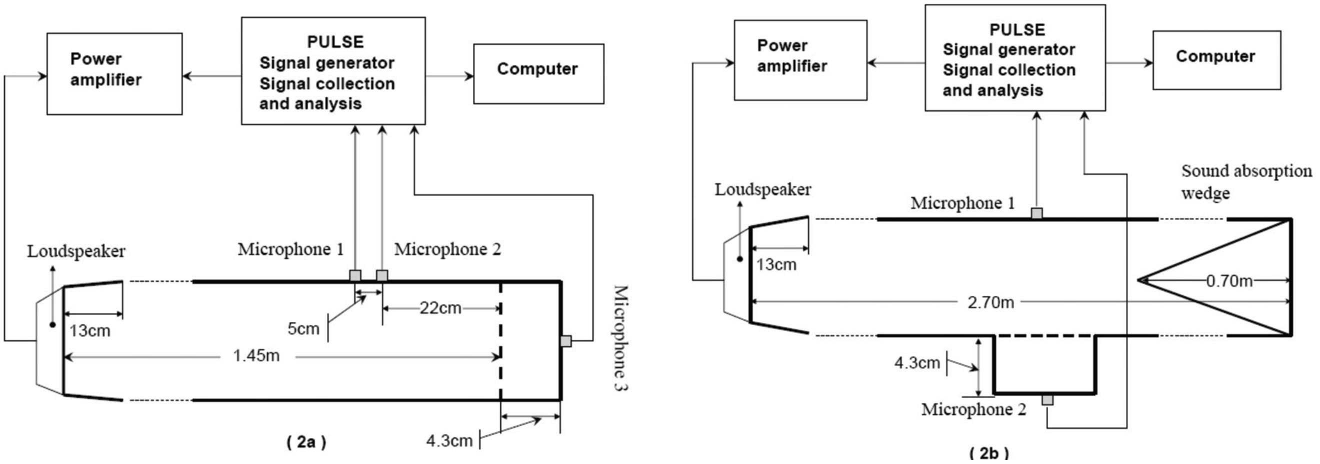


FIG. 2. Configurations of two experimental setups for measuring NR and the acoustic impedance of the perforated panel: (a) the impedance-tube setup and (b) the branch-cavity setup.

cm further away; and microphone 3 is placed at the bottom of the back cavity. In the branch-tube setup in Fig. 2(b), a sound absorption wedge is placed at the end of the main tube, and two microphones are used: Microphone 1 is mounted in the wall right in front of the panel and microphone 2 at the bottom of the back cavity. Signals from microphones are processed with a B&K Pulse multi-analyzer system 3560C, which also generates the white noise and sine signals fed to the high-intensity loudspeaker placed at the other end of the main tube through an OBEY-DPA-500B power amplifier. The loudspeaker mainly used is McCauley 6520 high-frequency compression driver. A laboratory-made high-intensity loudspeaker is also used for better low-frequency (below 500 Hz) performance.

Transfer-function method is applied to determine the normalized specific acoustic impedance of the perforated panel (Chung and Blaser, 1980; ISO 10534-2, 1998) and the sound pressure level on the front surface of the perforated panel. With the transfer function H_{12} of the complex sound pressure from microphone 1 to microphone 2 and the autospectrum S_{11} of the sound pressure recorded by microphone 1, the magnitude of the sound pressure at the front side of the panel is derived as $|p_2|=\sqrt{S_{11}}|e^{jkx_{12}}-H_{12}|/2|\sin(kx_{12})|$, in which k is the wave number and x_{12} is the distance between the two microphones. The difference in level between p_2 and the sound pressure, p_0, recorded by microphone 3 at the bottom of the back cavity gives the NR through the perforated device.

The acoustic impedance of the perforated panel can be also measured in the branch-tube setup according to the method developed by Melling (1973). The impedance equals $(H_{12}-\cos(kd))/j\,\sin(kd)$, where H_{12} is the ratio of the sound pressure on the front side of the panel to the sound pressure at the bottom of the back cavity and d is the depth of the branch cavity.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Measured noise reduction through the two sets of perforated-panel devices

The measured NR through the two sets of perforated-panel devices exhibits different patterns. For series I, increase in attenuation with sound pressure level is only present in a narrow bandwidth of around 200 Hz where a trough locates; moreover, the value of NR in this region is negative, indicating a noise increase in the cavity. For series II, significant level-dependent attenuation is achieved in wideband.

1. Series I

Figure 3 shows the NR through the device with a 8.5-mm-thick panel with 16 orifices of 1-mm diameter under increasing total sound pressure levels from around 90 dB to about 140 dB, which is the typical pattern of the measured NR through series I. Figure 4 compares the measured NR of two panels with different thicknesses (the 8.5-mm-thick and a 5-mm-thick panel with the same number of orifices) under total sound pressure levels at around 90 dB. It demonstrates that larger thickness leads to greater NR in the middle- and high-frequency ranges. This pattern is also present on all panels in this series under all sound pressure levels.

The trough in the NR curve results from a resonance owing to the interaction of the acoustic mass (inertance) in the orifices with the volume of air in the back cavity. Figure

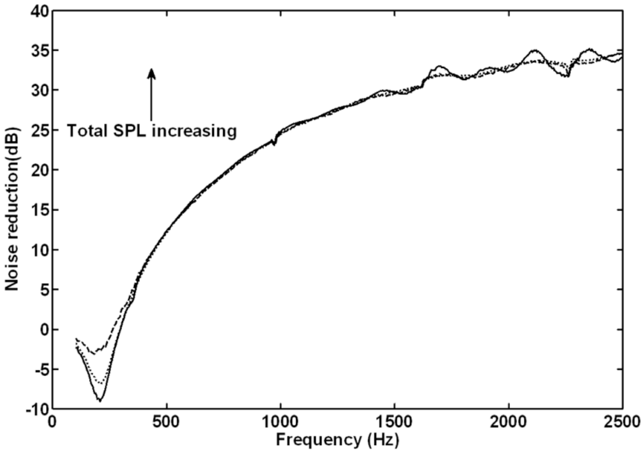


FIG. 3. (Color online) NR measured under different sound pressure levels for the 8.5-mm-thick panel with 16 orifices in series I. Total sound pressure level increase from the curve in solid, then dotted line, and then dashed line.

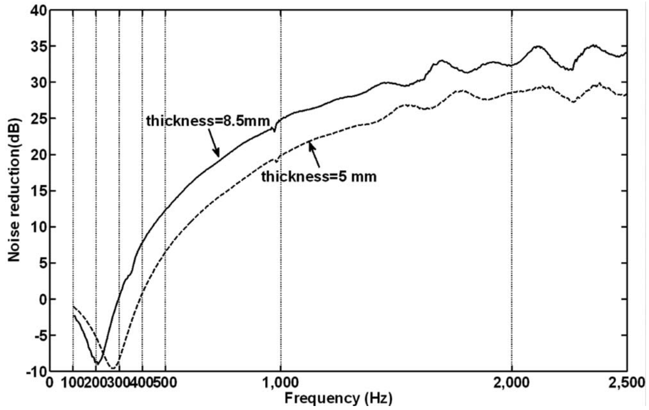


FIG. 4. NR measured under the lowest sound pressure level for the 8.5-mm-thick panel (in solid line) and the 5-mm-thick panel (in dashed line) with 16 orifices in series I.

4 shows that for the 8.5-mm-thick panel, the resonant frequency is at about 200 Hz and at about 300 Hz for the 5-mm-thick panel. In the frequency range below 700 Hz, the back cavity can be lumped as a volume of capacitance c, and the modified equivalent circuit in the low-frequency range is shown in Fig. 5. The resistance and the inertance (normalized, specific) of the perforated panel are represented as r and m, respectively.

The transfer matrix across the cavity now is

$$\begin{bmatrix} p_1 \\ V_1/A \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ j\omega c & 0 \end{bmatrix} \begin{bmatrix} p_0 \\ 0 \end{bmatrix},$$

and the frequency response of the system now becomes

$$\frac{p_2}{p_0} \equiv H(\omega) = (1 - \omega^2 mc) + j\omega cr. \qquad (8)$$

Here, ω is the radian frequency, and when the frequency goes to zero, the magnitude of the transfer function is $1(|H(0)|=1)$, which indicates a NR of 0 dB at zero frequency. This value indicates the low-frequency limit of the devices.

This transfer function has an extreme at frequency ω_0, when $d|H(\omega_0)|/d\omega=0$, and this condition gives

$$\omega_0 = \sqrt{\frac{1}{mc} - \frac{r^2}{2m^2}}. \qquad (9)$$

Physically, this extreme corresponds to the displacement resonance of the mass inside the orifices. At resonance, the magnitude of the oscillating movement of the air inside the orifices reaches the maximum, and the air in the back cavity is also compressed and expanded maximally. The microphone in the back cavity therefore receives the most amplified sound pressure.

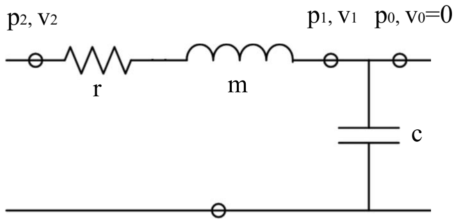


FIG. 5. The low-frequency acoustic circuit model of the perforated-panel device when the back cavity can be lumped as a volume of capacitance c. The resistance and the inertance of the perforated panel are represented as r and m, respectively.

For a perforated panel of thickness t and perforation ratio p and with the back cavity of depth L, m equals t/pc_0 and c equals L/c_0 in the circuit model. From the measured impedance data, we know that the normalized specific resistance r of the 8.5-mm-thick panel is about 2.5 and that of the 5-mm-thick panel is about 2. Insert these values into formula (9), and the resonant frequency is derived as 239 Hz for the device with the 8.5-mm-thick panel and 309 Hz for the one with the 5-mm-thick panel. They are close to the frequency where the trough locates in Figs. 3 and 4, and the discrepancy may result from inaccurate estimations for r, which actually varies with frequency. The errors in obtaining the exact dimension of small orifices may also contribute to the discrepancy.

This trough in NR was also mentioned by Allen and Berger (1990), who found in their device that when orifice diameters exceed 0.012 in. (0.3 mm), the resonance in the frequency range between 500 Hz and 1 kHz can increase the sound level at the eardrum by about 6 dB or more, compared to that of an unprotected ear.

Allen and Berger (1990) further explained that "smaller orifice diameters can increase the attenuation at low sound levels in the frequency range below 1 kHz by providing increased linear resistance." This explanation is in accord with formula (9), which shows that when the resistance r is large enough, ω_0 becomes imaginary and the resonance can be eliminated.

2. Series II

All panels in series II have nine orifices of 0.4-mm diameter, with various thicknesses and in different conditions. The 0.05-mm-thick panel with nine orifices of 0.4-mm diameter is the model-improved configuration. To eliminate panel vibration, a 0.5-mm-thick and a 1-mm-thick supporting panel with nine orifices whose diameters are 5 mm are glued orifice-to-orifice to the 0.05-mm-thick perforated panel, respectively. The additional acoustic mass (normalized, specific) provided by the orifices in the 0.5-mm-thick supporting panel is about 6% of the mass of the orifices in the original panel, and that of the 1-mm-thick supporting panel is about 12%. Both are negligible compared to the acoustic mass of the original panel, but experimental results show that only the 1-mm-thick supporting panel is strong enough to eliminate panel vibration.

This composite panel with a 1-mm-thick supporting panel is the only panel in series II showing no panel vibration, and it demonstrates the best level-dependent attenuation among all the panels being tested. Measured NR in Fig. 6 shows that significant increase in NR is achieved in the whole frequency span from 100 Hz to 2.5 kHz.

Panel vibration brings a trough in the NR curve to all other panels in this series. For the 0.5-mm-thick panel, a trough is present at about 800 Hz and with a bandwidth narrower than the resonance of the panels in series I. Compared to the panel with the same thickness but without ori-

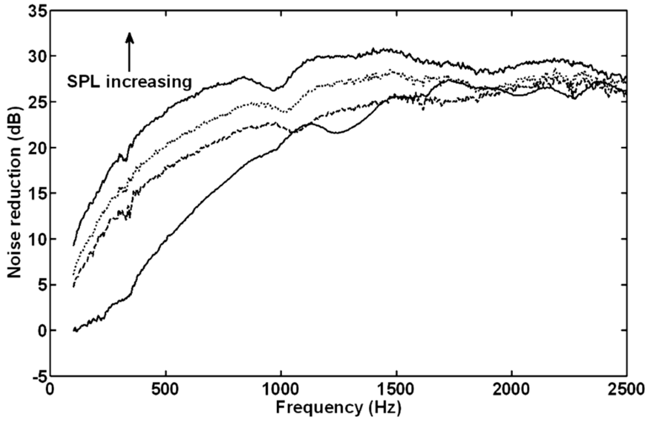


FIG. 6. (Color online) NR measured under different sound pressure levels for the 0.05 mm-thick panel glued to a 1 mm-thick panel in series II. Total sound pressure level increase from the curve in solid, then dotted line, then dashed line, and then solid.

fices, the trough location is about 100 Hz lower, and its bandwidth is wider because the perforated panel has lost mass and gained additional damping owing to the orifices. The trough locates at about 500 Hz for the 0.1-mm-thick perforated panel and at about 350 Hz (with a second trough at 700 Hz) for the 0.05-mm-thick perforated panel.

Analysis of the coupling of panel vibration with the impedance-tube system is not conducted here as it affects the sound attenuation negatively. It may be enough to know that panel vibration should be avoided in such a device.

B. Evaluation of the noise reduction model

In order to examine the validity of the model, the left hand side of formula (6), which is the measured NR, is compared with its right hand side, which is called the NR calculated with the measured acoustic impedance. The second approach calculating the right hand side uses theoretically predicted acoustic impedance by models found in the literature. Only linear impedance models are considered, as this paper addresses applications of the model in the linear range, and it is in the linear range where a resonance needs to be eliminated. In the nonlinear range at high sound pressure levels, this resonance is smoothed out by the large nonlinear resistance. This resonance is believed to hinder useful level-dependent attenuation.

For the NR calculated with measured acoustic impedance, the impedance measured with Melling's approach (Melling, 1973) is found not appropriate, as it is intrinsically the same as the method of calculating the NR. In Melling's method (Melling, 1973), the ratio of the sound pressure on the front side of the perforated panel to the sound pressure at the bottom of the cavity is measured, and its decibel is exactly the NR resorted to in this paper. When the so derived impedance is then used to calculate NR, the calculated NR and the measured value are exactly the same.

The acoustic impedance measured with the transfer-function method is however applicable, but the valid frequency range of the transfer-function method is limited by the pressure nodes presented at the two microphone locations when the surface of the perforated-panel device is highly reflective. At this point, the branch-tube setup is superior because it is free of pressure nodes with an almost traveling-wave field, and the acoustic impedance measured in this setup alone can agree in larger frequency range with theoretically predicted impedances.

For this reason, the measured NR is compared with the NR calculated with the impedance, which is measured with the transfer-function method in the impedance-tube setup, and the impedance measured in the branch-tube is merely compared with the theoretically predicted impedance.

As the NR model is derived under linear acoustics, its validity is also checked at increased sound pressure levels. The NR calculated with the measured impedance has the advantage that the nonlinear acoustic impedances at high sound pressured levels are measurable. For the NR calculated with theoretically predicted impedance, comparison with measurement is limited in the linear range, as this study only considers linear models.

On the other hand, in the design process it is not the acoustic impedance, but the dimensions of the orifices that are the design variables. The acoustic impedance of the perforated panel is an intermediate variable. Therefore, NR calculated with theoretically predicted acoustic impedance is more meaningful as it connects the design variables directly with the design goal—NR.

Consequently, two linear impedance models are first introduced, which give predictions close to the impedance of the two panel series, respectively. Next, comparison is made among the three quantities, measured NR, the NR calculated with the measured impedance, and the NR calculated with linear impedance models.

For the panels in series I, the acoustic impedances measured at low sound pressure levels agree in varied frequency ranges with Crandall's formula (Crandall, 1926, pp. 229–241), according to which the linear specific acoustic impedance of an orifice is estimated as

$$z = \frac{\Delta p}{\bar{u}} = j\omega\rho t\left[1 - \frac{2}{\kappa a}\frac{J_1(\kappa a)}{J_0(\kappa a)}\right]^{-1},$$

where $\bar{u}$ is the average particle velocity in the orifice, t is the thickness of the perforated panel, a is the radius of the orifices, $\kappa^2 = -j\rho\omega/\mu$, and μ is the kinetic viscosity of the air. The non-dimensional parameter $|\kappa|a$, which is $\sqrt{2}$ times the ratio of the radius to the boundary layer thickness in the orifice, measures the importance of viscosity in the orifice. A smaller value of this parameter suggests that viscosity plays a more important role compared to the inertia of the fluid. This formula estimates the specific acoustic impedance of an orifice. To derive the specific acoustic impedance of the perforated panel defined in the NR model, it is divided by the perforation ratio, which is the ratio of the total area of the orifices to the area of the panel.

The calculated (in solid line) and measured (in dashed line; in the impedance-tube setup and in the branch-tube setup) impedances (normalized, specific) of the 8.5-mm-thick panel with 16 orifices of diameter in 1 mm are shown in Fig. 7, which shows that the calculated resistance accords better with measurement than the calculated reactance. Impedance measured with the transfer-function method in the

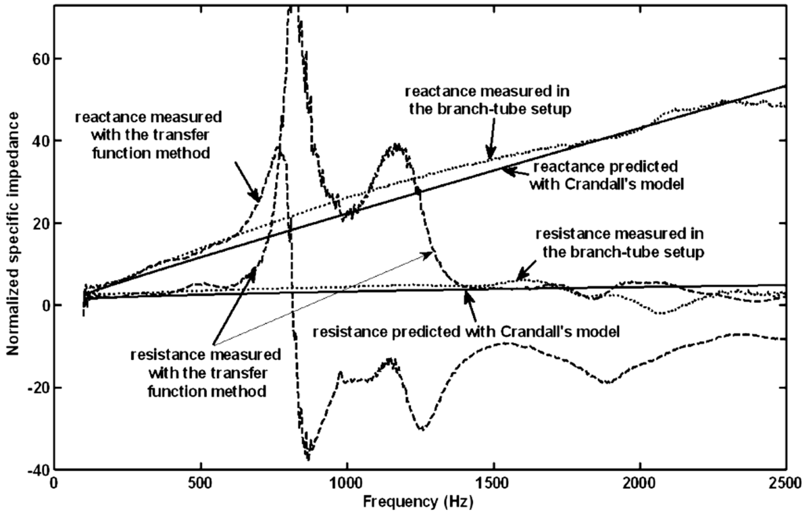


FIG. 7. (Color online) Normalized specific linear acoustic impedance of the 8.5-mm-thick panel with 16 orifices in series I predicted with Crandall's model (in solid line), measured in the impedance-tube setup (in dashed line) and measured in the branch-tube setup (in dotted line) under the lowest testing sound pressure level.

impedance tube is also shown in dotted line, and it agrees with the theoretical value only in the low-frequency range.

For the composite (0.05+1)-mm-thick panel (and it is the only one without panel vibration in series II), the agreement with Crandall's (Crandall, 1926) is poor, but it goes better with Thurston's end correction (Thurston, 1952). Thurston (1952) suggested an end correction for thin panels, on which radiation impedance is more significant, and it is in the form

$$z_e = j\rho\omega\delta\left[1-\frac{2}{\kappa a}\frac{J_1(\kappa a)}{J_0(\kappa a)}\right]^{-1},$$

in which $\delta = 16a/3\pi$. Stinson and Shaw (1985) found that this correction agrees well with their experimental results on perforated panels, which had circular orifices with 0.1–0.3 mm radii and 0.038–0.38 mm thicknesses. This range almost includes the dimensions of the panel being discussed. The measured impedance (in dotted line) of this panel is shown in Fig. 8, along with the impedance calculated with both Crandall's (in dashed line) and Thurston's predictions (in solid line). Thurston's prediction better accords with the measured value than Crandall's.

Next, measured NR and the NR calculated with measured impedance at the lowest and a high sound pressure level for the two panels are shown in Figs. 9–12, respectively. Figures 9 and 10 also include the NR calculated with theoretically predicted acoustic impedance. For the 8.5-mm-thick panel in Fig. 9, the impedance is predicted with Crandall's model; for the composite (0.05+1)-mm-thick panel in Fig. 10, it is predicted with Thurston's end correction. No trough is found either in the predicted or measured NR curves for the composite panel, and the resonance is successfully eliminated.

In general, better agreement between the measured and calculated NR is found on the thicker panel, and a possible explanation is that for the composite (0.05+1)-mm-thick panel, the impedance measured at the lowest testing sound pressure level may not stay in the linear impedance range. Detailed interpretation can be found in Sec. IV C.

For both panels, the calculated NR accords with the measured value better at lower sound pressure levels, and larger discrepancy is found at higher sound pressure levels. This trend is better demonstrated in Fig. 13, where the measured and calculated NR is plotted versus sound pressure level when sine signals are applied.

The larger discrepancy at higher sound pressure level may result from larger errors in the impedance measurement for a more reflective termination in the impedance tube at higher levels, or the model is less accurate when nonlinearity is more significant. The current study cannot determine whether the model is valid or not in the nonlinear range until the impedance can be measured more accurately at higher sound pressure levels.

C. Threshold of level-dependent attenuation

Another phenomenon investigated experimentally is the threshold of level-dependent attenuation, which is also the threshold from linear to nonlinear acoustic impedance of the orifice. It has been reported that when the sound pressure level across the orifice is above certain threshold, the resistance no longer maintains a constant value but increases with sound pressure level, and the reactance just decreases slightly. This variation in the acoustic impedance with inci-

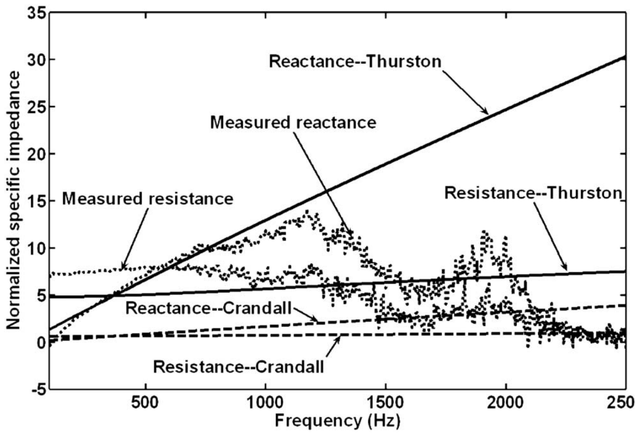


FIG. 8. Normalized specific linear acoustic impedance of the 0.05-mm-thick panel glued to a 1-mm-thick supporting panel measured in the branch-tube setup (in dotted line) and predicted with Crandall's impedance model (in dashed line) and with Thurston's end correction (in solid line).

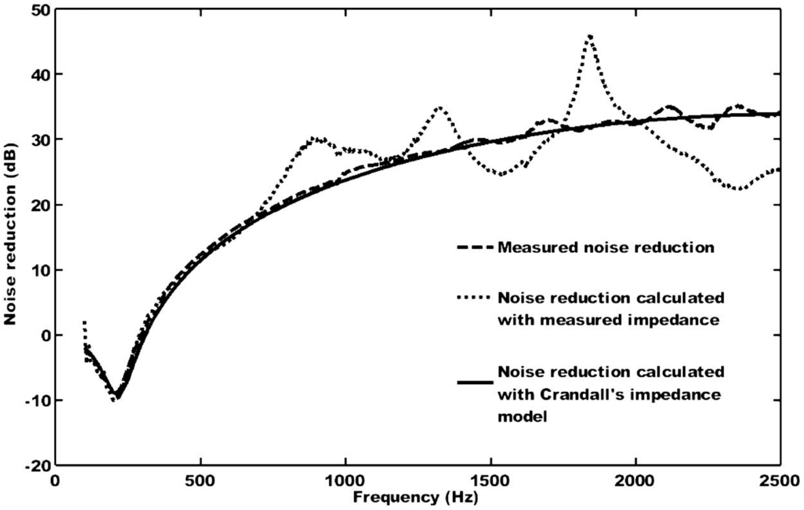


FIG. 9. Measured NR (in dashed line) under the lowest testing sound pressure level for the device with the 8.5-mm-thick panel with 16 orifices in series I, NR calculated with measured impedance (in dotted line), and NR calculated with Crandall's impedance model (in solid line).

dent sound pressure level is called the acoustic nonlinearity of an orifice, and it has been studied since 1935 (Sivian, 1935).

Panton and Goldman (1976) examined the measured data of the nonlinear impedance of thin orifices from several reported studies and correlated them with a few non-dimensional variables using dimensional analysis. For resistance, the correlation is expressed as

$$\frac{R}{\rho(\nu\omega)^{1/2}}=f\left(\frac{t}{d},\frac{u_o}{(\nu\omega)^{1/2}},\frac{d}{\Delta}\right),$$

in which R is the resistance, t is the thickness of the orifice, u_o is the acoustic particle velocity in the orifice, ν is the kinematic coefficient, and Δ is the boundary layer thickness in the orifice as introduced in Crandall's formula. For thin orifices ($t/d\rightarrow 0$), the non-dimensional resistance keeps a constant value until $u_o/(\nu\omega)^{1/2}=3$, then it increases as u_o is further increased. When $u_o/(\nu\omega)^{1/2}>100$, the non-dimensional resistance increases with a slope of +1, and the correlation can be simplified as $R\sim 1.18\rho v$, which is also called the steady-state flow region.

Based on this result, the threshold level of a particular perforated panel in the impedance-tube setup can be estimated. The particle velocity u_o in the orifices of the perforated panel and the particle velocity u_1 in the main tube are related by the law of mass conservation, which suggests that $u_o\sim u_1/p$. Here, p is the perforation ratio of the panel. Take the order of the specific impedance in the main tube as ρc_0; then the threshold level at 200 Hz for a 16-orifice (with 1-mm diameter) panel is estimated as 95 dB, and the threshold level for a 9-orifice (with 0.4-mm diameter) panel is 75 dB [as Panton and Goldman (1976) studied only thin panels satisfying $t/d\rightarrow 0$, it is stricter to only apply it to the composite panel).

To find experimental evidence for this empirical rule, sine signals are applied and NR through the two panels is measured under various sound pressure levels, and the results are plotted in Fig. 13. Sine signals of 200 and 600 Hz are applied to the device with the composite (0.05+1)-mm-thick panel, and the 200-Hz signal is also applied to the device with the 8.5-mm-thick panel. Although the measured NR does not rise exactly at the predicted threshold level, the pattern of the increase in NR with sound pressure level agrees with the prediction qualitatively. Threshold level for the 200-Hz excitation is lower than that for the 600-Hz excitation on the composite (0.05+1)-mm-thick panel, which demonstrates that for the same panel, the increase in NR starts at lower sound pressure

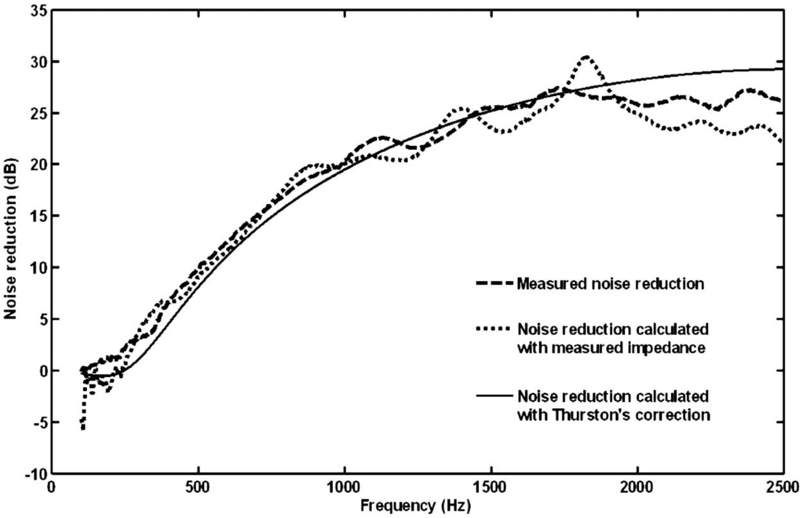


FIG. 10. Measured NR (in dashed line) under the lowest testing sound pressure level for the device with the 0.05-mm-thick panel glued to a 1-mm-thick panel in series II, NR calculated with the measured impedance (in dotted line), and NR calculated with Thurston's end correction (in solid line).

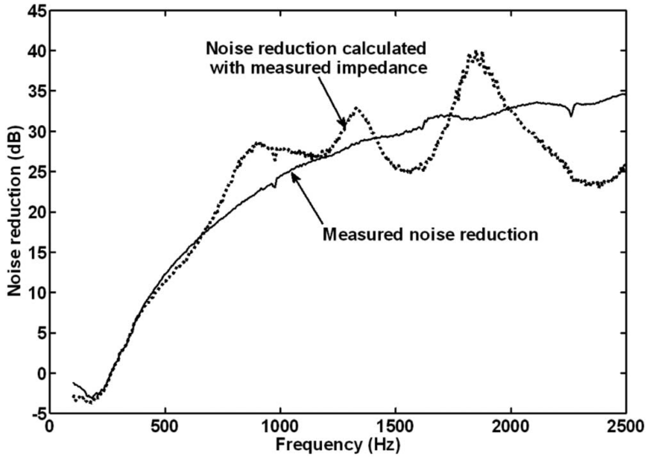


FIG. 11. Measured NR (in solid line) under a higher testing sound pressure level for the device with the 8.5-mm-thick panel with 16 orifices in series I and NR calculated with measured impedance (in dotted line).

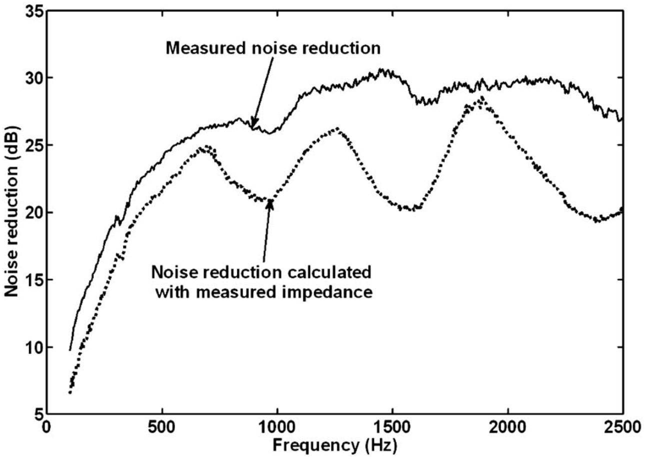


FIG. 12. Measured NR (in solid line) under a higher testing sound pressure level for the device with the 0.05-mm-thick panel glued to a 1-mm-thick panel in series II and NR calculated with the measured impedance (in dotted line).

level for the excitation signal with lower frequency. For the sine signal with the same frequency of 200 Hz, the increase in NR starts at lower sound pressure level for the composite (0.05+1)-mm-thick panel, which has a smaller perforation ratio. The two facts both agree with the empirical rule.

D. Potential contributions to the design of a perforated earplug

As an initial step in quantifying the physical principle governing the level-dependent attention of a perforated earplug, the essential acoustical elements of the earplug being mounted inside an ear canal are modeled with a simplified large-scale perforated-panel device, for which a mathematical model is established to calculate the NR through the device being placed in an impedance tube. In a word, the first step quantifies the attenuation of the perforated device in a simplified acoustic test fixture.

The dimensions of the devices tested extensively in this study are different from those of an average human ear canal, with a cross-sectional area about 30 times that of the ear canal, and the depth of the back cavity is about twice of the length of the ear canal; however, the normalized formula derived with the model is believed to be applicable to the ear canal, as plane-wave propagation can be also assumed in the center of the ear canal over the frequency range from 100 Hz to 2.5 kHz investigated in this paper (Rabbitt and Holmes, 1988).

This argument is further supported by measurements on a device with the dimensions of the ear canal. Figure 14 compares the model prediction (in solid line) given by formula (6′) for a rigid termination and the NR (in dashed line) measured in the impedance-tube setup for a perforated-panel device whose back cavity is 33.3 mm in depth and the diameter of its circular cross-sectional area is 7.36 mm. The thickness of the perforated panel applied is 0.62 mm, and it has a single orifice with a diameter of 0.8 mm. The NR calculated with the measured impedance is plotted in dotted line. The model-predicted NR is very close to the measured NR, and the three curves converge below 500 Hz. As the microphone at the bottom of the back cavity occupies almost the whole bottom surface, the surface of the 1/4 in. 40BD G.R.A.S. microphone is apparently closer to a rigid surface at lower

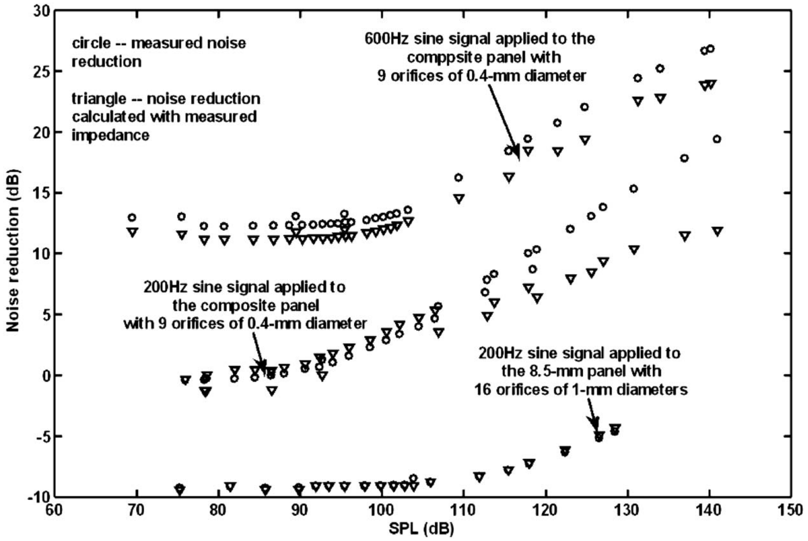


FIG. 13. Measured NR (in circle) and NR calculated with measured impedance (in triangle) for the two panels above, subject to 200 and 600 Hz sine signal excitations and under different sound pressure levels.

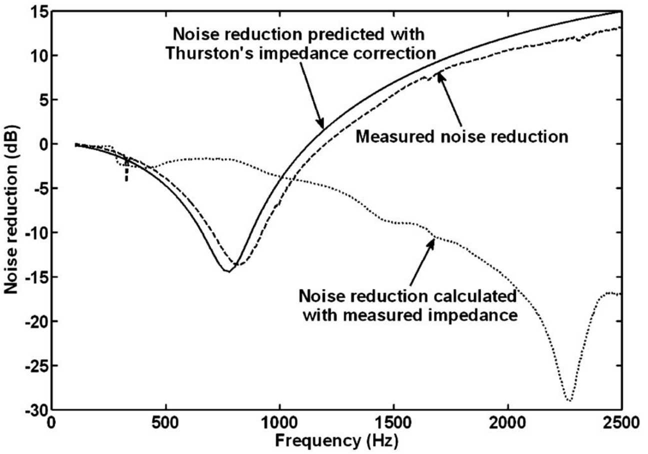


FIG. 14. Measured NR (in dashed line) under the lowest testing sound pressure level for a perforated-panel device with back cavity similar to the average human ear canal, NR calculated with the measured impedance (in dotted line), and NR calculated formula (6′) and with Thurston's end correction (in solid line).

frequency. Were the surface impedance of the microphone known, then formula (6″) can be applied, and better agreement between measured and predicted NR is expected.

In general, the model can predict the NR of a perforated device in the linear acoustic impedance range, with the aid of Crandall's impedance formula for the impedance of a thick orifice or Thurston's end correction for a thin orifice. The model can interpret a resonance on improper designs, which is demonstrated by the experimental results from the devices with perforated-panel series I.

The model also supplies design guidelines to eliminate the resonance. Formula (9) suggests that the resonant radian frequency ω_0 become imaginary if the geometry satisfies

$$\frac{t}{pL} < \frac{r^2}{2}. \qquad (10)$$

For a fixed back cavity, L is a constant, so both reducing the thickness t and increasing the resistance of the orifice r (by choosing a smaller orifice) are possible options.

Consequently, panels with smaller and thinner orifices are chosen for perforated-panel series II. The specific dimensions are determined in a simple way as the main purpose of the current study is to check the validity of the model, instead of finding the optimal configuration. First, the minimum diameter of the orifice most convenient for our workshop to drill is 0.4 mm. Next, the thinnest panel available is 0.05 mm thick, and 0.1-mm-thick and 0.5-mm-thick panels are also used for comparison. When these dimensions are taken into formula (9), the calculated resonance frequencies are all imaginary. Experimental results prove that resonance is absent in the NR of the composite (0.05+1)-mm-thick panel; moreover, the level-dependent performance is improved significantly: the increase in NR reaches 10–15 dB in the frequency range below 1 kHz from around 90 to 140 dB of total sound pressure level. The trough in the NR of the 0.1-mm-thick and 0.5-mm-thick panels results from panel vibration, not from the resonance.

For a small enough $(|\kappa|a<1)$ orifice, the resistance may be large enough to allow a larger thickness, and formula (10) still holds. The attenuation of the device generally increases with thickness, but the increase in attenuation with sound pressure level varies little. In a high-level impulsive noise environment, reduced attenuation at low levels could create less interference with speech and warning sounds when the high-level sound is not present, but the attenuation in the presence of high-level sound is reduced accordingly. In this situation, the NR model may help to determine the optimal range of thicknesses which can balance the attenuation requirements for high-level and low-level sounds.

The model also shows that the attenuation of the perforated-panel device is a function of the length of the back cavity. For an earplug, this suggests that the attenuation is related to the distance between the earplug and the eardrum. For the same earplug, difference in the length of individual ear canals or different locations of the earplug in the same ear canal could lead to different attenuations at the eardrum. Therefore, for a custom-made earplug, the length of the ear canal may need to be considered as a relevant design parameter.

The threshold level in the impedance-tube setup discussed previously is lower than that of a regular perforated earplug because the flow velocity in the orifices is accelerated to a greater degree owing to the small perforation ratio in the impedance-tube setup. Although this setup is unrealistic for a real earplug, the same mechanism could be referred to in lowering the threshold level of an earplug. For example, further raising the flow velocity in the orifice of the earplug relative to the velocity in the free field may lead to a lower threshold of level-dependent attenuation.

V. CONCLUSION

An equivalent circuit model is developed to calculate the attenuation of a perforated-panel device in a simplified acoustic test fixture, and experimental results on two sets of large-scale devices and one device with the dimensions of the ear canal prove that such a model is feasible, especially in the linear range.

The model suggests that the attenuation of a perforated earplug is not solely determined by the resistance of the orifice (a concept which has long existed as a common design guideline) but also by an incorporated effect of the acoustic filter comprised of the acoustic impedance of the orifice and other elements in the earplug-ear-canal system.

In general, the model provides a tool for efficient design and may be able to inspire the creation of more effective configurations. It is expected that the head and details of the ear canal, such as the impedance of the wall, the eardrum, and the middle ear, can be added to the equivalent circuit model, and the sound pressure received by the ear can be further quantified.

Application of the model is restricted by the unknown dependence of the nonlinear acoustic impedance of the orifices on the sound pressure level and the dimensions of the orifice. In situations such as evaluating the attenuation performance of an earplug against high-level impulsive noise or optimizing the dimensions of the orifice for best level-dependent performance, the ability to predict how the acoustic impedance of an orifice varies with incident sound pressure levels is necessary. For example, if the nonlinear impedance of an orifice could be estimated analytically, then the model may help to identify the optimal dimensions of the orifice. With formula (6), the level-dependent performance, which is defined as the increase in NR with a given amount of change in sound pressure, can be expressed as $d(\mathrm{NR})/dp$. It can be further represented as $d(\mathrm{NR})/dp = d(\mathrm{NR})/dz_{\mathrm{pp}\_N} \times dz_{\mathrm{pp}\_N}/dp$, in which the value of $d(\mathrm{NR})/dz_{\mathrm{pp}\_N}$ is derived with the model, and dz_{pp_N}/dp requires a nonlinear impedance model.

ACKNOWLEDGMENTS

The contribution of Meng Tian and Professor Bilong Liu in initiating the project is greatly appreciated. We would also like to show our sincere thanks to reviewers for their helpful comments and references. At last, we would like to thank Feng Peng and Yao Yin for supportive discussions and experimental assistance.

ISO 10534-2 (**1998**). “Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes. Part 7: Transfer-function method,” International Organization for Standardization, Geneva.

Allen, C. H., and Berger, E. H. (**1990**). “Development of a unique passive hearing protector with level-dependent and flat attenuation characteristics,” Noise Control Eng. J. **34**, 97–105.

Chung, J. Y., and Blaser, D. A. (**1980**). “Transfer function method of measuring in-duct acoustic properties. I. Theory,” J. Acoust. Soc. Am. **68**, 907–913.

Crandall, I. B. (**1926**). *Theory of Vibrating Systems and Sound* (Van Nostrand, New York, pp. 229–241.

Dancer, A. L., Buck, K., Hamery, P. J. F., and Parmentier, G. (**1999**). “Hearing protection in the military environment,” Noise Health **2**, 1–15.

Melling, T. H. (**1973**). “An impedance tube for precision measurement of acoustic impedance and insertion loss at high sound pressure levels,” J. Sound Vib. **28**, 23–54.

Mosko, J. D., and Fletcher, J. L. (**1971**). “Evaluation of the gunfender earplug: Temporary threshold shift and speech intelligibility,” J. Acoust. Soc. Am. **49**, 1732–1733.

Munjal, M. L. (**1987**). *Acoustics of Ducts and Mufflers* (Wiley, New York), pp. 75–83.

Panton, R. L., and Goldman, A. L. (**1976**). “Correlation of nonlinear orifice impedance,” J. Acoust. Soc. Am. **60**, 1390–1396.

Price, G. R. (**1983**). “Relative hazard of weapons impulses,” J. Acoust. Soc. Am. **73**, 556–566.

Rabbitt, R. D., and Holmes, M. H. (**1988**). “Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane,” J. Acoust. Soc. Am. **83**, 1064–1080.

Sivian, L. J. (**1935**). “Acoustic impedance of small orifices,” J. Acoust. Soc. Am. **7**, 94–101.

Stinson, M. R., and Shaw, E. A. G. (**1985**). “Acoustic impedance of small, circular orifices in thin plates,” J. Acoust. Soc. Am. **77**, 2039–2042.

Thurston, G. B. (**1952**). “Periodic fluid flow through circular tubes,” J. Acoust. Soc. Am. **24**, 653–656.

Zhang, G., Liu, B., Xiong, J., and Li, X. (**2006**). “An investigation of the nonlinear characteristics of a perforated plate configuration under broadband noise excitation,” Proceedings of Internoise 2006, pp. 652–661.

Spherical loudspeaker array for local active control of sound

Boaz Rafaely[a)]
Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

(Received 30 January 2008; revised 9 March 2009; accepted 10 March 2009)

Active control of sound has been employed to reduce noise levels around listeners' head using destructive interference from noise-canceling sound sources. Recently, spherical loudspeaker arrays have been studied as multiple-channel sound sources, capable of generating sound fields with high complexity. In this paper, the potential use of a spherical loudspeaker array for local active control of sound is investigated. A theoretical analysis of the primary and secondary sound fields around a spherical sound source reveals that the natural quiet zones for the spherical source have a shell-shape. Using numerical optimization, quiet zones with other shapes are designed, showing potential for quiet zones with extents that are significantly larger than the well-known limit of a tenth of a wavelength for monopole sources. The paper presents several simulation examples showing quiet zones in various configurations.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3110131]

PACS number(s): 43.50.Ki, 43.38.Hz [ADP] Pages: 3006–3017

I. INTRODUCTION

Spherical loudspeakers have been designed and employed as omni-directional sources for room acoustics applications.[1–4] These are typically designed from a matched set of loudspeaker units, mounted around the surface of a sphere, to form the shape of a dodecahedron, for example. Recently, these loudspeakers have been used as multiple-channel sources, by driving each loudspeaker unit individually. These multiple-channel spherical arrays have been proposed for music and sound field synthesis,[5] where numerical optimization has been used to control the source radiation pattern.[6–8] Due to its multiple-channel nature, the spherical loudspeaker array is capable of producing high-order sound fields, in spite of the fact that it is a spatially-compact source. These attractive properties of the spherical loudspeaker array may become useful for active control of sound.

Multiple-channel systems have been previously proposed for active sound control.[9] In one approach, array of individual loudspeakers positioned at various locations in an enclosure, such as an airplane cabin, was used to achieve global sound reduction at the noise frequencies using modal control.[10] In another approach, a multipole source was used to cancel the sound of another source, such as the noise from a transformer, as initially suggested by Kempton,[11] and further studied by Bolton *et al.*[12] The two approaches have been compared by a recent study,[13] concluding that each can be useful depending on the primary source characteristics. This paper proposes and investigates the use of a spherical loudspeaker array for active control of sound, and since this is a compact, high-order source, it can be viewed as a realization of a multipole source as suggested by Kempton[11] for global control.

While the research presented in the papers above aimed at global active control, local active control has also been of great interest. The latter was studied using a single loudspeaker to create a quiet zone that was shown to have an extent of a tenth of a wavelength for a reduction of 10 dB in the noise level.[14] Increase in the extent of the quiet zone was possible when two closely-located loudspeakers were used to cancel both pressure and particle velocity.[15] Following this approach, the spherical loudspeaker array can be viewed as an expansion from the two-loudspeaker compact source to a multiple-loudspeaker compact source for local active control of sound.

The paper offers several contributions compared to previous work, as summarized below.

- An extended model for the spherical loudspeaker array is developed, based on previous work,[16] showing that the source can produce pressure sound fields that can be represented by spherical harmonics up to a finite order related to the number of loudspeaker elements.
- This paper makes use of spherical harmonics,[17] rather than Taylor series expansion,[11] to describe the source and the produced sound field. The spherical harmonics representation is complete due to the orthogonality and completeness properties of the spherical harmonics,[17] unlike the Taylor series expansion.[17,18] Nevertheless, we show that similar to Kempton's work,[11] global noise reduction can be achieved outside the region of the sources. Previous work on active control of sound also used spherical harmonics,[13,19] but for representing the primary sound field rather than a multipole secondary source.
- While the approach proposed by Kempton[11] did not lead to the study of active sound control inside the region of the sources due to the divergence of the Taylor series, in this work such local control is investigated by taking advantage of the spherical harmonics representation, allowing analytical analysis of performance for this case.
- In addition to the problem of canceling the sound pressure field due to a single point source, which is the main concern of the multipole studies,[11,12,20] in this work we con-

[a)] Electronic mail: br@ee.bgu.ac.il. URL: http://www.ee.bgu.ac.il/~br sider primary sound fields composed of a single plane-wave and a composition of plane-waves, representing, for example, enclosed sound fields and sound fields due to one or many sources in the far-field.

To summarize, the aim and contribution of this paper is the investigation of the potential performance of the spherical loudspeaker array for active control of sound, in particular, for local active control. It is shown that the "natural" quiet zone using this source has the shape of a spherical shell. Using numerical optimization, quiet zones with other shapes can be created, showing to have an extent, which far exceeds the tenth of a wavelength limit. The paper derives expressions for the primary and secondary sound fields around the spherical source using spherical harmonics analysis, and presents formulations for the optimal secondary source, concluding with numerical investigation using computer simulations.

II. SOUND FIELD REPRESENTATION BY SPHERICAL HARMONICS

Spherical harmonics decomposition is employed in this paper to describe the primary (existing) sound field and the secondary sound field produced by the spherical loudspeaker array. Therefore, spherical harmonics decomposition, based on the work by Williams,[17] is briefly introduced. The standard spherical (r,θ,ϕ) coordinate system[21] is used throughout the paper. Consider a sound pressure function $p(k,r,\theta,\phi)$, defined over the surface of a virtual sphere of radius r, which is square integrable over (θ,ϕ), with k the wavenumber, then its spherical Fourier transform, denoted by $p_{nm}(k,r)$, and the inverse transform, are given by[22]

$$p_{nm}(k,r)=\int_0^{2\pi}\int_0^{\pi} p(k,r,\theta,\phi)Y_n^{m*}(\theta,\phi)\sin\theta d\theta d\phi, \qquad (1)$$

$$p(k,r,\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} p_{nm}(k,r)Y_n^m(\theta,\phi). \qquad (2)$$

The spherical harmonics $Y_n^m(\cdot,\cdot)$ are defined by

$$Y_n^m(\theta,\phi)\equiv\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{im\phi}, \qquad (3)$$

where n is the order of the spherical harmonics, $P_n^m(\cdot)$ is the associated Legendre function, and $i=\sqrt{-1}$.

Note that the sound fields considered in this paper are harmonic over time, such that $p(k,r,\theta,\phi,t)=p(k,r,\theta,\phi)e^{-i\omega t}$, where ω is the angular frequency and t represents time.[17]

Four types of sound fields are considered in this paper. These include sound fields due to a single plane-wave, multiple plane-waves, a point source inside a measurement sphere, and a point source outside a measurement sphere. In addition, the effect of scattering when a rigid sphere is positioned in these sound fields is also formulated.

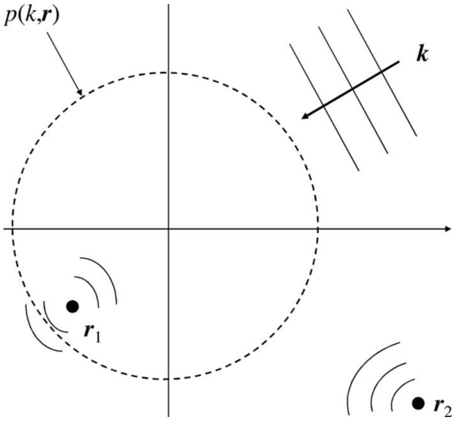


FIG. 1. Illustration of the pressure sound field on the surface of a virtual sphere of radius r, $p(k,\boldsymbol{r})$, with $\boldsymbol{r}=(r,\theta,\phi)$, due to three possible types of sources: (a) a plane-wave of wave-vector $\boldsymbol{k}=(k,\theta_w,\phi_w)$, (b) a point source at $\boldsymbol{r}_1=(r_1,\theta_1,\phi_1)$ with $r_1<r$, and (c) a point source at $\boldsymbol{r}_2=(r_2,\theta_2,\phi_2)$, with $r_2>r$.

A. Sound field due to a single plane-wave

When the sound field $p(k,r,\theta,\phi)$ is produced by a single plane-wave with amplitude $a(k)$, and an incident direction (θ_w,ϕ_w), then the pressure can be written as $p(k,r,\theta,\phi)=a(k)e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, where the wave-vector can be represented in spherical coordinates as $\boldsymbol{k}=(k,\theta_w,\phi_w)$ and the spatial position is given by $\boldsymbol{r}=(r,\theta,\phi)$.[17] See Fig. 1 for an illustration. Equation (1) can be employed to calculate the spherical Fourier transform of $p(k,r,\theta,\phi)$ over (θ,ϕ), with $p_{nm}(k,r)$ in this case given by[17]

$$p_{nm}(k,r)=4\pi i^n a(k)j_n(kr)Y_n^{m*}(\theta_w,\phi_w), \qquad (4)$$

where $j_n(\cdot)$ is the spherical Bessel function.

B. Sound field due to multiple plane-waves

In the case of a more general sound field, a plane-wave decomposition representation can be employed by assuming that the sound field is composed of an infinite number of plane-waves arriving from all directions with spatial amplitude density of $a(k,\theta_w,\phi_w)$. The coefficients p_{nm} can be calculated in this case by taking an integral over all directions (θ_w,ϕ_w) of Eq. (4),[23]

$$\begin{aligned}p_{nm}(k,r)&=\int_0^{2\pi}\int_0^{\pi}4\pi i^n a(k,\theta_w,\phi_w)j_n(kr)Y_n^{m*}(\theta_w,\phi_w)\\&\quad\times\sin\theta_w d\theta_w d\phi_w\\&=4\pi i^n a_{nm}(k)j_n(kr),\end{aligned} \qquad (5)$$

with $a_{nm}(k)$ the spherical Fourier transform of $a(k,\theta_w,\phi_w)$.

C. Sound field due to a point source at $r_1<r$

Another sound field considered in this work is that due to a point source. Given a point source at location $\boldsymbol{r}_1=(r_1,\theta_1,\phi_1)$ with amplitude $a(k)$, then the pressure produced by the point source at location $\boldsymbol{r}=(r,\theta,\phi)$ is given by $p(k,r,\theta,\phi)=a(k)e^{ik|\boldsymbol{r}-\boldsymbol{r}_1|}/|\boldsymbol{r}-\boldsymbol{r}_1|$. See Fig. 1 for an illustration of a source located at r_1 inside the measurement sphere of

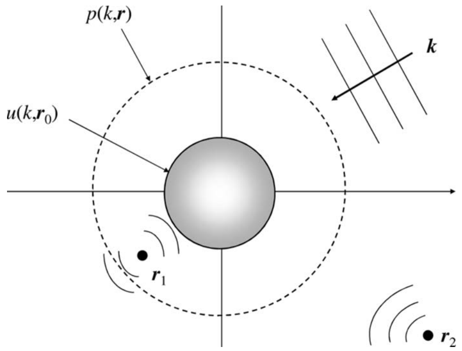


FIG. 2. Same as Fig. 1, but with an additional sphere of radius r_0 centered at the origin, representing either (a) a rigid sphere having zero radial surface velocity, $u(k,\boldsymbol{r}_0)=0$, with $\boldsymbol{r}_0=(r_0,\theta_0,\phi_0)$, or (b) a spherical source with radial surface velocity given by $u(k,\boldsymbol{r}_0)$.

radius r. The spherical Fourier transform of the pressure $p(k,r,\theta,\phi)$ taken over the surface of a sphere defined by (θ,ϕ) is given by[17]

$$p_{nm}(k,r)=4\pi ika(k)j_n(kr_1)h_n(kr)Y_n^{m^*}(\theta_1,\phi_1), \tag{6}$$

where $h_n(\cdot)$ is the spherical Hankel function of the first kind.

D. Sound field due to a point source at $r_2 > r$

In this case the point source is located at $\boldsymbol{r}_2 =(r_2,\theta_2,\phi_2)$ outside the measurement sphere, such that $r_2 > r$; see Fig. 1 for an illustration. The spherical Fourier transform of the pressure for this case is given by[17]

$$p_{nm}(k,r)=4\pi ika(k)j_n(kr)h_n(kr_2)Y_n^{m^*}(\theta_2,\phi_2). \tag{7}$$

E. Sound fields with scattering from a rigid sphere

The four types of sound fields described above due to a single plane-wave, multiple plane-waves, a point source at $r_1 < r$, and a point source at $r_2 > r$, are now studied further when a rigid sphere of radius r_0 is positioned at the origin, as illustrated in Fig. 2. The rigid sphere represents an inactive spherical source, having a zero radial surface velocity on its surface, such that $u=0$ in the figure. The sound field will now include, in addition to the incident field, a contribution due to scattering from the rigid sphere.

For a sound field due to a single plane-wave and multiple plane-waves that include the effect of the rigid sphere, Eqs. (4) and (5) will have a scattering term added. These equations can be rewritten by replacing $j_n(kr)$ with a new term $b_n(k,r,r_0)$, to get

$$p_{nm}(k,r)=4\pi i^n a(k)b_n(k,r,r_0)Y_n^{m^*}(\theta_w,\phi_w) \tag{8}$$

for a single plane-wave, and

$$p_{nm}(k,r)=4\pi i^n a_{nm}(k)b_n(k,r,r_0) \tag{9}$$

for multiple plane-waves. The new term is given by[17]

$$b_n(k,r,r_0)=j_n(kr)-\frac{j_n'(kr_0)}{h_n'(kr_0)}h_n(kr), \tag{10}$$

where $j_n'(\cdot)$ and $h_n'(\cdot)$ represent derivatives. The term with the spherical Bessel function $j_n(kr)$ is due to the incident field, and the term with the spherical Hankel function $h_n(kr)$ is due to the scattered field.

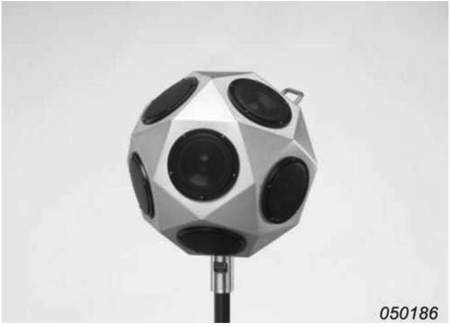


FIG. 3. A spherical loudspeaker array with 12 loudspeaker units, as an example for source S1. [Figure reproduced with permission from Brüel & Kjær.]

For sound fields due to point sources that include scattering from the rigid sphere, Eqs. (7) and (6) can be rewritten with j_n replaced with b_n, yielding

$$p_{nm}(k,r)=4\pi ika(k)b_n(k,r_1,r_0)h_n(kr)Y_n^{m^*}(\theta_1,\phi_1) \tag{11}$$

for the case $r_1 < r$, and

$$p_{nm}(k,r)=4\pi ika(k)b_n(k,r,r_0)h_n(kr_2)Y_n^{m^*}(\theta_2,\phi_2) \tag{12}$$

for the case $r_2 > r$.

III. A MODEL FOR THE SPHERICAL LOUDSPEAKER ARRAY

A model for the spherical loudspeaker array is developed in this section. The aim of the model is to predict the sound field produced by the spherical loudspeaker array, necessary for the active sound control study presented in Secs. IV–VII. Throughout the development, three sources will be considered:

*S*1. An actual spherical loudspeaker array with L loudspeakers arranged around the surface of a sphere. Figure 3 shows an example of such a source with 12 loudspeakers.

*S*2. A rigid sphere covered by L spherical caps each imposing a radial surface velocity of u_l, $l=0,\ldots,L-1$, at the segment on the sphere surface covered by the cap. Figure 4

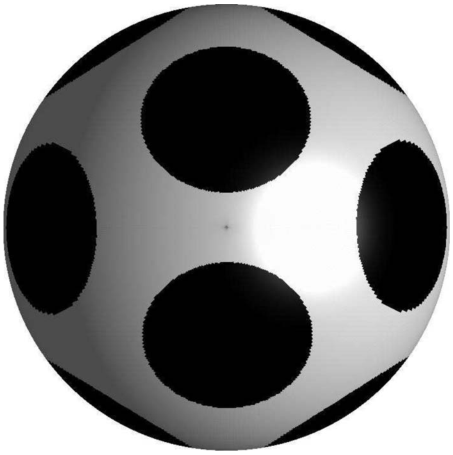


FIG. 4. A spherical source composed of a rigid sphere covered by 12 vibrating caps, as an example for source S2.

shows an example of such a source with 12 caps.

*S*3. A sphere with a continuous radial velocity distribution $u(k,r_0,\theta_0,\phi_0)$ defined over the entire sphere surface.

Source S3 will be used in this paper due to its simplicity and analytical analysis it facilitates. It is argued in this section that under some conditions, source S3 produces sound pressure field similar to source S2, and source S2 similar to source S1, so the use of source S3 in this study may be a useful first step to studying the potential performance of real loudspeaker arrays for active control of sound.

The spherical loudspeaker array, S1, has been shown to produce a sound pressure field similar to that of a rigid sphere covered with vibrating cap elements, as in source S2. The validity of this comparison has been studied by Meyer and Meyer,[24] where good agreement was achieved between measurements from S1 and simulations using S2 for a single cap.

The sound pressure field produced by source S2 has been previously studied theoretically.[16,17] The model facilitating prediction of the sound pressure field from cap velocities is briefly presented here. The spherical Fourier transform $u_{nm}(k,r_0)$ of the velocity on the surface of the sphere, $u(k,r_0,\theta_0,\phi_0)$, due to a single cap centered at the north pole of a rigid sphere of radius r_0, covering the surface defined by $\theta_0 \le \alpha$, is given by[17]

$$\begin{aligned} u_{nm}(k,r_0) &= \int_0^{2\pi}\int_0^{\pi} u(k,r_0,\theta_0,\phi_0) Y_n^{m^*}(\theta_0,\phi_0)\sin\theta_0 d\theta_0 d\phi_0 \\ &= 2\pi\delta_m\sqrt{\frac{2n+1}{4\pi}}u_0(k,r_0)\int_0^{\alpha} P_n(\cos\theta_0)\sin\theta_0 d\theta_0 \\ &= \delta_m\sqrt{\frac{4\pi}{2n+1}}\frac{u_0(k,r_0)}{2}[P_{n-1}(\cos\alpha)-P_{n+1}(\cos\alpha)], \end{aligned} \tag{13}$$

where $P_n(\cdot)$ is the Legendre polynomial, and the velocity on the sphere surface is given by

$$u(k,r_0,\theta_0,\phi_0) = \begin{cases} u_0(k,r_0), & 0 \leqslant \theta_0 \leqslant \alpha \\ 0, & \alpha < \theta_0 \leqslant \pi, \end{cases} \tag{14}$$

such that u_0 represents the cap velocity. The last line in Eq. (13) has been derived using a result from Williams,[17] in which case for $n=0$ the integral is evaluated directly to give $u_{00}=\sqrt{\pi}u_0(k,r_0)(1-\cos\alpha)$. Note that the dependence of u_{nm} on δ_m, the Kronecker delta function, is a result of the azimuth symmetry of the cap velocity.

Following the configuration of source S2, L caps are positioned on the sphere surface, at locations (θ_l,ϕ_l), for $l=0,\ldots,L-1$. The radial velocity u due to all caps can be formulated as a convolution between the velocity due to a single cap located at the north pole, and L delta functions located at each cap center.[16] Using the results that a delta function on the sphere, $\delta(\phi-\phi_l)\delta(\cos\theta-\cos\theta_l)$, transforms into a spherical harmonics $Y_n^{m^*}(\theta_l,\phi_l)$,[17] and that convolution transforms into a normalized product in the spherical harmonics domain,[22,23] we get

$$u_{nm}(k,r_0) = \frac{4\pi^2}{2n+1}[P_{n-1}(\cos\alpha)-P_{n+1}(\cos\alpha)] \times\sum_{l=0}^{L-1} u_l(k,r_0)Y_n^{m^*}(\theta_l,\phi_l), \tag{15}$$

where u_l is the radial velocity of the lth cap. Now, the sound field produced by source S2 having a radial surface velocity u_{nm} is given by[17]

$$p(k,r,\theta,\phi) = i\rho_0 c\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{h_n(kr)}{h_n'(kr_0)}u_{nm}(k,r_0)Y_n^m(\theta,\phi), \tag{16}$$

such that

$$p_{nm}(k,r) = i\rho_0 c\frac{h_n(kr)}{h_n'(kr_0)}u_{nm}(k,r_0). \tag{17}$$

Equations (15) and (16) provide the relation between caps' velocity and the pressure sound field, and form the model for source S2 that can be used to predict the sound pressure field given the radial velocity of each cap. Given a desired pressure sound field, numerical optimization can be used to compute u_l to achieve that desired sound field with minimal error.[16] Note that with L spherical caps, only L spherical harmonics in u_{nm} or p_{nm} can be controlled, defined in the range $0\le n\le N$, $-n\le m\le n$, with N satisfying $(N+1)^2\le L$.

In this paper we take the analytical analysis one step further by relating sources S2 and S3. We recognize the summation in Eq. (15) as a quadrature approximation to Eq. (1),[25] which can be designed to satisfy

$$\sum_{l=0}^{L-1} u_l(k,r_0)Y_n^{m^*}(\theta_l,\phi_l) = \tilde{u}_{nm}(k,r_0), \quad n\leqslant N, \tag{18}$$

such that u_l are samples from the continuous velocity function $\tilde{u}$, which is the inverse spherical Fourier transform of $\tilde{u}_{nm}$, i.e., $u_l(k,r_0)=\tilde{u}(k,r_0,\theta_l,\phi_l)$ for $l=0,\ldots,L-1$. Since we are interested in controlling u_{nm} up to order N, there exist sampling configurations for which $\tilde{u}_{nm}$ can be recovered from u_l with great accuracy. The reader is referred to Refs. 25 and 26 for further reading on sampling and aliasing on the sphere. We can now rewrite Eq. (15) as

$$u_{nm}(k,r_0) = \frac{4\pi^2}{2n+1}[P_{n-1}(\cos\alpha)-P_{n+1}(\cos\alpha)] \times\tilde{u}_{nm}(k,r_0), \quad n\leqslant N. \tag{19}$$

Equations (18) and (19) relate sources S2 and S3, i.e., radial caps' velocities in source S2 and the spherical Fourier transform of the radial surface velocity in source S3. It is therefore clear that for order-limited radial velocity, sources S2 and S3 can be used to produce the same sound pressure field. In the remainder of this paper we therefore use source S3, and assume that the spherical loudspeaker array produces order-limited radial surface velocities. This means that u_{nm}

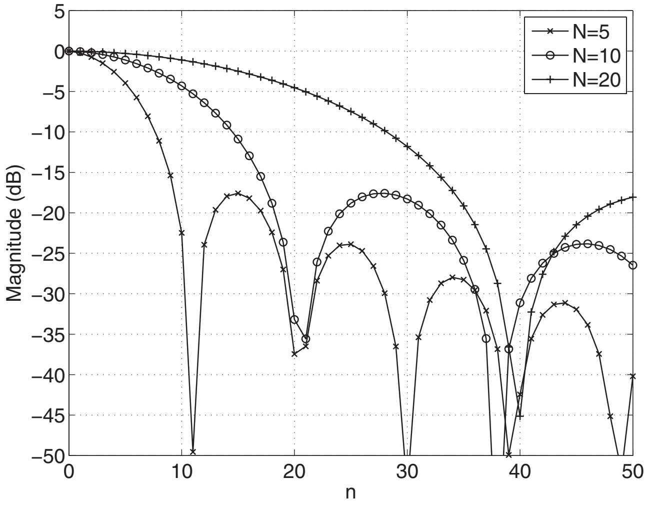


FIG. 5. Magnitude of $g(n,\cos\alpha)$ for three values of N.

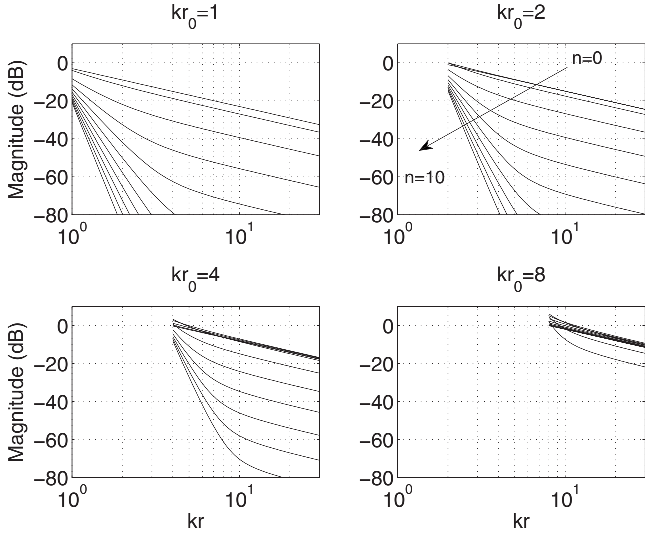


FIG. 6. Magnitude of $h_n(kr)/h_n'(kr_0)$ for $n=0,\ldots,10$ and for four values of kr_0, as a function of kr.

for $n \leq N$ can be used as an input to source S3, and Eq. (16) is used as the analytical model for predicting the sound pressure field produced by the source given u_{nm}.

Few comments are in place.

(i) A real loudspeaker array S1, and source S2 with a finite number of caps, may produce sound pressure fields that can be controlled up to some order N in the spherical harmonics domain. Therefore, no control over the higher-order harmonics of the sound field $n > N$ may be possible. It may therefore be useful to show that the magnitude of these higher-order components can be made negligible. This is discussed below.

(ii) The caps themselves provide some form of attenuation of the high-order harmonics. Consider the term left of the summation in Eq. (15),

$$g(n,\cos\alpha) = \frac{4\pi^2}{2n+1}[P_{n-1}(\cos\alpha) - P_{n+1}(\cos\alpha)]. \qquad (20)$$

A limit for the number of caps that can fit onto the surface of a sphere is given by the ratio of the surface of the sphere, $4\pi r_0^2$, and the surface of a single cap, $2\pi r_0^2(1-\cos\alpha)$,[27] i.e., $2/(1-\cos\alpha)$. On the order hand, if we are interested in harmonics up to order N, which satisfy $(N+1)^2 \leq L$, we can write an approximate relation between N and cap angle α, i.e., $\cos\alpha \approx 1-2/(N+1)^2$. Using this relation, we can investigate g as a function of n for a given order N. Figure 5 shows the magnitude of g, illustrating that the caps provide a low-pass filter along n. However, orders just above N are still significant, and only orders above $2N$ are significantly attenuated. Therefore, the cap filter may not be sufficient to attenuate orders just above N.

(iii) Further reduction in the high-order harmonics will occur when sound is radiated from the sphere, following Eq. (17), due to the term $h_n(kr)/h_n'(kr_0)$. The magnitude of this term for $n=0,\ldots,10$ is illustrated in each sub-figure of Fig. 6, and for four values of kr_0. The figure shows that for $n \leq kr_0$, the magnitude is about constant along n, starting to decrease for $n > kr_0$. This means that a good design choice would be $kr_0 \approx N$. Choosing $kr_0 < N$ means that high orders, such as N and $N-1$, will have small magnitude and so may be difficult to use without introducing excessive noise. Therefore, choosing $kr_0 \approx N$ will ensure that all orders in the range $n \leq N$ can radiate sound efficiently, and also that orders higher than N are attenuated and therefore will not corrupt the desired pressure sound field. Also note that close to the sphere surface, high orders have higher magnitude, which means that smaller spheres can be used to radiate high harmonic orders in the near-field, although at the same time attenuation of orders $n > N$ is also reduced.

(iv) Under the assumption that only orders up to N are controlled in p_{nm}, we can accurately generate only sound fields of finite order N. Now, we know from previous studies that sound fields composed of plane-waves[23] will have insignificant harmonic orders $n > N$ only for $kr < N$. This means that a spherical source of order N can be used to cancel plane-wave sound fields only at a distance from the source that satisfies $kr < N$. This sets a limit to the useful radial range of the spherical source for active control of sound. Combining this limit with $kr_0 \approx N$ introduced in comment (iii) above, it is clear that the useful range satisfies $r \approx r_0$, i.e., generate quiet zones only very close to the source. In practice, quiet zones farther from the source can be produced by choosing $r > r_0$ (with $kr \approx N$ and $kr_0 < N$), but this may come at the expense of increased transducer noise amplification, which in turn may require low-noise transducers.

IV. SOUND FIELD AROUND THE SPHERICAL SOURCE

When the spherical source is placed in an existing primary sound field, the total sound field around the source is composed of the incident primary field, the primary field scattered from the spherical source, and the secondary field radiated from the spherical source. In this section the total sound field is formulated for the various primary sound fields described in Sec. II, only here it is assumed that the sphere is not rigid but possess a given radial surface velocity, u. This is illustrated in Fig. 2 with $u \neq 0$.

A. Primary plane-wave

The incident sound field due to a plane-wave with amplitude $a(k)$ and incident direction $\boldsymbol{k}=(k,\theta_w,\phi_w)$ is given in the spherical harmonics domain by Eq. (4) for free-field propagation, and by Eq. (8) when a rigid sphere of radius r_0 is centered at the origin. When a spherical source of radial surface velocity $u(r_0,\theta_0,\phi_0)$ replaces the rigid sphere, the boundary conditions on the sphere surface change from zero radial velocity for the rigid sphere to the spherical source radial velocity, u. The spherical source radial surface velocity has to balance the velocity contributions from the incident and scattered fields due to the plane-wave, such that

$$u_{\text{inc}}(k,r_0,\theta_0,\phi_0)+u_{\text{scat}}(k,r_0,\theta_0,\phi_0)=u(k,r_0,\theta_0,\phi_0), \quad (21)$$

where u_{inc} is the contribution due to the incident sound field and u_{scat} is the contribution due to the scattered sound field. The velocity boundary condition can be written in terms of sound pressure derivative according to the equation of momentum conservation, or Euler's equation[17]

$$\left.\frac{\partial p_{\text{inc}}(k,r,\theta_0,\phi_0)}{\partial r}\right|_{r=r_0}+\left.\frac{\partial p_{\text{scat}}(k,r,\theta_0,\phi_0)}{\partial r}\right|_{r=r_0}$$
$$=i\rho_0 cku(k,r_0,\theta_0,\phi_0), \quad (22)$$

where p_{inc} and p_{scat} are the incident and scattered pressure fields, given by[17]

$$p_{\text{inc}}(k,r,\theta,\phi)=a(k)\sum_{n=0}^{\infty}\sum_{m=-n}^{n}4\pi i^n j_n(kr)Y_n^{m*}(\theta_w,\phi_w)Y_n^m(\theta,\phi) \quad (23)$$

and

$$p_{\text{scat}}(k,r,\theta,\phi)=a(k)\sum_{n=0}^{\infty}\sum_{m=-n}^{n}c_{nm}h_n(kr)Y_n^{m*}(\theta_w,\phi_w)Y_n^m(\theta,\phi), \quad (24)$$

where c_{nm} are the coefficients of the scattered sound field. Substituting Eqs. (23) and (24) into Eq. (22), and writing the resulting equation in the spherical harmonics domain, we get

$$a(k)Y_n^{m*}(\theta_w,\phi_w)[4\pi i^n j_n'(kr_0)+c_{nm}h_n'(kr_0)]$$
$$=i\rho_0 cu_{nm}(k,r_0). \quad (25)$$

The spherical harmonics coefficients of the scattered sound field can now be written as

$$c_{nm}=\frac{i\rho_0 c}{a(k)Y_n^{m*}(\theta_w,\phi_w)h_n'(kr_0)}u_{nm}(k,r_0)-\frac{4\pi i^n j_n'(kr_0)}{h_n'(kr_0)}, \quad (26)$$

and the total sound field is given by

$$p_{nm}(k,r)=a(k)Y_n^{m*}(\theta_w,\phi_w)[4\pi i^n j_n(kr)+c_{nm}h_n(kr)]. \quad (27)$$

Now, $p(k,r,\theta,\phi)$ can be computed as the spherical Fourier transform of $p_{nm}(k,r)$. Substituting Eq. (26) into Eq. (27), we get after rearranging terms

$$p_{nm}(k,r)=4\pi i^n a(k)Y_n^{m*}(\theta_w,\phi_w)b_n(k,r,r_0)$$
$$+i\rho_0 c\frac{h_n(kr)}{h_n'(kr_0)}u_{nm}(k,r_0). \quad (28)$$

The term on the left is the familiar term for the incident plus scattered sound fields from a rigid sphere [Eq. (8)], while the term on the right denotes radiation from a spherical source [Eq. (17)]. It is therefore clear that the total sound field is a composition of the incident sound field, the scattering from a rigid sphere, and the radiation from the spherical source.

B. Multiple primary plane-waves

In the case of a more general sound field, a plane-wave decomposition representation can be assumed as in Eq. (5) and when applied to Eq. (28) we get

$$p_{nm}(k,r)=4\pi i^n a_{nm}(k)b_n(k,r,r_0)+i\rho_0 c\frac{h_n(kr)}{h_n'(kr_0)}u_{nm}(k,r_0). \quad (29)$$

C. Primary point source at $r_1<r$

Using a similar derivation, the sound field around the source due to a point source located at $r_1<r$ can also be formulated; see Fig. 2 for an illustration. First, c_{nm} in this case can be formulated in a way similar to the derivation in Eqs. (21)–(26), using Eq. (6) for the incident sound field due to a point source instead of Eq. (23). Then, combining the incident fields as in Eq. (6) with the scattered field defined by c_{nm} we get

$$p_{nm}(k,r)=4\pi ika(k)Y_n^{m*}(\theta_1,\phi_1)h_n(kr)b_n(k,r_1,r_0)$$
$$+i\rho_0 c\frac{h_n(kr)}{h_n'(kr_0)}u_{nm}(k,r_0). \quad (30)$$

This equation is similar to Eq. (11), but in addition includes the radiation of the spherical source.

D. Primary point source at $r_2>r$

The derivation for the sound field around a spherical source due to a point source located at $r_2>r$, outside the measurement sphere of radius r, is similar to the case of a point source at $r_1<r$. The difference is that Eq. (7) is used instead of Eq. (6) for the incident sound field due to the point source, to give

$$p_{nm}(k,r)=4\pi ika(k)Y_n^{m*}(\theta_2,\phi_2)h_n(kr_2)b_n(k,r,r_0)$$
$$+i\rho_0 c\frac{h_n(kr)}{h_n'(kr_0)}u_{nm}(k,r_0). \quad (31)$$

V. ACTIVE CONTROL OF SOUND

Active control of sound is realized by superimposing the primary sound field, p_{pri}, with the secondary sound field, p_{sec}, produced by the spherical source. Ideally, the superposition should produce a total sound field of zero pressure, or at least a reduced pressure compared with the primary sound field. The total sound field, p_{tot}, can be written in the spherical harmonics domain as

$$p_{nm_{\text{tot}}}(k,r) = p_{nm_{\text{pri}}}(k,r) + p_{nm_{\text{sec}}}(k,r). \tag{32}$$

Note that the primary field is composed of incident and scattered contributions, as previously presented. In this section active control of sound using the spherical secondary source and for the four primary fields as presented in Sec. IV and illustrated in Fig. 2 is studied analytically.

A. A primary plane-wave

In the case of a primary plane-wave, the total pressure is given by Eq. (28) as follows:

$$p_{nm_{\text{tot}}}(k,r) = 4\pi i^n a(k) Y_n^{m*}(\theta_w,\phi_w) b_n(k,r,r_0) + i\rho_0 c \frac{h_n(kr)}{h_n'(kr_0)} u_{nm}(k,r_0). \tag{33}$$

The total pressure can be set to zero if the following holds:

$$u_{nm}(k,r_0) = -\frac{4\pi i^{n-1} a(k) Y_n^{m*}(\theta_w,\phi_w) b_n(k,r,r_0) h_n'(kr_0)}{\rho_0 c h_n(kr)}. \tag{34}$$

Satisfying the condition that sets the total pressure to zero is possible only for a specific value of r, as the required velocity u_{nm} is now a function of r. This suggests that the natural quiet zone for this configuration is a shell of a chosen radius r. However, there is no guarantee for perfect cancellation outside this shell. This configuration could be useful when a quiet zone is required in the vicinity of the secondary source, both located in a plane-wave primary field, e.g., in the far-field of a noise source.

B. Multiple primary plane-waves

This case is similar to the single plane-wave case, except that here the term $a_{nm}(k)$ representing a distribution of plane-waves is used instead of the term $a(k)Y_n^{m*}(\theta_w,\phi_w)$ for a single plane-wave. The total pressure is given by Eq. (29) as follows:

$$p_{nm_{\text{tot}}}(k,r) = 4\pi i^n a_{nm}(k) b_n(k,r,r_0) + i\rho_0 c \frac{h_n(kr)}{h_n'(kr_0)} u_{nm}(k,r_0). \tag{35}$$

The total pressure can be set to zero if the following will hold:

$$u_{nm}(k,r_0) = -\frac{4\pi i^{n-1} a_{nm}(k) b_n(k,r,r_0) h_n'(kr_0)}{\rho_0 c h_n(kr)}. \tag{36}$$

The resulting quiet zone has a shell-shape in this case as well.

C. Primary monopole at $r_1 < r$

In this case active control of sound is required at radial distances satisfying $r > r_1$, and so Eq. (30) is used for the total primary and secondary sound fields. Rearranging Eq. (30) we get

$$p_{nm_{\text{tot}}}(k,r) = h_n(kr)\left[4\pi ika(k) Y_n^{m*}(\theta_1,\phi_1) b_n(k,r_1,r_0) + \frac{i\rho_0 c}{h_n'(kr_0)} u_{nm}(k,r_0)\right], \tag{37}$$

and the source velocity required to set the total pressure to zero is given by

$$u_{nm}(k,r_0) = -\frac{4\pi ka(k) Y_n^{m*}(\theta_1,\phi_1) b_n(k,r_1,r_0) h_n'(kr_0)}{\rho_0 c}. \tag{38}$$

Selecting this source velocity will produce zero pressure in the range $r > r_1$. This is consistent with the analysis presented by Kempton,[11] although here we used spherical harmonics and a spherical loudspeaker array, rather than Taylor series expansion and a multipole. This configuration is therefore useful when the secondary source can be placed near the primary source with aim of reducing the sound pressure level outside the region of the sources. Unlike the cases of primary plane-waves, in this case active control can be achieved in the entire space outside a sphere of radius r that encircles both the primary point source and the secondary spherical source, therefore achieving global control at this region.

D. Primary monopole at $r_2 > r$

In this case active control of sound is required at radial distances satisfying $r < r_2$. Unlike the work by Kempton,[11] in which the Taylor series did not converge in this range, in this work, due to initial separation of the solution into the two cases, and the use of spherical harmonics, we can find a solution for this case. Here we use Eq. (31) for the total sound field, rewritten here

$$p_{nm_{\text{tot}}}(k,r) = 4\pi ika(k) Y_n^{m*}(\theta_2,\phi_2) h_n(kr_2) b_n(k,r,r_0) + i\rho_0 c \frac{h_n(kr)}{h_n'(kr_0)} u_{nm}(k,r_0), \tag{39}$$

and the source velocity required to set the total pressure to zero is given by

$$u_{nm}(k,r_0) = -\frac{4\pi ka(k) Y_n^{m*}(\theta_2,\phi_2) h_n(kr_2) b_n(k,r,r_0) h_n'(kr_0)}{\rho_0 c h_n(kr)}. \tag{40}$$

Unlike the case $r_1 < r$, in this case satisfying the condition that sets the total pressure to zero is possible only for a given value of r, similar to the plane-wave case. The quiet zone for this configuration is therefore a shell of a chosen radius r, as well.

VI. NUMERICAL DESIGN OF QUIET ZONES

A more general numerical approach can be used to find solutions beyond those presented in the analytical analysis. The spherical Fourier transform of Eq. (32) can be calculated to find the total pressure $p(k,r,\theta,\phi)$. The latter can then be sampled at Q points in space, (r_q,θ_q,ϕ_q), such that

$$p_{\mathrm{tot}}(k,r_q,\theta_q,\phi_q)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} p_{nm_{\mathrm{tot}}}(k,r_q)Y_n^m(\theta_q,\phi_q), \qquad (41)$$

where $p_{nm_{\mathrm{tot}}}$ has been derived for the different primary sound fields considered in Sec. IV, and is given by Eqs. (33), (35), (37), and (39). We next substitute Eq. (33) in Eq. (41), therefore considering the case of a single plane-wave sound field. A derivation for the other cases is straightforward, and will not be presented. Substituting Eq. (33) in Eq. (41), we get

$$\begin{aligned}p_{\mathrm{tot}}(k,r_q,\theta_q,\phi_q)={}&\sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n a(k)Y_n^{m^*}(\theta_w,\phi_w)b_n(k,r_q,r_0)\\&\times Y_n^m(\theta_q,\phi_q)\\&+\sum_{n=0}^{N}\sum_{m=-n}^{n} i\rho_0 c\frac{h_n(kr_q)}{h_n'(kr_0)}u_{nm}(k,r_0)\\&\times Y_n^m(\theta_q,\phi_q). \qquad (42)\end{aligned}$$

Equation (42) can be written in a matrix form as

$$\mathbf{p}=\mathbf{s}+\mathbf{A}\mathbf{u}, \qquad (43)$$

where $\mathbf{p}$ is a $Q\times 1$ vector of the total pressure samples, given by

$$\mathbf{p}=[p_0,p_1,\ldots,p_{Q-1}]^T, \qquad (44)$$

with

$$p_q=p_{\mathrm{tot}}(k,r_q,\theta_q,\phi_q), \qquad (45)$$

and $\mathbf{s}$ is a $Q\times 1$ vector of the primary pressure sound field samples, which also includes scattering from the source under rigid boundary conditions $(u_{nm}=0)$, given by

$$\mathbf{s}=[\ s_0,s_1,\ldots,s_{Q-1})]^T, \qquad (46)$$

where each element is given by the first summation in Eq. (42) as

$$s_q=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n a(k)Y_n^{m^*}(\theta_w,\phi_w)b_n(k,r_q,r_0)Y_n^m(\theta_q,\phi_q). \qquad (47)$$

We note that the second summation in Eq. (42) is of limited order as we have assumed that $u_{nm}=0\,\forall\, n>N$, due to the limited-order source. Vector $\mathbf{u}$ of dimensions $(N+1)^2\times 1$ is therefore given by

$$\mathbf{u}=[u_{00},u_{1(-1)},u_{10},u_{11},\ldots,u_{NN}]^T, \qquad (48)$$

where the explicit dependence of $u_{nm}(k,r_0)$ on k and r_0 has been omitted for notation simplicity. Finally, matrix $\mathbf{A}$ of dimensions $Q\times(N+1)^2$ is defined by the elements A_{qj} given by

$$A_{qj}=i\rho_0 c\frac{h_n(kr_q)}{h_n'(kr_0)}Y_n^m(\theta_q,\phi_q), \quad j=n^2+n+m, \quad n\leqslant N, \quad -n\leqslant m\leqslant n. \qquad (49)$$

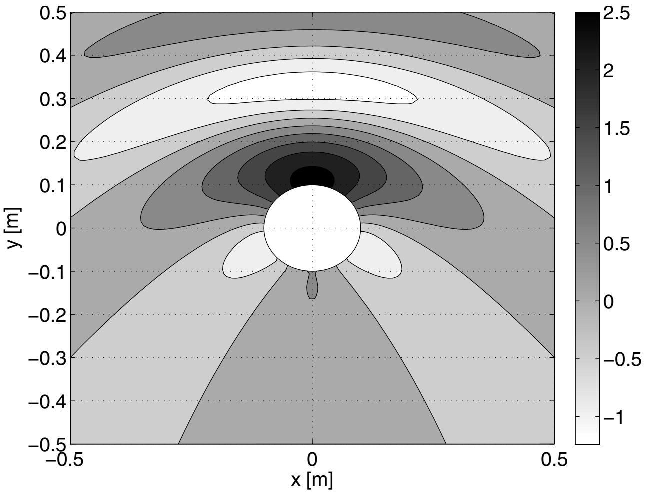


FIG. 7. Magnitude in decibels of a single unit amplitude plane-wave primary sound field with the secondary source switched off. 0 dB refers to the level of the incident pressure field.

Now, a least-squares solution to the source velocity can be computed by setting the total pressure to zero, if possible, or minimizing the 2-norm of the total pressure[28]

$$\mathbf{u}=-\mathbf{A}^{\dagger}\mathbf{s}, \qquad (50)$$

where $\mathbf{A}^{\dagger}$ is the pseudoinverse of $\mathbf{A}$.[28]

VII. SIMULATION EXAMPLES

Several simulation examples are presented in this section to study the quiet zones produced by the spherical source. A spherical source of radius $r_0=0.1$ m is assumed to generate sound by controlling its surface radial velocity. The primary sound field is composed of a unit amplitude harmonic plane-wave, propagating to $(\theta_w,\phi_w)=(90°,270°)$, with frequency 500 Hz. When the source surface is at rest, it is assumed rigid, such that the overall sound field in the vicinity of the sphere is a superposition of the incident plane-wave and the scattered field. Figure 7 shows a contour plot of the incident plus scattered sound field at a cross-section along $\theta=90°$ (the x-y plane), with the spherical source switched off. The figure shows that the variations in the sound field are relatively small, with higher variations close to the source.

A. Plane-wave primary field with monopole secondary source

In the first active control simulation, the spherical source is assumed a monopole source, by limiting its surface veloc-

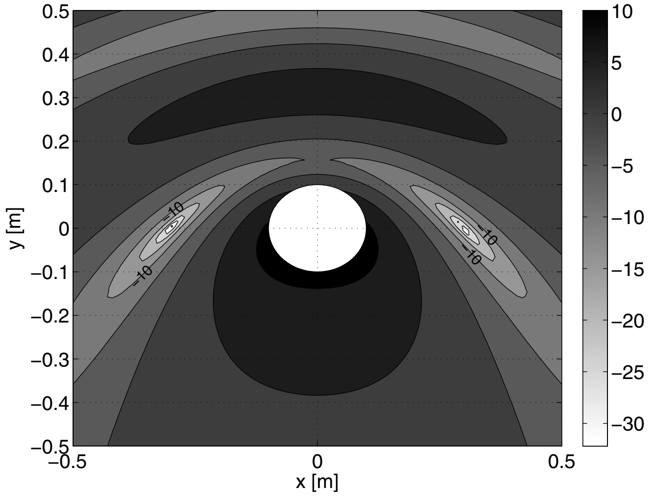


FIG. 8. Magnitude in decibels of the sound attenuation, Eq. (51), with a monopole secondary source designed to cancel the total pressure at $(r,\theta,\phi)=(0.3,90°,0°)$. Primary sound field as in Fig. 7.

ity to order $N=0$. The surface velocity is designed to cancel the pressure at a single point, $(r,\theta,\phi)=(0.3,90°,0°)$, or $(x,y)=(0.3,0)$ in Fig. 8. At a frequency of 500 Hz, a tenth of a wavelength is about 6.8 cm, which is the expected extent of the −10 dB quiet zone with a monopole source.[14] Figure 8 shows the magnitude, in decibels, of sound attenuation, calculated as

$$ATT = 20\,\log_{10}\left(\frac{p_{\text{tot}}}{p_{\text{pri}}}\right). \qquad (51)$$

The figure shows that the width of the −10 dB quiet zone is about a tenth of a wavelength, while its extent is larger along its length due to the symmetry of the sound fields. Due to the same symmetry, another quiet zone is present around $\phi=180°$.

B. Spherical secondary source with shell quiet zone

In the following simulation, the surface velocity of the spherical source was computed to cancel the total sound field at a radius of $r=0.25$ m, where the surface velocity with spherical harmonics up to order $N=6$ was assumed. Figure 9 shows that indeed a significant attenuation is achieved around $r=0.25$, with a larger extent around $\phi=270°$. This is since around $\phi=270°$ the spherical source is located downstream compared to the primary source and so the two sound fields can be more easily matched.[9] Although the downstream attenuation is significant, the extent of the upstream attenuation is around a tenth of a wavelength, similar to the extent when using a monopole.

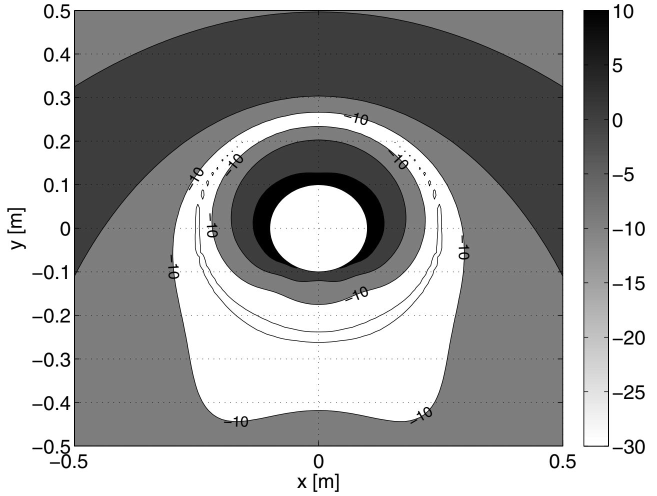


FIG. 9. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to cancel the total pressure at a radius of $r=0.25$. Primary sound field as in Fig. 7.

The next simulation has the aim to create a larger zone of quiet. For this purpose, Eq. (50) is used to compute the surface velocity of the source, by minimizing the total pressure at a set of locations in the range $r\in[0.2,0.3]$, $\theta=90°$ and $\phi\in[0°,360°)$. Attenuation levels are presented in Fig. 10, showing a large quiet zone, much larger than the conventional extent of a tenth of a wavelength achieved using a monopole secondary source. Nevertheless, it should be noted that some sound enhancement is also notably farther from the source. Improved performance of the spherical source compared to a monopole source is explained by the ability of the spherical source to produce complex secondary sound fields that better match the primary field. Indeed, it has been previously shown that improvement in active control performance is achieved by increasing the number of sources from one to two, canceling pressure and pressure gradient at one point,[15] where about doubling in the size of the quiet zone was achieved. In this work, with the use of a multiple-channel source, further improvement is evident. Furthermore, the potential ability of the spherical source to control the pressure sound field in three dimensions suggests a potential for improved flexibility in the shape and orientation of the quiet zone. This is in contrast to the two-source arrangement or a linear source array, where the sound field can be controlled in one or two dimensions only.

C. Spherical secondary source with spatially-confined quiet zone

Although the results presented in Fig. 10 may be useful, the shape of the quiet zone, i.e., spherical shell, may not be

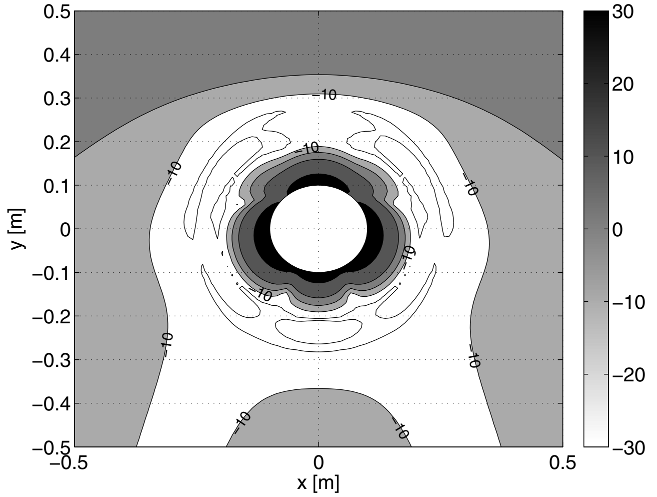


FIG. 10. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to minimize the total pressure at a range of points around $r\in[0.2,0.3]$. Primary sound field as in Fig. 7.

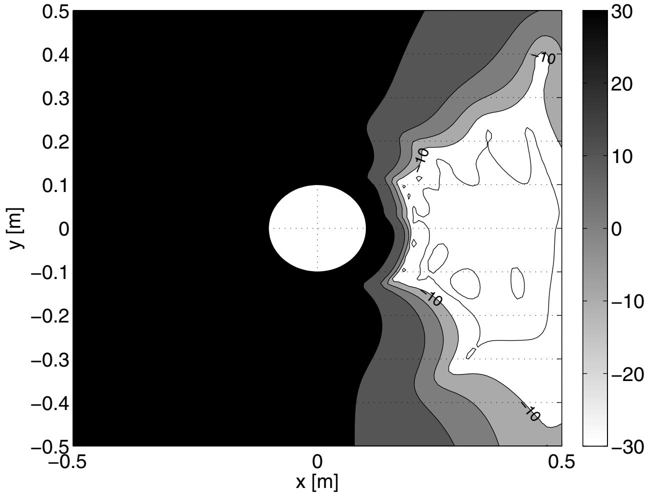


FIG. 11. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to minimize the total pressure at a range of points around $r \in [0.2, 0.5]$ and $\phi \in [-30°, 30°]$. Primary sound field as in Fig. 7.

the desired shape in practice. In many applications, the required quiet zone may be defined over a confined space, in which listeners may occupy. Therefore, in this simulation, Eq. (50) is used for the computation of the source velocity, only here the minimization points are distributed in the range $r \in [0.2, 0.5]$, $\theta = 90°$ and $\phi \in [-30°, 30°]$, defining a space around $(x, y) = (0.35, 0)$. Figure 11 shows the attenuation results. A large quiet zone is achieved, with dimensions of around 30–50 cm, much larger than the conventional-size quiet zones. However, the large quiet zone comes at the expense of significant amplification outside the quiet zone, which may not be tolerated in practice. Therefore, in the next simulation, diagonal regularization, also known as Tikhonov regularization, has been employed in the process of matrix inversion in Eq. (50) to avoid solutions with large magnitude. An example for the regularized solution is presented in Fig. 12. Indeed, the amplification away from the quiet zone has been reduced, in particular, in the region $x < 0$, but so did the extent of the quite zone.

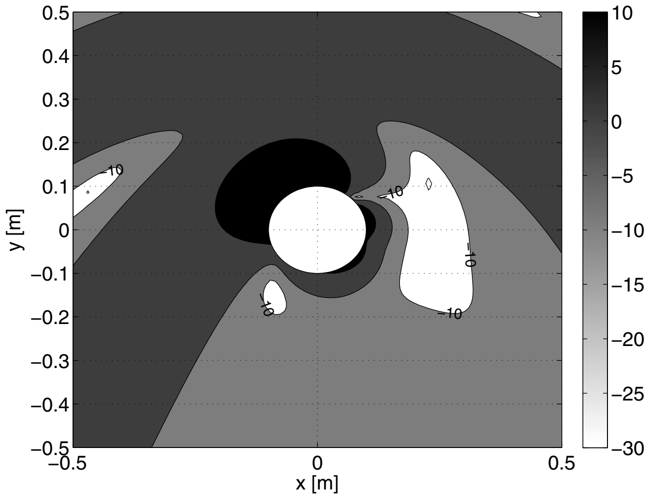


FIG. 12. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to minimize the total pressure as in Fig. 11 when incorporating diagonal regularization in matrix inversion. Primary sound field as in Fig. 7.

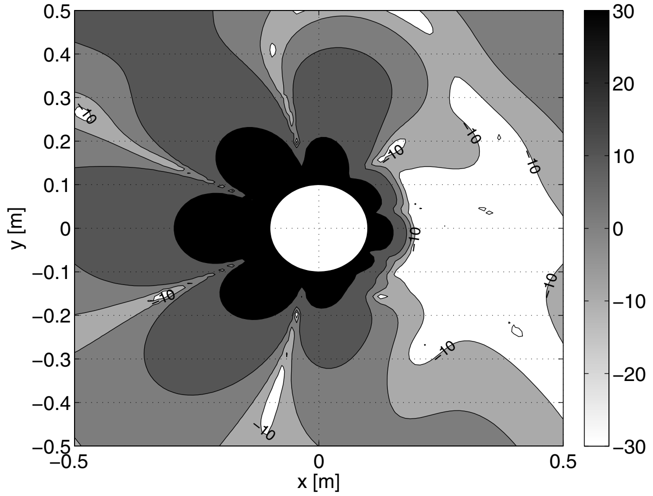


FIG. 13. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to minimize the total pressure as in Fig. 11 when regularization is employed by minimizing the pressure at additional points located all around the space illustrated in the figure. Primary sound field as in Fig. 7.

The next simulation is similar to the one presented in Fig. 12 only here regularization is achieved by adding more pressure minimization points, distributed throughout the entire space presented in the figure. The results presented in Fig. 13 show that a larger quiet zone was achieved compared to Fig. 12, but also at the expense of sound amplification, directed away from the quiet zone. Study of methods that reduce or spatially confine the amplification zones is proposed for future work.

D. Multiple plane-wave primary field with spherical secondary source

One of the potential advantages of the spherical source is its ability to produce complex sound fields. The theoretical formulation above suggested that a shell-like quiet zone, for example, can be achieved for any primary sound field, which

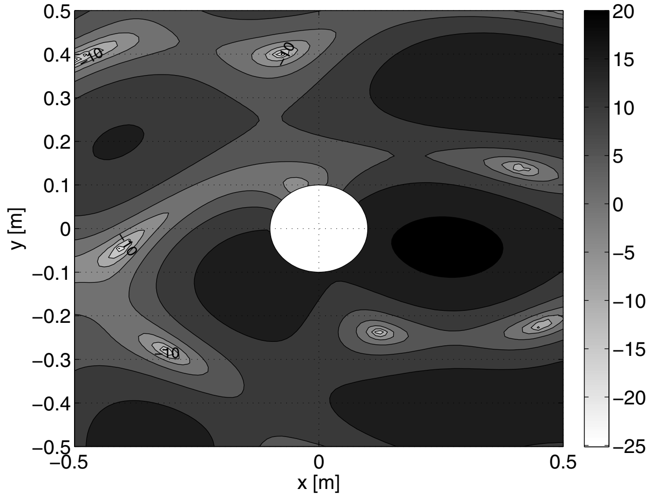


FIG. 14. Magnitude in decibels of a multiple plane-wave primary sound field with secondary source switched off.

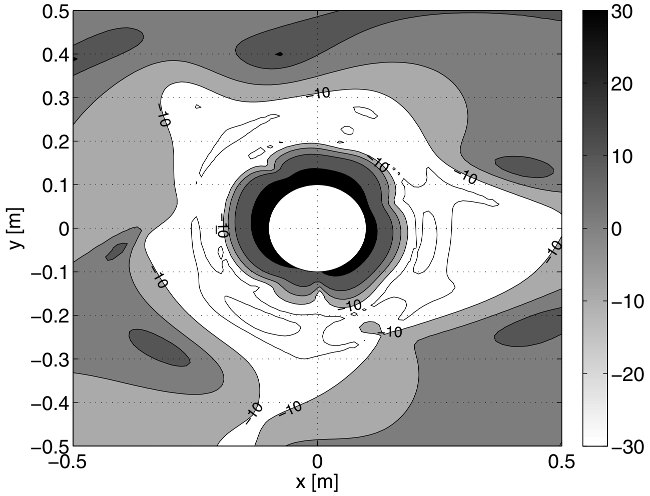


FIG. 15. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to minimize the total pressure as in Fig. 10, but with the primary sound field as in Fig. 14.

has a spherical Fourier transform. In the following example the primary sound field is composed of 20 plane-waves with unit magnitude and random phases and arrival directions, uniformly distributed within the entire range. Figures 14–16 are similar to Figs. 7, 10, and 13, but with the more complex primary sound field. Figure 14 shows the magnitude of the sound field, with a much larger variability compared to the single plane-wave field. Figure 15 shows that a large shell-like quiet zone is achieved even with the complex sound field, verifying the ability of the spherical source to control complex sound fields. Figure 16 shows that large quiet zones can also be achieved in a confined region with complex primary sound fields.

This simulation study presented the potential of the spherical source to produce large quiet zones, even in complex primary sound fields. However, further research is required to investigate ways in which improved quiet zones can be achieved while controlling the zones of sound amplification.

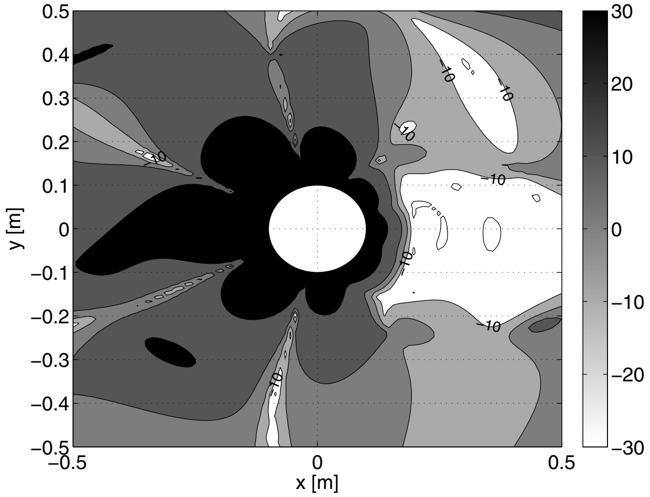


FIG. 16. Magnitude in decibels of the sound attenuation, Eq. (51), with a spherical secondary source designed to minimize the total pressure as in Fig. 13, but with the primary sound field as in Fig. 14.

VIII. CONCLUSION

A theoretical analysis of a spherical loudspeaker array for active control of sound has been presented. Formulation of optimal source velocity has been developed using spherical harmonics representation, showing that the natural cancellation region is a shell around the spherical source. Numerical solutions for source velocity have been proposed to cancel the mean-squared error of the residual pressure at selected locations. The spherical source has shown potential for significantly larger quiet zone compared to a monopole source. However, in some cases, this came at the expense of large amplification in the sound field, suggesting either excessive driving of the secondary source, or inability to suppress the power radiation of the primary source. The design of quiet zones while limiting the regions of amplifications, and the use of a practical spherical loudspeaker array for active control of sound are proposed for future research.

[1] ISO 3382:1997, Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters, International Organization for Standardization, Geneva, Switzerland (1997).

[2] G. K. Behler and S. Muller, "Technique for the derivation of wide band room impulse response," in Proceedings of the 31st Techniacustica Conference, Madrid (2000).

[3] T. W. Leishman, S. Rollins, and H. M. Smith, "An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound," J. Acoust. Soc. Am. **120**, 1411–1422 (2006).

[4] R. S. Martin, I. B. Witew, M. Arana, and M. Vorlander, "Influence of the source orientation on the measurement of acoustic parameters," Acta. Acust. Acust. **93**, 387–397 (2007).

[5] P. Kassakian and D. Wessel, "Design of low-order filters for radiation synthesis," in *Proceedings of the 115th Meeting* (Audio Engineering Society (AES), New York, 2003), p. 5925.

[6] P. Kassakian and D. Wessel, "Characterization of spherical loudspeaker arrays," in *Proceedings of the 117th Meeting* (Audio Engineering Society (AES), San Francisco, CA, 2004), p. 6283.

[7] R. Avizienis, F. Adrian, P. Kassakian, and D. Wessel, "A compact 120 loudspeaker element spherical loudspeaker array with programmable radiation patterns," in *Proceedings of the 120th Meeting* (Audio Engineering Society (AES), Paris, 2006), p. 6783.

[8] G. K. Behler, "Sound source with adjustable directivity (A)," J. Acoust. Soc. Am. **120**, 3224 (2006).

[9] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

[10] S. J. Elliott, P. A. Nelson, I. M. Stothers, and C. C. Boucher, "In-flight experiments on the active control of propeller-induced cabin noise," J. Sound Vib. **140**, 219–238 (1990).

[11] A. J. Kempton, "The ambiguity of acoustic sources—A possibility for active control?," J. Sound Vib. **48**, 475–483 (1976).

[12] J. S. Bolton, B. K. Gardner, and T. A. Beauvilain, "Sound cancellation by the use of secondary multipoles," J. Acoust. Soc. Am. **98**, 2343–2362 (1995).

[13] X. Qiu and C. H. Hansen, "Secondary acoustic source types for active noise control in free field: Monopoles or multipoles?," J. Sound Vib. **232**, 1005–1009 (2000).

[14] P. Joseph, S. J. Elliott, and P. A. Nelson, "Statistical aspects of active control in harmonic enclosed sound fields," J. Sound Vib. **172**, 629–655 (1994).

[15] J. Garcia-Bonito and S. J. Elliott, "Active cancellation of acoustic pressure and particle velocity in the near field of a source," J. Sound Vib. **221**, 85–116 (1999).

[16] F. Zotter and R. Holdrich, "Modeling a spherical loudspeaker system as a multipole source," in Proceedings of the 33rd German Annual Conference on Acoustics, Stuttgart (2007).

[17] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).

[18] T. Matrin and A. Roure, "Optimization of an active noise control system using spherical harmonics expansion of the primary field," J. Sound Vib. **201**, 577–593 (1997).

[19] M. Azarpeyvand, "Active noise cancellation of a spherical multipole source using a radially vibrating spherical baffled piston," Acoust. Phys. **51**, 709–720 (2005).

[20] T. A. Beauvilain and J. S. Bolton, "Sound cancellation by the use of secondary multipoles: Experiments," J. Acoust. Soc. Am. **107**, 1189–202 (2000).

[21] G. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 5th ed. (Academic, San Diego, CA, 2001).

[22] J. R. Driscoll and D. M. Healy, "Computing Fourier-transforms and convolutions on the 2-sphere," Adv. Appl. Math. **15**, 202–250 (1994).

[23] B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," J. Acoust. Soc. Am. **116**, 2149–2157 (2004).

[24] P. S. Meyer and J. D. Meyer, "Multi acoustics prediction program (MAPP) recent results," in Proceedings of the 16th Reproduced Sound, Institute of Acoustics Conference, Stratford (2000).

[25] B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**, 135–143 (2005).

[26] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," IEEE Trans. Signal Process. **55**, 1003–1010 (2007).

[27] E. W. Weisstein, "Spherical cap," "MathWorld, a Wolfram Web Resource," http://mathworld.wolfram.com/SphericalCap.html (last viewed 4/1/2009).

[28] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. (The John Hopkins University Press, Baltimore, MD, 1996).

Response to a change in transport noise exposure: A review of evidence of a change effect

A. L. Brown[a)]
Urban Research Program, Griffith School of Environment, Griffith University, Nathan 4111, Brisbane, Australia

Irene van Kamp
Centre of Environmental Health Research, National Institute for Public Health and the Environment, P.O. Box 1, 32700 BA Bilthoven, The Netherlands

(Received 3 April 2008; revised 8 December 2008; accepted 9 February 2009)

Environmental appraisals of transport infrastructure plans are generally conducted in situations where there will be a step change, or an abrupt change, in noise exposure. While there has been a number of studies of response to step changes in exposure, and seven previous reviews of subsets of these studies, understanding of human response to a change in noise exposure remains limited. Building largely on these previous reviews, this paper examines the evidence that when noise exposure is changed, subjective reaction may not change in the way that would be predicted from steady-state exposure-response relationships. The weight of evidence, while not incontrovertible, is that when exposure changes, responses show an excess response compared to responses predicted from steady-state exposure-response relationships. That is, there is a *change effect* in addition to an *exposure effect*—at least for road studies and at least where the change in exposure results from changes at the source. Further, there appears to be little, if any, adaptation of this excess response with time. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3095802]

PACS number(s): 43.50.Qp, 43.66.Lj [BSF] Pages: 3018–3029

I. INTRODUCTION

Step changes in transport noise occur in situations where (a) new roads and railways are constructed or existing ones closed; (b) there are major increases or decreases in road, rail, or air traffic; (c) noise mitigation measures are implemented in high noise environments; (d) new airport runways are constructed or existing ones closed; (e) there is a major change in the mix of road vehicle types, trucks, in particular; and where (f) there is a major rearrangements of flight paths.

These are always significant changes as far as the community, and authorities, are concerned, and a prediction of the response of the community to that change is an important part of assessment of the proposed changes.

Conventional wisdom is that human response to a step change to transport noise should be able to be predicted from exposure-response curves derived from studies where human response has been assessed over a range of noise conditions. In the past 30 years, many studies have established exposure-response relationships for transport noise. Schultz (1978) and Fidell *et al.* (1991) presented synthesized curves for all such surveys for which noise exposure (as DNL) and the percentage of highly annoyed persons were available. Miedema and Vos (1998) more recently provided synthesized curves separately for aircraft, road traffic, and railway noise, adding the results of some newer studies to the data used by Schultz (1978) and Fidell *et al.* (1991). This synthesis also established confidence limits on the estimate of human response to transport noise exposure (Miedema and Oudshoorn, 2001). The EC working group on health effects of environmental noise recommended (EC/DG ENV, 2002) the relationships presented by Miedema and Oudshoorn (2001) for estimating noise annoyance based on the noise exposure of dwellings.

However, the majority of human response measurements used in these syntheses is likely to have been conducted at sites at which the prevailing noise environment had changed little over preceding years, although we have not confirmed this by examining original site selection methodologies. Further, although again without access to sample selection methodologies, common procedures used in exposure-response studies would have been for respondents who had not been resident for a considerable period, say, a year, to be excluded from the samples, with researchers selecting these to comprise only those respondents chronically exposed to the particular noise dose of the dwellings sampled. Exposure-response curves derived from these studies thus reflect human response to noise in situations of *steady-state*, *constant*, or *unchanging* noise exposure.[1]

These same curves are now used extensively, in noise impact assessments, to estimate likely response of a population experiencing a change in noise exposure. The interest in this paper is whether these steady-state exposure-response relationships estimate human response to a change.

There is continuing interest in response to change (Anotec Consulting, 2003; Huybregts, 2003; Guski, 2004; Van Kempen and van Kamp, 2005; Klæboe *et al.*, 2006). Driving much of this interest is the predicted growth in land and surface traffic, the new infrastructure to accommodate this growth, and community response and health effects associ-

[a)]Electronic mail: lex.brown@griffith.edu.au ated with these changes (for example, Egan *et al.*, 2003). There is rapid growth in traffic at regional airports, and new runways are being planned at major EU airports such as Frankfurt, Schiphol and Heathrow, and throughout Asia.

There is now a number of studies that have examined human response where there has been a *step change*, or *abrupt change*, in noise exposure. The results suggest, although not invariably, that response may be different where there has been an increase or decrease in level, to that predicted from steady-state curves. In other words, human response to change in exposure may include a *change effect* as well as an *exposure effect* and the change effect manifests itself as an *excess response*.

Previously, excess response has been described by various terms such as *exaggerated response* (Huybregts, 2003), *over-reaction* (Fields, 1993; Job, 1988; Schreckenberg and Meis, 2007; Breugelmans *et al.*, 2007), or *overshoot* (Guski, 2004). Lambert *et al.* (1998) used the term *new infrastructure effect*. However, in the psychological literature, overreaction is defined as an exaggerated response or a reaction with unnecessary or inappropriate force, emotional display, or violence. We suggest that terms that carry such connotations be abandoned in favor of the more neutral *excess response*, replacing all of the above.[2] Kastka *et al.* (1995a) and Baughan and Huddart (1993) previously used a related term, *excess effects*. There are also a few reports in the change literature where the opposite of *excess response* has been observed, and while it is not a completely satisfactory antonym to excess response, we suggest that these observations be described as an *under response* to change in exposure.

This paper reviews the literature of studies of human response in situations where the level of transportation noise has changed. Past overviews, individual studies, and more recent results were examined, and we summarize in this paper the weight of evidence available on the existence, magnitude, and persistence of the change effect. In a companion paper (Brown and van Kamp, 2009), we examine the range of explanations that have been suggested for the excess-response phenomenon.

II. THE CHANGE LITERATURE

We examined the scattered, but growing, literature on human response to a change in noise exposure. Some 140 papers were located in a search of the literature, 1980–2006, including *Psycinfo*, *Toxline*, *Embase*, *Medline*, *SciSearch*, *Biosis*, and *Enviroline*. The search profile included keyword(s), in title and or abstract, related to *noise*, *change*, *sensitivity*, *annoyance*, and *(over-) reaction*. In addition, titles were selected from previous reviews of response to change and Internoise Proceedings (1985–2007). From the initial list, a selection process excluded studies where the noise source was not transport (road, air, or rail), the subjects were specific groups (e.g., the hearing impaired), the sounds were highly specific (e.g., sonic booms), or where there was no observed change in noise levels. Outcome measures were restricted to dissatisfaction, annoyance, nuisance, and activity interference/disturbance, or sleep indicators; studies of performance (e.g., in schools) were excluded. Laboratory studies were also excluded, as were studies that reported change in exposure but without reporting response.

Many investigations were reported in more than one paper; multiple studies were sometimes reported in a single paper; and several were re-analyzed and re-reported by several authors. We have chosen not to separate out these matters, but include in the tables below a set of some 40 papers which contain within them the body of studies and prior reviews on human response to a step-change in transport noise exposure over the past 3 decades.

III. THE STUDIES

A. Summary of the studies

Tables I and II provide an overview of the studies located. Over 20 papers (Table I) involved a decrement in exposure, more than 10 an increment in exposure (Table II), and 7 include both increments and decrements. Within the tables, studies have been ordered chronologically and it can be seen that there has been modest, though consistent, interest in the field over more than 3 decades. The tables contain a simple summary of the nature of the change in exposure (cause of the change and approximate magnitude of the change in level when reported) and the survey design—most involving before and after surveys, some also including control sites. The tables also indicate, where available, select observations on measured change effects: excess response or otherwise in community response to change, and adaptation of any excess response over time, although of course the observations from most studies are far more complex than what is able to be shown in this table. They show diverse and sometimes conflicting findings in the various studies.

The search of the change literature found seven prior reviews of these studies of change. Several were particularly detailed in their examination of work to date and provide an excellent foundation for examining the change literature. Most of these had, however, appeared only in the gray literature, and it is appropriate, given their limited dissemination, to summarize the key outcomes of the prior reviews below, then updating these with the findings of more recent studies where possible. The tables show the individual studies included in each of the earlier reviews.

B. Some characteristics of change studies

The studies involved road, air, and rail sources, although not equally, with well over half involving changes in road traffic noise levels, and about one-quarter air transport sources and the rest rail. Residents' responses to noise were those that occurred in their dwellings, although several of the more recent studies included additional responses outdoors.

1. Type of change

A step, or abrupt, change in noise exposure may occur through three different mechanisms. Type 1 changes result from a new or eliminated source, or change in intensity of the source. The majority of the studies (all of the air studies) were Type 1, resulting from changes in traffic flow rates, road bypass construction, or change in runway configura-

TABLE I. Change studies in which there was a decrease in noise exposure.

			Reviews									
Reference	Source	Type	F	V	H	SS	FEZ	S	HU	Noise change	Survey design	Select observations
Fidell and Jones, 1975	Air	1			H		FEZ			Flight paths change, −3 dB	Immed B, immed. A+1 mo. A	No reduction in effects
Lambert, 1978	Road	2	F		H					Barrier, variable up to −10dB	5 mo. B, 17 mo. A, + retro.	No excess response
Kastka and Paulsen, 1979	Road	2			H					7 sites barrier, −3 to −18 (mean −7dB)	~1 yr B, 1 yr A	Under response -but see Kastka *et al.* (1995a)
Kastka, 1981	Road	1		V	H		FEZ			30 sites traffic −8 to 3 (mean −1 dB)	~1 yr B, 1 yr A	Mostly excess response, variable across sites
Mackie and Davies, 1981	Road	1	F		H					10 towns, traffic relief bypass or traffic management, −3 to −5 dB	B, 3–6 mo. A	see Langdon, Griffiths, 1982
Langdon and Griffiths 1982	Road	1	F		H	SS				6 bypass sites −3 to −15 dB	3 mo. B, 4–6 mo. A	Excess response
Richard and Richter-Richard, 1984	Road	1								18 sites traffic calming to −12 (mean –3 dB)	B, 6 mo. A	Excess response,
Brown *et al.*, 1985	Road	1	F		H	SS				1 site traffic reduction −10 dB	15 mo. A, retrospective	Excess response
Babisch and Gebhardt, 1986	Road	1				SS				City traffic reduction −11 dB	B, 1.5 yr A No annoyance meas.	No excess response in disturbances
Griffiths and Raw, 1989	Road	1				SS	FEZ	S		Continuation of previous studies	17–22 mo. A, 7–9 yr A	No (22 mo.) some (7 yr) adaptation
Vincent and Champelovier, 1993	Road	2			H					Barrier −9 dB	1 mo. B, 1 mo. A	Noise was not the only environmental change
Kastka *et al.*, 1995a	Road	2				SS				12 sites barriers 0 to −13 (mean −4 dB)	1–2 yr B, 8–10 yr A	Under and excess response, No adaptation
Kastka *et al.*,1995b	Air	1		V						Airport source reduction −1.5 dB	2 yr A gradual change	Not noticed by community
Öhrström, 1997	Rail	2						S		Various countermeasures + new traffic	B, 3.5 yr A	Expectation possibly influences annoyance
Moehler *et al.*, 1997	Rail	1				SS			HU	Rail grinding source reduction −7 to −8 dB	1 mo. B, 1 and 12 mo. A	No estimate of change effect
Mital and Ramakrishnann, 1997	Road	2						S		1 barrier site	Not reported	Inadequate data on response
Klæboe *et al.*, 1998	Road	1				SS				8 areas, area-wide traffic improvement	B, 7 and 9 yr A	Excess response
Mehra and Lutz, 2000	Road	1								Bypass relief of urban area	B, 3 yr A, retrospective	No estimate of change effect
Öhrström and Skånberg, 2000	Road	1								Bypass tunnel, −9 to −14 dB	2–3 mo. B, 14 mo. A	No excess response
Stansfeld *et al.*, 2001	Road	1						S		Town traffic relief, bypass −2 to −4 dB	6 mo. B, 7–9 mo. A	Noise change too small for annoyance changes
Öhrström, 2004	Road	1								See Öhrström and Skånberg, 2000)	2–3 mo. B, 14 mo. A	No excess response[a], No adaptation
Nilsson and Berglund, 2006	Road	2								Barrier, variable –0 to −7.5 dB	9 mo. B, 15 mo. A	Excess response outdoors, not indoors[a]

[a]But see contrary evidence in Fig. 1 below.

B=before change, A=after change, mo.=month(s), yr=year(s), wk=week(s), and ~=approximately.

Reviews: F: Fields (1994); V: Vallet (1996); H: Horonjeff and Robert (1997); SS: Schuemer and Schreckenberg (2000); FEZ: Fields *et al.* (2000); S: Stansfeld *et al.* (2001); HU: Huybregts (2003).

TABLE II. Change studies in which there was an increase (or an increase and a decrease) in noise exposure.

			Reviews									
Reference	Source	Type	F	V	H	SS	FEZ	S	HU	Noise change	Survey design	Select observations
Jonsson and Sörensen, 1973	Road	1	F		H					New highway, increase not reported	6 mo. A, 18 mo. A	No adaptation
Nimura *et al.*, 1973	Rail	1	F		H					Shinkansen, 2 different lines	8 yr A one line, 4 mo. A other	Difference between lines implies adaptation
Weinstein, 1982	Road	1	F		H	SS				New road, +19 dB	B, 4 and 16 mo. A. Controls	No adaptation
Van Dongen and van den Berg, 1983	Rail	1			H	SS			HU	New line	2 mo. B, 4 mo.A,18 mo. A	Excess response, some adaptation lower levels
Fidell *et al.*, 1985[a]	Air	1	F		H		FEZ			Temporary flight changes, +9 to −19dB	B, weekly A approx, 5 rounds	Excess response (Raw and Griffiths, 1985 reanalysis)
Raw and Griffiths, 1985	Air	1	F		H	SS				Reanalysis of Fidell *et al.* (1985)	See Fidell *et al.*, 1985	Excess response
Griffiths and Raw, 1986[a]	Road	1	F		H	SS	FEZ	S		6 bypass 1 new road, min +/−3 dB	1–4 mo. B, 2–3 mo. A	Excess response
Brown, 1987	Road	1	F		H	SS				1 site traffic increase +6 to −11 dB	2 wk B, 2–7 mo. A, 12 mo. A	Excess response, no adaptation
Baughan and Huddart, 1993[a]	Road	⋯			H					9 decrease 4 increase −10 to +5 dB	1–2 mo. B, 1–2 mo. A	Excess response,
Gjestland *et al.*, 1995	Air	1					FEZ			Temporary flight changes for 3 wk	B, during, A	No response to temporary increase
Job *et al.*, 1996[a]	Air	1								New runway, likely changes only	B only, + expectation of change	Attitude influences reaction
Lambert *et al.*, 1994, 1996	Rail	1		V				S		New TGV line-increases unreported	3–4 yr A	Excess response, adaptation lower levels
Fidell *et al.*, 1996	Air	1				SS	FEZ			Gradual decrease over 3 yr −3 dB	Cross-sectional at end of period	<10% noticed change
Lambert *et al.*, 1998	Various	1							HU	Change as a result of new infrastructure	Compare responses beside infrastructure <5 yr and >10 yr	Change effect for roads, change effect for railways above 67 dB
Hatfield *et al.*, 1998a[a]; Hatfield *et al.*,1998b[a]	Air	1								New runway, increases and decreases	B, separate group A	Some excess response (home owners)
Fidell *et al.*,2000	Air	1								Airport closure and opening 3 airports	behavioral awakening B and A	No changes observed
Hatfield *et al.*, 2001[a]	Air	1								New runway	1–3 mo. A	No change effect, self-reported adaptation
Schreckenberg *et al.*, 2001	Rail	1							HU	Existing line +0 dB, new not yet built	1 yr B, 1 yr A	Expected future annoyance high
Fidell *et al.*, 2002	Air	1								New runway, 0 to +7 dB	15 mo. B, 21 mo. A	Excess response, not all attributable to change
Breugelmans *et al.*, 2007[a]	Air	1								New runway Schiphol. Increases to +4 dB, decreases to −6 dB (mean levels)	Panel study. B, and 3 A at yearly intervals. Control Group	Excess response for increase, not for decrease. Excess response is not explained by non-acoustical factors Generally no adaptation after 2 1/2 yr

[a] These studies also included sites at which there was a decrease in noise exposure.

B=before change, A=after change, mo.=month(s), wk=week(s), yr=year(s), and ∼=approximately.

Reviews: F: Fields (1994); V: Vallet (1996); H: Horonjeff and Robert (1997); SS: Schuemer and Schreckenberg (2000); FEZ: Fields *et al.* (2000); S: Stansfeld *et al.* (2001); HU:Huybregts (2003).

tions. Type 2 changes result from some (usually noise path) mitigation intervention. In Type 2 changes, there are no changes in the transport source flow rates or source noise emissions, just in exposure of the respondents. All of the Type 2 studies involved barriers, mostly along roads—one along a railway. Over three quarters of all change studies were Type 1. There could also be a Type 3 change in which an individual may relocate from one dwelling to another that has a different noise exposure. While this, equally, is a step-change in noise exposure, this type of change was not reported in any of the studies reviewed, though it might usefully be included in future studies seeking to investigate the nature of human response to change.

2. The change in noise exposure

Dimensions of the change in exposure include the direction of the change—increase or decrease, the magnitude of the change, and whether the change is a step change (say, from one day to the next) or whether it is gradual, and if gradual the rate of change—over several weeks, or over, say, a year. Some noise exposure changes may be temporary (such as shutting a runway for maintenance) whereas others are "permanent." Some studies of gradual and of temporary change are included in the studies in Tables I and II, but the majority was step changes.

Different studies used quite different noise metrics, but always those conventionally used to describe noise exposure from the particular transport sources. Indicative changes in level, to the extent that these can be summarized from the original studies, are shown in Tables I and II, with changes reported in decibels, irrespective of the specific noise metric used [in line with the approach suggested by Fields (1990)]. Many of the studies involved substantial change (5 dB or more), though some reported considerably less change, even purporting to be studies of change when accompanied by small changes of 2 dB or less. Part of the conflicting results from different studies of change may result from bundling together the results of studies in which the change in level has been large, and others where the change has been minimal and potentially non-noticeable by respondents.

3. Measures of response

The subset of the literature chosen was that reporting outcomes in terms of dissatisfaction, annoyance, and activity interference (disturbance) including self-reported sleep disturbance measures. The majority of change studies focused on self-reported annoyance/dissatisfaction. Given that most of the studies were conducted before the development of the ISO standard on measuring annoyance (ISO, 2001), they contain different approaches, scales, and measures in the reporting of human response.

4. Study designs

Some change studies have been cross-sectional, comparing results at change sites to those at control sites or against general exposure-response curves. However, most are longitudinal, involving either simple before and after designs, where exposure and response are measured at two points of time, or periodically over a much longer period out to several years after the change. Not all studies have been able to undertake measurements before the change occurred and retrospective assessments (recall) of prior situations have occasionally been utilized, as have assessment of respondent's expectation of the post-change situation.

5. The contexts

There are various contexts in which changes in noise occur. In addition to the major distinction between Type 1 and Type 2 changes, important contextual dimensions include the following: whether respondents knew the change to be permanent or temporary; whether the change occurred with, or immediately following, periods of intensive construction activities; the reason for the noise change; and the activities and attitudes of the authorities responsible for the change. Some noise changes, for example, may occur in contexts in which the authorities are seen to attempt to manage, minimize, or explain the change; others may occur without such interventions. Change situations may also involve community concern or protests over the matters which generated the change in noise exposure. These contextual matters are not examined further in the current paper, but are considered in a review of explanations postulated for the change effect (Brown and van Kamp, 2009).

IV. PREVIOUS REVIEWS OF CHANGE STUDIES

The seven prior reviews were conducted by Fields (1994), Vallet (1996), Horonjeff and Robert (1997), Schuemer and Schreckenberg (2000), Stansfeld *et al.* (2001), Fields *et al.*. (2000) and Huybregts (2003). Purpose, methodology, and comprehensiveness of these reviews varied.

Fields (1994) conducted a meta-analysis on 282 studies of community reaction to environmental noise, examining the effect of personal and what he called situational variables on reported effects of noise. A subset of these studies involved a change in noise exposure. Fields (1994) tested specific hypotheses: whether a new noise or a change in noise impacted annoyance more than would be predicted from reactions to a familiar, existing noise (excess response) and whether annoyance decreased as the time since an increase in noise level lengthened (adaptation of the change effect). His results are shown in Tables III and IV. In Table III, change in exposure was present in 14 of the studies he reviewed (for which there were 19 researcher findings included in the analysis). Of the 19 findings, 8 concerned airports.

Fields (1994) concluded, from the studies he reviewed, that the evidence was too unevenly divided to indicate whether changes in noise cause excess response to the change. While excess response was observed in 42% of the findings (32% of responses), under response was observed in 11% (34% of the responses), with the remainder showing no important difference. When the analysis included only standard surveys, the majority (some 60% of findings and responses) indicated excess response. The studies included increases in exposure as well as decreases, and Fields (1994) separately examined these, concluding that there was not a

TABLE III. Results of Fields (1994) meta-analysis examining community reaction to a change (increase or decrease) in noise exposure. "Findings" relate to observation by the original researchers, and "interviews" to the total number of respondents.

			Percent of findings or interviews that show when the noise exposure changes		
Type of findings	Findings or interviews	Number of findings or interviews	Respondents under-react to a change	No important difference	Respondents over-react to a change[a]
Standard and nonstandard[b]	Findings	19	11%	47%	42%
	Interviews/responses	14 097	34%	35%	32%
Standard only	Findings	13	8%	31%	61%
	Interviews/responses	7 737	11%	29%	60%

[a] We use the term of Fields (1994) *over-react* (and *under-react*) rather than our preferred terms *excess response* (and *under response*) in this table.
[b] Nonstandard studies were those which the author had classified as having some methodological weakness.

clear pattern distinguishing between the results where there were increments or decrements in noise exposures.

Following a change in noise exposure, the meta-analysis also found mixed evidence as to whether the levels of annoyance with the new levels of exposure decreased (adapted) over time. Table IV shows that while 43% of findings (49% of the response) show such adaptation, 43% of findings (13% of the responses) showed the opposite—annoyance increased over time since the change—and the remainder showed no increase or decrease.

Vallet (1996) did not undertake a similar systematic review of studies of change, instead making a series of observations, mostly from airport noise studies, on the effects of change and the methodology of change studies. Our paraphrasing of his observations is as follows.

- Annoyance shows some inertness to the physical change in levels both for gradual change (gradual increase as traffic loads increased or gradual decrease as noise emission levels of individual aircraft decreased) and for step changes such as new airports or runways. He suggested, though on what evidence is unclear, that a minimum of 6 dBA L_{eq} of the noise exposure is necessary before there is a change in the annoyance level.
- Physiological measures showed a larger and quicker response to the change than did annoyance, where there was a time delay in a decrease in annoyance (at Los Angeles) and a clear excess response to new aircraft noise (in Paris).
- Non-acoustical variables such as fear, noise sensitivity, attitudes toward the source, belief that authorities could prevent or reduce the noise, and advertising campaigns all influence the noise response and need to be investigated in change studies.

Horonjeff and Robert (1997) undertook an extensive review, building largely on the work of Fields (1994), identifying 23 change studies in 51 citations, covering road (12 studies), rail (2), and air (9) transport sources. The purpose of their review was as precursor to the design of further studies to develop a predictive model for the effect of change in noise exposure around airports, and included a detailed examination of the purpose, methodology, results, and limitations in the studies conducted to that time. Of particular interest was their synthesis of the magnitude of any change effect measured in the studies they reviewed. Such a synthesis required them to make approximations (described in the original paper) to overcome the difficulties presented by different acoustic measures, response scales, and available baseline[3] responses from which to estimate the change effect. This analysis by Horonjeff and Robert (1997) is a pivotal review as it contains quantitative estimates of excess response in the change studies. We report their analysis of excess response below, and add newer data to it from studies conducted since their review.

The Horonjeff and Robert (1997) synthesis was in terms of a decibel-equivalent estimate (see Fields, 1990) of the magnitude of the *change* effect—what they called the *abrupt-change effect*. The decibel-equivalent change effect is the change, in decibels, on the exposure-response curve, *ad-*

TABLE IV. Results of Fields (1994) meta-analysis examining whether annoyance with a new source increases with time after an increase in noise exposure.

			Percent of findings or interviews that show when the noise exposure changes		
Type of findings	Findings or interviews	Number of findings or interviews	Annoyance increases	No important difference	Annoyance decreases
Standard and nonstandard	Findings	7	43%	14%	43%
	Interviews/responses	1581	13%	38%	49%
Standard only	Findings	6	33%	17%	50%
	Interviews/response	1450	6%	41%	53%

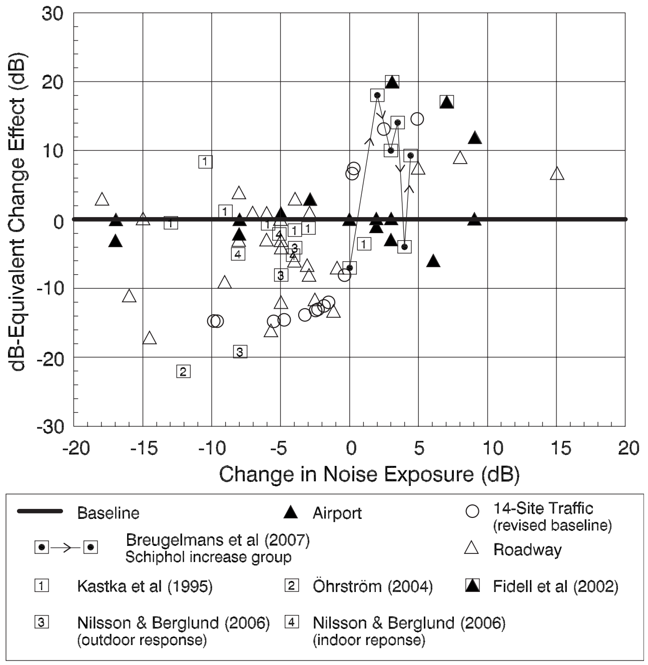


FIG. 1. Decibel-equivalent change effect (excess response) for different levels of change in noise exposure. Most of the data points are from the original review by Horonjeff and Robert (1997). The data points plotted with square symbols have been added by the present authors from more recent studies—see Sec. V. The line with arrows connects the data points tracing the trajectory of excess response of one panel over $2\frac{1}{2}$ years (six remeasurements) following an increase in aircraft noise at Schiphol airport (Breugelmans *et al.*, 2007)—and was also added by the present authors. Most of the studies (first and third quadrants) indicate an excess response, though a few (second and fourth quadrants) show the opposite—an under response to the change. (after Horonjeff and Robert, 1997).

ditional to the before to after change in exposure, which would predict the change in response between the before and after conditions. Their synthesis is shown in Fig. 1 [which included adjusting the baseline used in one study: the *14-Site Traffic* study of Baughan and Huddart (1993)]. Figure 1 shows the change in exposure on the horizontal axis (−18 dB decreases to +15 dB increases) and the decibel-equivalent change effect on the vertical axis. In the words of Horonjeff and Robert (1997):

> Data points that lie along the horizontal axis (0 dB-Equivalent change effect) indicate cases in which the baseline dose-response curve for the study correctly estimates the change in annoyance: no abrupt change in effect was observed.…
>
> Data points in the first and third quadrants … indicate cases in which there was an abrupt-change effect, and the effect resulted in a change in annoyance greater than that predicted from the baseline curve.

Figure 1 shows that much of the data falls into the first and third quadrants, that is, the majority of results support the existence of an excess-response change effect. There are no data points representing a strong under response, while there are large excess responses to small level changes of 1–2 dB. Results from more recent studies (see Sec. V below) are also plotted in this figure.

Horonjeff and Robert (1997) also found that nine studies designed to measure the decay of the excess response generally failed to find evidence of decay, that is, there was little evidence of *adaptation* or *habituation* of the change effect. Most first post-change interviews were conducted 3–7 months after the change (one at 0.5 months, and one at 12 months), with last post-change interviews conducted 16–96 months after the change. There is no supporting evidence that excess response attenuates with time.

Horonjeff and Robert (1997) drew an extensive set of conclusions with respect to methodology for change studies—among which is that it is critical that change studies need to be longitudinal as it is difficult to base any conclusions about change on cross-sectional studies. They also noted the need for site-specific baseline exposure-response curves and for accounting for the possible influence of mediating variables (such as the specific neighborhood, demographic variables, attitudes, and expectations of the respondents concerning the noise source).

Schuemer and Schreckenberg (2000) reviewed 14 of what they termed exemplary studies on step changes in transport noise. They concluded the following.

- In ten studies an excess response to change in noise levels was confirmed—for both incremental and decremental noise changes.
- There was no evidence of adaptation of the excess response.
- The authors also noted that the characterization of a change as *before* and *after* may be simplistic, identifying that most situations are more likely to have several different before situations (for example, before announcement, after announcement, and construction) and after situations (immediately upon change, days or weeks after, and years after).

Fields *et al.* (2000) reported a rigorous analysis of previous studies and reviews of change that was aimed at developing tools to design further studies to estimate the step-change effect in community response to a change in noise levels, particularly for airports. They made a large number of recommendations and observations, and only a selection of these is reported here.

- Several previous studies of changes in noise environments have been conducted where the changes in exposure were too small to have a reasonable probability of being detected (e.g., Fidell *et al.*, 1985 and Gjestland *et al.*, 1990). They suggested that almost nothing can be learnt about the effects of change unless those changes involve considerable changes in noise exposure;
- Permanent changes are likely to produce different reactions to temporary changes [for example, temporary runway closure for repairs (Fidell *et al.*, 1985) and short-term military exercises (Gjestland *et al.*, 1995)].
- Response to changes in source levels may be quite different to change effects generated by path attenuation (house insulation or barrier construction);
- Other mediating variables that affect annoyance can be spatially or temporally correlated with the noise change

and can distort measures of the effect of change. Both cross-sectional studies and longitudinal studies have problems in isolating the change effect, with longitudinal studies preferred utilizing repeat measures of the same subjects, though with the need to control for the correlated variables to the maximum extent possible.
- Evidence is presented, on the basis of reanalysis of existing studies, that panel studies increase the precision of estimates of differences in responses between two points of time, and that panel studies do not appear to introduce survey-resurvey bias, particularly if repeat surveys are at least 1 month apart. However, they noted there is a lack of consistency in this result among some of the surveys analyzed, and they suggested that panel studies should be strengthened by the addition of non-panel respondents in the repeat surveys.

The Stansfeld *et al.* (2001) review was prepared to provide evidence to inform a rapid prospective Health Impact Assessment for the Mayor of London's Ambient Noise Strategy. The report focused on non-auditory health impacts, causal pathways to explain health effects, and interventions that can improve health. Change studies were briefly reviewed in the context of a section dealing with interventions, but no specific mention was made of excess response or adaptation. The authors' conclusions with respect to annoyance and sleep disturbance were as follows.

- The literature on intervention studies is sparse.
- Intervention studies were of variable quality.
- Intervention such as noise barriers and reduction of road traffic noise levels at source by approximately 10 dBA seem to decrease levels of annoyance in communities exposed to road traffic.
- Relatively few community studies assessing the impact of noise reduction on sleep have been conducted. Those studies that have been carried out with a reduction between 6 and 14 dBA resulted in both subjective and objective improvements in sleep.

Huybregts (2003) reviewed seven previous studies of change in railway noise and concluded that there is no reason to doubt that, similar to other forms of transport noise, there is an excess response (he termed it an exaggerated community response) when railway noise exposure changes, though he suggested further work is needed to understand both magnitude and duration of the change effect.

V. MORE RECENT CHANGE STUDIES

A range of field studies of change has been conducted since the various reviews (see Tables I and II) and show similar diversity in study designs, and results as found in the earlier studies. Inclusion of some of the results from these later studies in the previous syntheses is possible.

Decibel-equivalent magnitudes of the change effect have been able to be estimated by the current authors by applying the same methodology used by Horonjeff and Robert (1997). These are included in Fig. 1. Two of the seven sites in the study Fidell *et al.* (2002) of change in aircraft noise levels experienced sufficient increase in exposure to allow decibel-equivalent change effects to be estimated [we used the FICON (1992) exposure-response curve to estimate the change effect from the reported data]. Nilsson and Berglund (2006) and Öhrström (2004) reported studies of decrease in road traffic noise exposure, the first from the placement of a barrier, and the second from a reduction in traffic flow. These authors suggested that there was no excess response to the change indoors, but our reanalysis [using the Miedema and Oudshoorn (2001) and Miedema and Vos (1998) exposure-response curves respectively] suggests that there was a large change effect at three of the "sites" (actually three "distance from road categories"—change in noise exposure of more distant categories could not be estimated from the paper) in the barrier study and at the one site in the traffic reduction study. Kastka *et al.* (1995a) revisited the barrier sites reported previously (Kastka and Paulsen, 1979), reporting new data and readjusting their steady-state exposure-response baseline. Kastka *et al.* (1995a) examined residents' responses in 1988 and 1976 to barriers that had been constructed after the first survey. We have calculated decibel-equivalent changes at their seven barrier sites (using their noise disturbance score and their before exposure-response relationship—Table 10 in Kastka *et al.* (1995a)—and these have also been shown in Fig. 1. At five of the sites, there is a small excess response, but an under response, one large, at two sites.

A recent longitudinal study of response to noise around Schiphol Airport incorporates the most comprehensive and purpose-designed study of change to date, though detailed results are not yet widely reported (Ministry of Transport, Public Works and Water Management, 2005). Surveys of effects of aircraft noise exposure were conducted around Schiphol in 1996, 2002, and 2005 (Houthuijs *et al.*, 2007). A new runway at the airport was opened in February 2003, and a panel of 640 persons, whose exposure was likely to change as a result of the new runway, was selected from the 2002 survey group. This panel was resurveyed annually over the 2 1/2 years following the change, with half of the panel surveyed in northern hemisphere springs and half in autumns, giving six data points subsequent to the change (Breugelmans *et al.*, 2007). In total, 478 respondents completed four panel interviews, one before the change and three after the change. The panel was made up of three subgroups: one experiencing an increase in exposure, one a decrease, and one as control experiencing negligible change. Results from the first (before change) panel round were used to derive a baseline exposure-response relationship based on noise exposure (L_{den}) over the previous 12 months.

Breugelmans *et al.* (2007) reported significant excess response for the subgroup experiencing the increase in exposure. Decibel-equivalent change effects have been estimated by the current authors based on the before-change exposure-response relationship, and the trajectory of excess response for this subgroup over the six repeats is shown in Fig. 1. Excess response was observed from the second round of surveys and continued throughout the study. There was a drop in the penultimate round but a return to large excess-response in the latest round. The subgroup experiencing the decrease in exposure, and the control group experiencing negligible change, did not exhibit excess response in any of the survey rounds.

This study was particularly notable in that, together with post-change longitudinal measurement of annoyance (and sleep disturbance and self-reported change in general health), longitudinal measurements were also made on a wide range of non-acoustical factors including living satisfaction, noise sensitivity, expectations about the airport, neighborhood quality and future noise levels, fear of aircraft crashes, and negative attitudes toward the airport. A generalized linear mixed model analysis found that the excess response was not diminished significantly when adjusted for all of the non-acoustic factors. The findings suggest that the change effect was driven primarily by the change in noise exposure.

Overall, these more recent studies show magnitudes of change effect excess response (Fig. 1) in line with those reported in the original synthesis by Horonjeff and Robert (1997).

Klæboe *et al.* (1998) also reported an excess response, with a decibel-equivalent change effect of 5–8 dB resulting from an area-wide reduction in noise exposure (and a reduction in other factors such as air pollution exposure) from traffic reductions affecting a whole town area in Oslo from 1987 to 1994/96. This result was the aggregate across all exposure levels ($50-75L_{\mathrm{eq,24\ h}}$, particularly those less than $70L_{\mathrm{eq,24\ h}}$) but the results cannot be included in Fig. 1 as changes in noise exposure at individual sites were not reported.

Other observations from the more recent change studies include suggestions that expectation of increased annoyance had a noticeable effect on the level of annoyance before noise countermeasures were realized (Öhrström, 1997; Schreckenberg *et al.*, 2001); a bypass achieving a mean reduction in noise of 2–4 dB produced no noticeable change in annoyance (Stansfeld *et al.*, 2001) and suggestions that adaptation occurs soon after changes in noise exposure (Hatfield *et al.*, 2001).

VI. DISCUSSION

A. Excess response to change

1. Airport studies and studies of gradual change in exposure

As already noted by Horonjeff and Robert (1997), the results for the airport studies were, in general, quite different to those for the road studies. With the exception of our estimate of large excess response in the studies by Fidell *et al.* (2002) and by Breugelmans *et al.* (2007), the change effect in the airport studies was very small—in some cases, an under response—compared to the predominance of excess response in the road studies. While this may demonstrate a difference in response to change between aircraft noise and roadway noise, another, and perhaps more obvious, explanation is that the difference may be an artifact of the nature of the particular noise changes that occurred at most of the airports studied.

Horonjeff and Robert (1997) also noted that most of the airport studies they reviewed either involved temporary changes in noise exposures (Fidell *et al.*, 1985; Raw and Griffiths, 1985; Gjestland *et al.*, 1995) or small changes of 3 dB or less in noise exposure (Fidell and Jones, 1975; Fidell *et al.*, 1985). Some airport change studies (Fidell *et al.*, 1996; Kastka *et al.*, 1995b) and some road change studies (Stansfeld *et al.*, 2001) also had the acoustic characteristic of a gradual change in noise exposure. As Fields *et al.* (2000) previously noted, these are very different situations to where there is an abrupt or step change in exposure.

In summary, because of the potentially confounding effect of the limited magnitude and different nature of the changes that occurred in the various airport studies for which data are available, it would be inappropriate to draw conclusions from these studies about response to change around airports. Further studies involving change at airports that do not have these constraints [as, for example, the Schiphol study reported by Breugelmans *et al.* (2007) and Houthuijs *et al.* (2007)] will be necessary to examine whether there might be any difference between response to change for different transport modes. The same applies for situations, for any mode, for where there has been a gradual change in exposure as against a step-change.

2. Type 2 changes

Studies of both Type 1 and Type 2 changes were included in the reviews, and there is some evidence that people may respond differently in Type 2 changes, reporting less response and little or no change effect (Griffiths and Raw, 1986). Langdon and Griffiths (1982) re-examined the results of Kastka and Paulsen’s (1979) longitudinal study of barriers and found under response, explaining this as due to the differential effect of noise reductions by barriers rather than reductions of the noise source. However, as noted above, using the new data for these sites from Kastka *et al.* (1995a), there was a small excess response at the majority of the sites (see Fig. 1) but an under response at two sites. Vincent and Champelovier (1993) reported that noise annoyance shows only a small reduction for a 9 dB drop of noise levels resulting from barrier construction at their one site, though, in fact, the reduction in percent highly annoyed appears generally in line with expected reductions from steady-state exposure-response curves, that is, no excess response. No excess response to change was also suggested in the longitudinal study by Lambert (1978) of the effect of a single barrier. While Nilsson and Berglund (2006) reported no excess response in a barrier study, re-analysis by the current authors (Sec. V above) suggests that there was. Baughan and Huddart (1993) also noted that the change effect may not be present in Type 2 changes

While Fields *et al.* (2000) concluded that studies aimed at evaluating the effect of noise-shielding interventions (barriers, double glazing) rarely lead to findings of an excess response, evidence of the presence and direction of change effects in Type 2 studies to date is ambiguous. A reasonable conclusion at this stage is that the results of Type 1 and Type 2 studies should be separated in any future analysis of change studies given the mixed evidence above regarding excess response in Type 2 studies.

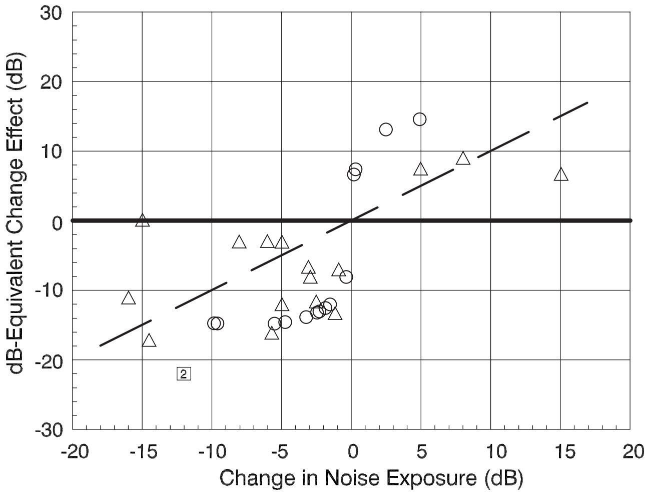


FIG. 2. Decibel-equivalent excess response change effect for Type 1 changes of road traffic sources only. The broken line indicates a change effect of the same magnitude (decibel-equivalent) as the change in noise exposure. Legend as in Fig. 1.

3. Type 1 road changes

Given the conclusions above regarding airport change studies, and Type 2 change studies, it is reasonable to examine a specific subset of the studies—those where the source was road traffic and where the nature of the change in exposure was Type 1 changes.

Figure 2 shows the same data set as Fig. 1, but deletes aircraft noise studies and Type 2 road studies where the noise exposure change occurred through modification of the transmission path—usually by the installation of barriers. The remaining studies are those where the change in noise exposure has resulted from changes in the road source itself—the construction of new roads, either as new sources or providing traffic relief on existing roads—or some other change in traffic flow. The weight of evidence from this subset of field studies is that there is an excess-response change effect in annoyance responses for these changes in noise exposure. All available studies demonstrate, with remarkable consistency, an excess response in situations of both increments and decrements of noise exposure: respondents whose noise exposure has increased report more annoyance than expected from steady-state studies; respondents whose noise exposure has decreased report less annoyance than expected from steady-state studies. The effect is present even for quite small changes in noise exposure.

The broken line shown in Fig. 2 is not a line of best-fit as we have chosen not to suggest a predictive relationship between noise change and its associated change effect from the studies reviewed—given the differences between the studies in terms of metrics and designs, and the approximations necessary to estimate the change effect from the data reported in them. The broken line indicates where the magnitude of the change effect is equal to the magnitude of the change in exposure. Figure 2 clearly shows that, in road studies with Type 1 changes, the excess-response change effect (decible-equivalent) tends to be greater (often much greater) than the change in noise level exposure itself.

4. Annoyance versus activity interference in change responses

The results reported in Fig. 2 are for annoyance responses. Several authors have found, or suggest, that activity interferences (speech interference, closing windows, etc.) may not display the same level of excess response as do annoyance measures when noise exposure changes.

Kastka *et al.* (1995a) found that interference effects of noise did not show the same change effects that were observed in annoyance scores in their longitudinal study of the effect of noise reduction by barriers. Klæboe *et al.* (1998) also provided evidence from their road traffic reduction study in Oslo that, while there was a significant excess-response change effect for traffic noise annoyance, there was no similar systematic change effect in the reporting of disturbances/inconveniences from road traffic noise. Babisch and Gebhardt (1986) reported that there was no excess response to a reduction of some 11 dB in road traffic noise levels in inner city Berlin, 1.5 years after the change, as their result did not deviate from steady-state response data collected in Hamburg. However, as Klæboe *et al.* (1998) pointed out, this study had measured changes in interference activities only, not annoyance, and such a result is in line with the findings above of no observed excess response in activity interference. Breugelmans *et al.* (2007) also found no excess response in sleep disturbance, despite a large excess response in annoyance. However, Öhrström (2004) reported good correlations between reductions in annoyance scores and reductions in activity disturbances in her longitudinal study of a large decrease in road traffic noise exposure.

B. Adaptation after the change

Only a small number of recent studies contributed data tracking respondents' reactions for an extended period after the change. Griffiths and Raw (1987, 1989) extended a previous longitudinal study of response to reduced traffic noise (Griffiths and Raw, 1986). They found that some 40% of the large excess response to change they had originally measured 2 years after the reduction was still present 7–9 years after the change. Moehler *et al.* (1997) showed that annoyance reductions achieved by noise reductions through rail grinding still persisted in a third survey, 12 months after the initial survey following the noise reductions. Klæboe *et al.* (1998) reported similar exposure-response curves at two time periods (2 years apart), after area-wide improvements had reduced exposure, indicating persistence of the excess-response change effect. In the longitudinal study at Schiphol airport, Breugelmans *et al.* (2007) reported no sustained adaptation in excess response to increased exposure over 2 1/2 years after the change. There is thus no evidence from recent work to alter the conclusions reached by Horonjeff and Robert (1997), in their review of nine longitudinal studies, that there is little evidence that excess response attenuates within several years of the change.

VII. CONCLUSIONS

Studies of human response to a change in noise exposure over 3 decades have produced results which suggest that human response to change is not in line with what would be expected from steady-state exposure-response curves. Building on previous syntheses of change studies, this review concludes that a change effect of excess response to a step-change in road traffic noise occurs in addition to the exposure effect. This occurs in noise annoyance responses though not in activity interference responses. Consistent evidence of a similar change effect for aircraft noise and railway noise changes is lacking but, rather than this indicating that human response to change is different between different transportation noise sources, we suggest that this may be a result of the nature of the noise changes available in most aircraft and railway noise change studies to date: generally small, gradual, or temporary.

A change effect is unequivocally present in the results of the road traffic noise studies where the intensity of the road traffic source changes through changes in traffic volume on the source roads (Type 1 changes). For these types of change situations, the decibel-equivalent magnitude of the excess responses (both the excess benefit arising from reductions in exposure and the excess disbenefits arising from increases in exposure) can be greater, often much greater, than the change in noise levels itself. For changes resulting from the insertion of barriers or other path mitigation interventions (Type II changes), the evidence for a change effect is not clear. The excess-response change effect does not appear to attenuate over time—even years after the change.

Further studies of change are required as many of the studies of change reviewed have been characterized by weak or inappropriate designs. The difficulty, of course, is to find “natural laboratories” where such change research can be conducted. It will only be through the careful design of change studies that it will be possible to unravel, and explain, this phenomenon, and provide decision makers with practical and quantitative advice on community response to changes in noise exposure. The presence, magnitude, and persistence of the excess response warrants consideration of a change effect in assessing the impact of infrastructure changes and in policy making with respect to such changes.

[1] While we refer to exposure-response relationships derived under these conditions as *steady state* curves, there is a caveat to this descriptor. We observe later in this paper that change effects are persistent; hence any previous change in exposure at a site should continue to influence the response of those long-term residents who had experienced the change, and thus be “mixed” with the response of other residents at the site who had moved in since any change in exposure. We examine some limited evidence on the differential response of long-term residents and newcomers in Brown and van Kamp (2009).

[2] While we use the preferred term excess response throughout this paper, where we are reviewing the work and results of other authors, it is often more appropriate to use the terminology they have adopted; most often over-reaction.

[3] The baseline response in any change study is the *steady-state* exposure response curve applicable to the sites in that study before change in exposure occurred. The origin of these baselines varied from study to study, ranging from local baseline exposure-response curves derived within the study itself, to control site exposure-response curves, to synthesized exposure-response curves (mostly Schultz, 1978)—see Horonjeff and Robert (1997) for full details. The baseline provides the datum from which change effects are estimated.

Anotec Consulting (**2003**). “Studies on current and future aircraft noise exposure at and around community airports. Summary for policy makers,” Document No. PAN012-6-0.

Babisch, W., and Gebhardt, S. (**1986**). “Gestortheitsreaktionen durch Verkehrslamr—Eine ‘vorher/nachher’-untersuchung. (Annoyance reactions caused by traffic noise—A ‘before/after’-study),” ZfLärmbek. **33**, 38–45.

Baughan, C., and Huddart, L. (**1993**). “Effects of traffic noise changes on residents’ nuisance ratings,” in Proceedings of the Sixth International Congress on Noise as a Public Health Problem, Noise & Man 1993, Nice, July, Vol. **2**, pp. 585–588.

Breugelmans, O., Houthuijs, D., van Kamp, I., Stellato, R., van Wiechen, C., and Doornbos, G. (**2007**). “Longitudinal effects of a sudden change in aircraft noise exposure on annoyance and sleep disturbance around Amsterdam Airport,” in Proceedings of the ICA, Madrid, Paper No. ENV-04-002-IP.

Brown, A. L. (**1987**). “Responses to an increase in road traffic noise,” J. Sound Vib. **117**, 69–80.

Brown, A. L., Hall, A., and Kyle-Little, J. (**1985**). “Response to a reduction in traffic noise exposure,” J. Sound Vib. **98**, 235–246.

Brown, A. L., and van Kamp, I. (**2009**). “Response to a change in transport noise exposure: Competing explanations of change effects,” J. Acoust. Soc. Am. **125**, 905–914

EC/DG ENV (**2002**). “Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise,” OJ L 189, 18.7.2002, EU Parliament, Brussels, pp. 12–25.

Egan, M., Petticrew, M., Ogilvie, D., and Hamilton, V. (**2003**). “New roads and human health: A systematic review,” Am. J. Public Health **93**, 1463–1471.

FICON (**1992**). Federal Agency Review of Selected Airport Noise Analysis Issues, Federal Interagency Committee on Noise, FICUN, Wahington, DC.

Fidell, S., Barber, D., and 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., Horonjeff, R., Mills, J., Baldwin, E., Teffeteller, S., and Pearsons, K. (**1985**). “Aircraft annoyance at three joint air carrier and general aviation airports,” J. Acoust. Soc. Am. **77**, 1054–1068.

Fidell, S., and Jones, G. (**1975**). “Effects of cessation of late-night flights on an airport community,” J. Sound Vib. **42**, 411–427.

Fidell, S., Pearsons, K., Tabachnik, B. G., and Howe, R. (**2000**). “Effects on sleep disturbance of changes in aircraft noise near three airports,” J. Acoust. Soc. Am. **107**, 2535–2547.

Fidell, S., Silvati, L. and Haboly, E. (**2002**). “Social survey of community response to a step change in aircraft noise exposure,” J. Acoust. Soc. Am. **111**, 200–209.

Fidell, S., Silvati, L., and Pearsons, K. (**1996**). “On the noticeability of small and gradual declines in aircraft noise exposure levels,” in Proceedings of Internoise 1996, Liverpool, UK, Book 5, pp. 2247–2252.

Fields, J. M. (**1990**). “Policy-related goals for community response studies,” Environ. Int. **16**, 501–514.

Fields, J. M. (**1993**). “Effect of personal and situational variables on noise annoyance in residential areas,” J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M. (**1994**). “A review of an updated synthesis of noise/annoyance relationships,” NASA Report No. CR-194950, NASA Langley Research Center, Hampton, VA.

Fields, J. M., Ehrlich, G. E., and Zador, P. (**2000**). “Theory and design tools for studies of reactions to abrupt changes in noise exposure,” NASA Report No. CR-2000-210280, NASA Langley Research Center, Hampton, VA.

Gjestland, T., Liasjø, K., and Granøien, I. (**1995**). “Community response to noise from short-term military aircraft exercises,” J. Sound Vib. **182**, 221–228.

Gjestland, T., Liasjø, K., Granøien, I., and Fields, J. M. (**1990**). “Response to noise around Oslo Airport Fornebu,” DELAB Report No. STF40 A90189, Civil Aviation Administration, Oslo.

Griffiths, I. D., and Raw, G. J. (**1986**). “Community and individual response to changes in traffic noise exposure,” J. Sound Vib. **111**, 209–217.

Griffiths, I. D., and Raw, G. J. (**1987**). “Community and individual response to changes in traffic noise exposure,” in *Environmental Annoyance: Characterization, Measurement, and Control*, edited by H. S. Koelega (Elsevier, Amsterdam), pp. 333–343.

Griffiths, I. D., and Raw, G. J. (**1989**). “Adaptation to changes in traffic noise exposure,” J. Sound Vib. **132**, 331–336.

Guski, R. (**2004**). “How to forecast community annoyance in planning noisy facilities," Noise Health **6**, 59–64.

Hatfield, J., Job, R. F. S., Carter, N. L., Peploe, P., Taylor, R., and Morell, S. (**1998b**). "Attitude-mediated reaction to noise influences physiological responses: Evidence supporting causality," in Proceedings of Internoise 1998, Christchurch, New Zealand, Paper No. 440.

Hatfield, J., Job, R. F. S., Carter, N. L., Peploe, P., Taylor, R., and Morell, S. (**2001**). "The role of adaptation in responses to noise exposure: Comparison of steady state with newly high noise areas," in Proceedings of the Fourth European Conference on Noise Control, Euronoise PATRA, January.

Hatfield, J., Job, R. F. S., Peploe, P., Carter, N. L., Taylor, R., and Morell, S. (**1998a**). "Demographic variables may have a greater modifying effect on reaction to noise when noise exposure changes," in Proceedings of the Noise Effects 1998, Seventh International Congress on Noise as a Public Health Problem, Sydney, Vol. 2, pp. 527–530.

Horonjeff, R. D., and Robert, W. E. (**1997**). "Attitudinal response to changes in noise exposure in residential communities," NASA Report No. CR-97-205813, National Aeronautics and Space Administration, Washington, DC.

Houthuijs, D., Breugemans, O., van Kamp, I., and van Wiechen, C. (**2007**). "Burden of annoyance dues to aircraft noise and non-acoustical factors," in Proceedings of Internoise 2007, Istanbul, Paper No. 838472.

Huybregts, C. (**2003**). "Community response to changes in railway noise exposure—A review," in Proceedings of Wespac VIII, Melbourne, Australia, April.

ISO/TS 15666:2003 Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys (**2003**).

Job, R. F. S. (**1988**). "Over-reaction to changes in noise exposure: The possible effect of attitude," J. Sound Vib. **126**, 550–552.

Job, R. F. S., Topple, A., Carter, N. L., Peploe, P., Taylor, R., and Morell, S. (**1996**). "Public reactions to changes in noise levels around Sydney Airport," in Proceedings of Internoise 1996, Liverpool, UK, Book 5, pp. 2419–2424.

Jonsson, E., and Sörensen, S. (**1973**). "Adaptation to community noise—A case study," J. Sound Vib. **26**, 571–575.

Kastka, J. (**1981**). "Zum Einfluss verkehrsberuhigender maßnahmen auf lärmbelastung und lärmbelästigung (The influence of traffic calming measures on noise load and noise annoyance)," ZfLärmbek **28**, 25–30.

Kastka, J., Borsch-Galetke, E., Guski, R., Krauth, J., Paulsen, R., Schuemer, R., and Oliva, C. (**1995b**). "Longitudinal study on aircraft noise effects at Düsseldorf airport 1981–1993," in Proceedings of the 15th ICA, Trondheim, Vol. IV, pp. 447–451.

Kastka, J., Buchta, U., Ritterstaedt, R., Paulsen, R., and Mau, U. (**1995a**). "The long term effect of noise protection barriers on the annoyance response of residents," J. Sound Vib. **184**, 823–852.

Kastka, J., and Paulsen, R. (**1979**). "Untersuchung über die subjektive und objektive wirksamkeit von schallschutzeinrichtungen und ihre nebenwirkungen auf die anlieger (A study into the subjective and objective effectiveness of noise barriers and their side effects on residents)," Institut für Hygiene, Universität Düsseldorf.

Klæboe, R., Engelien, E., and Steinnes, M. (**2006**). "Context sensitive noise impact mapping," Appl. Acoust. **67**, 620–642.

Klæboe, R., Kolbenstvedt, M., Lercher, P., and Solberg, S. (**1998**). "Changes in noise reactions—Evidence for an area effect?," in Proceedings of Internoise 1998, Christchurch, New Zealand, pp. 16–18.

Lambert, J., Champelovier, P., and Vernet, I. (**1996**). "Annoyance from high speed train noise: A social survey," J. Sound Vib. **193**, 21–28.

Lambert, J., Champelovier, P., and Vernet, I. (**1998**). "Assessing the railway bonus: The need to examine the 'new infrastructure' effect," in Proceedings of Internoise 1998, Christchurch, New Zealand.

Lambert, J., Champelovier, P., Vernet, I., Annequin, C., and Baez, D. (**1994**). "Community response to high speed train noise in France," in Proceedings of Internoise 1994, Yokohama, pp. 125–128.

Lambert, R. F. (**1978**). "Experimental evaluation of a freeway noise barrier," Noise Control Eng. **11**, 86–94.

Langdon, F. J., and Griffiths, I. D. (**1982**). "Subjective effects of traffic noise exposure, II. Comparisons of noise indices, response scales and the effects of changes in noise levels," J. Sound Vib. **83**, 171–180.

Mackie, M., and Davies, C. H. (**1981**). "Environmental effects of traffic change," TRRL Laboratory Report No. 1015, Transport and Road Research Laboratory, Crowthorne, UK.

Mehra, S. R., and Lutz, C. (**2000**). "Berechnung und subjective wahrnehmung der lärmpegeländerung aufgrund einer neu erstellten umgehungsstraße (Measurement and subjective perception of noise level changes due to a new roadway)," ZfLärmbek **47**, 58–67.

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.

Ministry of Transport, Public Works and Water Management (**2005**). "Evaluatie Schipholbeleid: Schiphol beleefd door omwonenden (Evaluation Schiphol: Schiphol perceived by its residents)," Directorate-General Transport and Aviation.

Mital, A., and Ramakrishnann, A. S. (**1997**). "Effectiveness of noise barriers on an interstate highway: A subjective and objective evaluation," J. Hum. Ergol (Tokyo) **26**, 31–38.

Moehler, U., Hegner, A., Schuemer, R., and Schuemer-Kohrs, A. (**1997**). "Effects of railway-noise reduction on annoyance after rail-grinding, in Proceedings of Internoise 1997, Budapest, Vol. II, pp. 1021–1026.

Nilsson, M. E., and Berglund, B. (**2006**). "Noise annoyance and activity disturbance before and after the erection of a roadside noise barrier," J. Acoust. Soc. Am. **119**, 2178–2188.

Nimura, T., Sone, T., and Kono, S. (**1973**). "Some considerations on noise problem of high-speed railway in Japan," in Proceedings of Internoise 1973, Copenhagen, pp. 298–307.

Öhrström, E. (**1997**). "Community reactions to railway traffic-effects of countermeasures against noise and vibration," in Proceedings of Internoise 1997, Budapest, pp. 1065–1070.

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**, 719–729.

Öhrström, E. and Skånberg, A. (**2000**). "Adverse health effects in relation to noise mitigation—A longitudinal study in the city of Göteborg," in Proceedings of Internoise 2000, Nice, Vol. 4, pp. 2112–2115.

Raw, G. J., and Griffiths, I. D. (**1985**). "The effect of changes in aircraft noise exposure (Letter to the editor)," J. Sound Vib. **101**, 273–275.

Richard, J., and Richter-Richard, H. (**1984**). "Erfahrungen mit dem einsatz verkesrsberuhigender massnahmen zur larmminderung (Experiences from traffic calming measures to reduce noise loads)," ZfLärmbek. **31**, 11–14.

Schreckenberg, D., and Meis, M. (**2007**). "Noise annoyance around an international airport planned to be extended," in Proceedings of Internoise 2007, Istanbul, Turkey.

Schreckenberg, D., Schuemer, R., and Moehler, U. (**2001**). "Railway-noise annoyance and 'misfeasance' under conditions of change," in Proceedings of Internoise 2001, The Hague, The Netherlands, Paper No. 344.

Schuemer, R., and Schreckenberg, D. (**2000**). "Anderung der larmbelastung bei massnahme bedingter stufenweise veranderter gerauschbelastung—Hinweise auf einige befunde und interpretationsansatze (The effect of stepwise change of noise exposure on annoyance)," ZfLärmbek. **47**, 134–143.

Schultz, T. J. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.

Stansfeld, S. A., Haines, M. M., Curtis, S. E., Brentnall, S. L., and Brown, B. (**2001**). *Rapid Review on Noise and Health for London* (St. Bartholmew's and the Royal London School of Medicine and Dentistry, Queen Mary, University of London, London).

Vallet, M. (**1996**). "Annoyance after changes in airport noise environment," in Proceedings of Internoise 1996, Liverpool, UK, Vol. 5, pp. 2329–2335.

Van Dongen, J. E. F., and van den Berg, R. (**1983**). "De gewenning aan het geluid van een nieuwe spoorlijn (Getting used to noise from a new railway line)," Report No. RL-HR-03-02, Ministry of Housing, Spatial Planning and the Environment, The Hague.

Van Kempen, E. E. M. M., and van Kamp, I. (**2005**). "Annoyance from air traffic noise. Possible trends in exposure-response relationships," Report No. 01/2005 MGO Evk, RIVM, Bilthoven, The Netherlands.

Vincent, B., and Champelovier, P. (**1993**). "Changes in the acoustic environment: need for an extensive evaluation of annoyance," in Proceedings Noise and Man 1993, Sixth International Congress on Noise as a Public Health Problem, Vol. 2, pp. 425–428.

Weinstein, N. D. (**1982**). "Community noise problems: Evidence against adaptation," J. Environ. Psychol. **2**, 87–97.

A stochastic model for the noise levels

A. Giménez[a)]
Ciencias Físicas, Matemáticas y de la Computación, Universidad CEU-Cardenal Herrera, C/San Bartolomé 55, Alfara del Patriarca (Valencia) 46115, Spain

M. González
Economia y Empresa, Universidad CEU-Cardenal Herrera, C/San Bartolomé 55, Alfara del Patriarca (Valencia) 46115, Spain

(Received 1 April 2008; revised 5 March 2009; accepted 8 March 2009)

Accurate predictions of environmental noise levels are necessary to implement noise reduction strategies in urban areas. In this paper, a stochastic model is introduced to describe and predict the L_{den}, L_{day}, L_{evening}, and L_{night} levels. A Gaussian Ornstein–Uhlenbeck model is used to represent the dynamics of the noise levels, where the mean-reversion properties and seasonal volatility for each day of the week are studied separately.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3109980]

PACS number(s): 43.50.Rq, 43.60.Cg [BSF] Pages: 3030–3037

I. INTRODUCTION

A chief objective of environmental planning is the reduction in traffic noise levels in urban areas. These noise levels are difficult to predict accurately; however, there have been two approaches to estimate noise levels. In the first approach[1–11] the noise levels are determined by using an established mathematical expression with traffic conditions, such as number of vehicles, vehicle type, speed, and pass frequency as inputs. The second approach[12–16] is based on estimating the statistics of the measured noise levels. The main goal is to obtain statistical distributions and confidence intervals. In these studies the objective is to estimate errors for time periods shorter than the time of measurement, or to obtain the corresponding time estimate for equivalent noise levels of a given duration with a given margin of error. Typically, these time estimates are short (in the order of hours and the sampling time used to obtain the estimate is in the order of minutes). Nevertheless, only a few studies[17–19] have considered the analysis of time series, perhaps due to the high cost and manpower required to measure noise levels for periods of the order of years. In these studies the adjustment of the noise levels was done using autoregressive integrated moving average (ARIMA) models.

In the present article we propose a stochastic model to estimate the seasonal variation and differences in the volatility and reversion speed for each day of the week for long periods of time. Typically, stochastic approaches are used when the underlying physics is unknown or random noise is an important component of the process. They are often used in scientific disciplines such as economics, physics, and the life sciences.

The paper is organized as follows. In Sec. I we present the stochastic model for the daily noise level variations and the proposed Ornstein–Uhlenbeck model is introduced in Sec. II, followed by a discussion of the results obtained for each noise level. Section III is a short discussion of our major conclusions.

II. MODELING NOISE

To find a stochastic process describing the evolution of noise levels we use a noise database of 5 years of noise data gathered from a street in Madrid, Spain. The noise at this location is primarily due to urban traffic. The data consist of hourly noise levels, $L_{\text{eq},1\ \text{h}}$, obtained from January 1, 2001 to December 31, 2005. These data permitted calculation of the following metrics:

(1) L_{day}: L_{eq} between 7:00 am and 7:00 pm,
(2) L_{evening}: L_{eq} between 7:00 pm and 11:00 pm,
(3) L_{night}: L_{eq} between 11:00 pm and 7:00 am, and
(4) L_{den}: according to the formula

$$L_{\text{den}}=10\log\frac{12\times 10^{L_{\text{day}}/10}+4\times 10^{(L_{\text{evening}}+5)/10}+8\times 10^{(L_{\text{night}}+10)/10}}{24}.$$

A. The mean noise level

From the noise level analysis we clearly observe a strong seasonal variation. The daily noise levels (L_{den}, L_{day}, L_{evening}, or L_{night}) seem to vary over the yearly mean noise level during the summers and the winters.[20–23] A first approximation would be to model the seasonal dependence with a sine-function $\sin(\omega t+\varphi)$, where t denotes the time measured in days. Neglecting leap years, we have $\omega=2\pi/365$. The phase angle determines the times of the yearly minimum and maximum noise levels. Moreover, we allow for a weak linear trend in the data. Higher order terms are not included because they would have an even weaker effect on the overall dynamics. We also have included dummy variables in our model in two ways. First, to account for events repeated each year such as bank and religious holidays. Second, to control the presence of outliers in the data such as times when road work was being done. The dummy variable, $S_{i,j,t}$, takes the value 0 or 1 depending on

[a)] Author to whom correspondence should be addressed. Electronic mail: algisan@uch.ceu.es

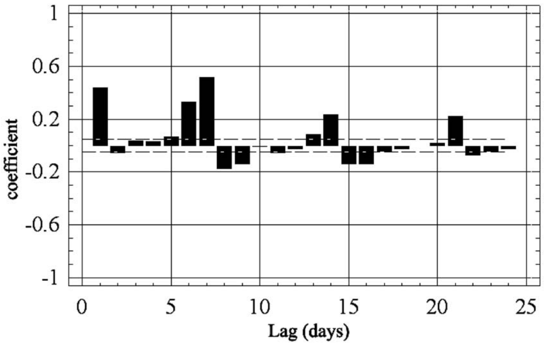


FIG. 1. ACF for L_{day}.

the day selected for each variable. Summing up, a deterministic model for the mean noise levels at time $L_{i,t}^{e}$ (where the index i denotes den, day, evening, or night, depending on the estimated noise metric) would have the form

$$L_{i,t}^{e} = A_i + B_i t + C_i \sin(\omega t + \varphi_i) + \sum_{j=1}^{j=n} \beta_{i,j} S_{i,j,t}, \qquad (1)$$

where the parameters A_i, B_i, C_i, φ_i, and $\beta_{i,j}$ are chosen so that the curve is a good fit to the data. Additionally, to account for possible differences in parameter A between days, we have added additional dummy variables for each day from Tuesday to Sunday ($\beta_{i,\text{T}}$, $\beta_{i,\text{W}}$, $\beta_{i,\text{TH}}$, $\beta_{i,\text{F}}$, $\beta_{i,\text{SA}}$, and $\beta_{i,\text{SU}}$), for the case in which these differences are significant (as discussed below). So, parameter A_i in the deterministic model presented is referenced to Monday. A good model would be one with less dummy variables. The estimation of these parameters is given in Sec. ID. In the following, we suppress index i showing, separately, the parameter values obtained for each noise metric when we discuss results.

B. Driving noise process of noise levels

Noise levels are not deterministic so to obtain a more realistic model we introduce noise into the deterministic model [Eq. (1)], by taking into account the daily volatility, σ_t. One choice is a standard Wiener process,[24] $(W_t, t \geq 0)$, which is a reasonable choice considering the mathematical tractability of the problem. An examination of the data series reveals a weekly autocorrelation in the data (the Monday value depends of the previous Monday, that of Tuesday, of the previous Tuesday, and so on) as can be seen in Fig. 1 for L_{day} data. Therefore, we make the assumption that σ_t is a piecewise constant function, with a constant value during each day, where $(\sigma_\mu)_{\mu=1}^{7}$ are positive constants. Thus, the driving noise process of the noise levels would be $(\sigma_t W_t, t \geq 0)$.

C. Mean-reversion

The noise levels cannot increase day after day for a long period of time. This means that our model should not allow the noise levels to deviate from their mean values for more than short periods of time, i.e., the predicted noise levels should have a "mean-reverting" property. Combining all the assumptions together, we model the noise levels by the following stochastic differential equation (SDE):

$$dL_t = a[L_t^e - L_t]dt + \sigma_t dW_t, \qquad (2)$$

where the real drift coefficient, a, determines the speed of the mean-reversion. The solution of such an equation is usually called an Ornstein–Uhlenbeck process.[25,26]

Equation (2) is the one that does not actually converge to L_t^e for long periods of time (see, for example, Refs. 27–31), and an extra deterministic term needs to be added to the drift term to obtain a process that really converges to Eq. (1)

$$\frac{dL_t^e}{dt} = Bt + \omega C \sin(\omega t + \varphi). \qquad (3)$$

This extra term adjusts the drift so that the solution of the SDE tends to L_t^e (as the noise level L_t^e is not constant). Starting at $L_s = x$, we now get the following model for the noise levels:

$$dL_t = \left\{ \frac{dL_t^e}{dt} + a[L_t^e - L_t] \right\} dt + \sigma_t dW_t, \quad t > s. \qquad (4)$$

The solution is

$$L_t = (x - L_s^e) e^{-a(t-s)} + L_t^e + \int_s^t e^{-a(t-x)} \sigma_x dW_x, \qquad (5)$$

where L_t^e is given by Eq. (1).

D. Parameter estimation

In this section we estimate the unknown parameters A, B, C, φ, β_j, σ, and a in Eqs. (1) and (2). First, to find numerical values of the constants in the deterministic model (1) we fit the function

$$Z_t = b_0 + b_1 t + b_2 \sin(\omega t) + b_3 \cos(\omega t) + \beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_n S_{n,t} \qquad (6)$$

to the noise level data using the method of least-squares. This means that we have to find the parameter vector $\xi = [b_0, b_1, b_2, b_3, \beta_1, \beta_2, \ldots, \beta_n]$ such that

$$\min_{\xi} \|Z - X\|^2, \qquad (7)$$

where Z is the vector with elements [Eq. (6)] and X is the data vector. The constants in Eq. (1) are then obtained from the relations

$$A = b_0,$$

$$B = b_1,$$

$$C = \sqrt{b_2^2 + b_3^2},$$

$$\varphi = \arctan\left(\frac{b_3}{b_2}\right),$$

TABLE I. Deterministic model for noise levels.

Level	Deterministic model (dB)	Dummy variables (dB)		
		β_1	β_2	R^2
L_{den}	$74.10-0.000\,28t+0.8089\sin(\omega t+1.24)$	−0.27	−2.33	0.7397
L_{day}	$71.12-0.000\,35t+0.8544\sin(\omega t+1.33)$	−2.59	−2.50	0.8027
L_{evening}	$69.71-0.000\,27t+0.8282\sin(\omega t+1.17)$	−1.03	−2.47	0.6206
L_{night}	$66.28-0.000\,27t+0.7843\sin(\omega t+1.22)$	0.61	−2.25	0.7977

$$\beta_j = \beta_j. \tag{8}$$

Second, we use two methods to obtain the volatility σ_μ from the daily data. The first method is based on the quadratic variation of the observed noise levels during the day, L_j (see, e.g., Ref. 32),

$$\sigma_\mu^2 = \frac{1}{N_\mu}\sum_{j=0}^{N_\mu}(L_{j+1}-L_j)^2, \tag{9}$$

where N_μ represents the total number of days.

The second method is based on discretizing Eq. (4) and thinking of the resultant equation as a regression equation. Indeed, during a given day μ, the discretized equation can be written as

$$L_j = L_j^e - L_{j-1}^e + a_\mu L_{j-1}^e + (1-a_\mu)L_j + \sigma_\mu \varepsilon_j, \quad j=1,\ldots,N_\mu, \tag{10}$$

where $\{\varepsilon_j\}_{j=1}^{N_\mu-1}$ is an independent and identically distributed variable.

With $\hat{L}_j = L_j - (L_j^e - L_{j-1}^e)$ we can write Eq. (10) as

$$\hat{L}_j = a_\mu L_{j-1}^e + (1-a_\mu)L_{j-1} + \sigma_\mu \varepsilon_j, \tag{11}$$

which can be seen as a regression of today's noise level on last week's noise level.

Thus, an efficient estimator of σ_μ is (see, e.g., Ref. 33)

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu - 2}\sum_{j=1}^{N_\mu}[\hat{L}_j - \hat{a}_\mu L_{j-1}^e + (1-\hat{a}_\mu)L_{j-1}]^2. \tag{12}$$

So, in method 2, we must first calculate the mean-reversion parameter, a_μ, to find the value of volatility, σ_μ. For the calculation of the mean-reversion parameter of method 2, we use the method of least-squares to fit a linear regression of Eq. (11).

Finally, we estimate the speed of reversion for method 1. Since the time between observations of the noise levels is obviously bounded away from zero, it is appropriate to estimate the mean-reversion parameter by using the martingale estimation function method suggested in Ref. 34. Based on observations collected during N_μ days, an efficient estimator of a_μ is obtained as a zero of the equation $G_N(a_\mu)=0$, where

$$G_N(a_\mu) = \sum_{j=1}^{N_\mu}\frac{b'(L_{j-1};a_\mu)}{\sigma_{\mu,j-1}^2}[L_j - E(L|L_{j-1})],$$

where b' denotes the derivative with respect to a_μ of the drift term in Eq. (4). Then, we only have to determine each of the terms $[L_j - E(L|L_{j-1})]$ by solving Eq. (11). Therefore

$$G_N(a_\mu) = \sum_{j=1}^{N_\mu}\frac{L_j^e - L_{j-1}}{\sigma_{\mu,j-1}^2}[L_j - (L_{j-1}-L_{j-1}^e)e^{-a_\mu} - L_j^e], \tag{13}$$

from which it is easily verified that

$$a_\mu = -\log\left(\frac{\sum_{j=1}^{N_\mu}\frac{L_{j-1}^e - L_{j-1}}{\sigma_{\mu,j-1}^2}(L_j - L_j^e)}{\sum_{j=1}^{N_\mu}\frac{L_{j-1}^e - L_{j-1}}{\sigma_{\mu,j-1}^2}(L_{j-1} - L_{j-1}^e)}\right). \tag{14}$$

Inserting in Eq. (14) the numerical values obtained for volatility by Eq. (9), we get the mean-reversion parameter for method 1.

III. RESULTS AND DISCUSSION

The results of the deterministic model, for each metric, are shown in Table I. These take Monday as the reference day for parameter A. The dummy variables β_1 and β_2 represent bank holidays and Christmas, respectively. For parameter A, the deterministic model was found to depend on the day selected, especially if the selected day is a Saturday or Sunday, and also depending on the selected noise level there are larger or smaller differences between days. For example, for L_{day} and L_{evening}, there is a significant difference in A between weekends and weekdays but not between weekdays. However, for L_{night} and L_{den}, there are significant differences in A for all days of the week with the exception of Monday and Tuesday, as can be seen in Table II. However, these

TABLE II. Difference in parameter A for each of the days referring to Monday's level.

Day	L_{den}	t-value	L_{day}	t-value	L_{evening}	t-value	L_{night}	t-value
				Parameter A (dB)				
Monday	74.10	1900	71.12	1403	69.71	1331	66.28	1440
				Difference with respect to Monday (dB)				
Tuesday, β_{T}	−0.06	−1.39	−0.06	−1.10	+0.03	+0.48	−0.09	−1.76
Wednesday, β_{W}	+0.19	+4.26	−0.04	−0.65	−0.01	−0.17	+0.39	+7.25
Thursday, β_{TH}	+0.35	+7.76	−0.08	−1.30	−0.06	−0.98	+0.68	+12.6
Friday, β_{F}	+0.88	+19.2	−0.19	−3.24	−0.04	−0.586	+1.54	+28.5
Saturday, β_{SA}	+1.01	+22.0	−1.82	−30.5	−0.66	−10.7	+2.23	+41.5
Sunday, β_{SU}	+0.91	+19.9	−2.72	−45.4	−0.71	−11.5	+2.28	+42.0

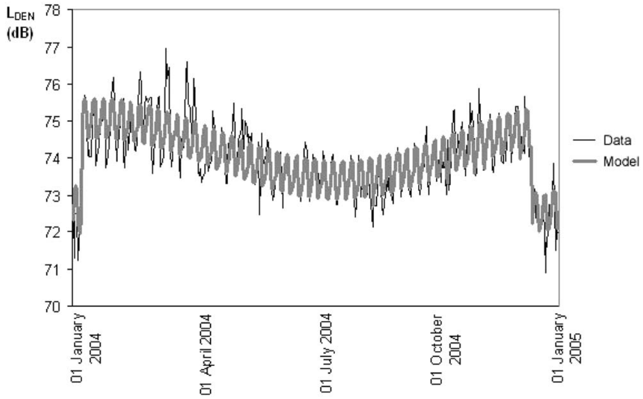


FIG. 2. Deterministic model over real data for L_{den} in year 2004.

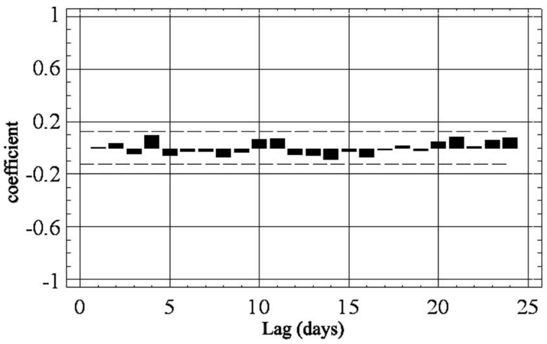


FIG. 3. Partial ACF for residuals of L_{den}.

differences are small, and in a practical sense, they appear to be insignificant. In general, the change is largest on the weekend. For the rest of the parameters B, C, and φ, we observe that the metric parameters are close to one another for each of the metrics. Finally, with respect to the dummy variables, we have found that the effect of bank holidays, β_1, differs depending on the metric: For L_{den}, L_{day}, and L_{evening} the dummy variable takes positive values while for L_{night} the dummy variable takes negative values. However, for holidays such as Christmas the dummy variable, β_2, takes similar values for each metric. Taking these considerations into account, the adjustment of the deterministic model for the L_{den} value in year 2004 is shown in Fig. 2. For volatility and speed of reversion, the values obtained are given by Tables III–VI for each metric, respectively. We observe that the L_{den} metric exhibits smaller differences between methods 1 and 2 in comparison to the other three metrics. It is also shown that method 2 presents less variation in volatility and speed of reversion from one day to another than method 1.

Taking into account all noise metrics, by method 1, volatility varies between 0.7 and 1.5 dB, and the difference between maximum volatility and minimum volatility is less than 0.6 dB. Referring to speed of reversion, values range (excluding Wednesday, which exhibits a higher speed of reversion) from 1.3 to 2.9 day^{-1}. By method 2, volatility ranges from 0.45 up to 0.8 dB, and the variation between metrics for any single day of the week is less than 0.2 dB. The speed of reversion presents ranges from 0.7 to 1 day^{-1} (including Wednesday) with a variation that is less than 0.2 day^{-1} when comparing between metrics for any single day of the week. With these results, one can take methods 1 and 2 to estimate the upper and lower limits, respectively, of the noise metrics.

To check the validity of the model we have processed the behavior of residuals. In Fig. 3 we observe the autocorrelation function (ACF) of residuals obtained from the model for L_{den}. It shows that all autocorrelations in the ACF plot of residuals are insignificant, which implies that the residuals are statistically independent. For all other metrics the ACF of residuals exhibits similar results as is shown in Fig. 3 for L_{den}. In Table VII, the residuals for L_{evening} are listed, and again they indicate that the model is adequate. A distribution fit to the residuals seems to be elliptical. We have used a generalized error distribution where a tail index of 1 indicates a Laplace distribution and a tail index of 2 indicates a normal distribution. In general, for all metrics, the tail index of the distribution is around 1.35.

Finally, Fig. 4 shows the predicted and actual L_{den} for 5000 simulations for the period from January 1, 2006 to February 12, 2006 (generalized error distribution with tail index of 1.35; each simulation for the overall time period). This figure shows that there are only small differences between actual and predicted L_{den}. The figure also shows the 1% and 99% (lower and upper) predictions for L_{den}.

IV. CONCLUSIONS

A novel model for describing and predicting the noise metrics L_{den}, L_{day}, L_{evening}, and L_{night} has been presented. A Gaussian Ornstein–Uhlenbeck model is used to describe the

TABLE III. Volatility (in dB) and speed of reversion (in day^{-1}) for level L_{den}.

	Volatility		Speed of reversion		Mean value	
Day	Method 1	Method 2	Method 1	Method 2	Volatility	Speed reversion
Monday	0.947	0.468	1.659	0.813	0.708	1.236
Tuesday	0.728	0.469	1.502	0.781	0.598	1.141
Wednesday	0.842	0.476	2.170	0.886	0.659	1.528
Thursday	1.015	0.489	1.655	0.809	0.752	1.232
Friday	0.859	0.532	1.530	0.784	0.696	1.157
Saturday	0.830	0.540	1.463	0.768	0.685	1.116
Sunday	0.987	0.569	1.442	0.764	0.778	1.103
Mean	0.887	0.506	1.631	0.801	0.696	1.216

TABLE IV. Volatility (in dB) and speed of reversion (in day^{-1}) for level L_{day}.

	Volatility		Speed of reversion		Mean value	
Day	Method 1	Method 2	Method 1	Method 2	Volatility	Speed reversion
Monday	1.283	0.621	1.939	0.858	0.952	1.399
Tuesday	1.248	0.612	1.457	0.769	0.930	1.113
Wednesday	1.389	0.624	4.372	1.013	1.007	2.692
Thursday	1.509	0.563	1.885	0.848	1.036	1.367
Friday	1.250	0.624	1.602	0.799	0.937	1.200
Saturday	1.019	0.715	1.621	0.802	0.867	1.212
Sunday	1.224	0.819	1.321	0.733	1.022	1.027
Mean	1.275	0.654	2.028	0.832	0.964	1.430

stochastic dynamics of the noise metrics based on empirical results. The deterministic model is based on a constant, A, with a linear trend in addition to a sine-function that describes seasonal differences. The main differences found for the deterministic model that compares the noise metrics L_{den}, L_{day}, L_{evening}, and L_{night} are in the constant, A, that depends on the day of the week. It is also shown that the random component includes volatility and mean-reversion properties. Mean volatility varies from about 0.7 dB to about 1 dB (depending on the noise metric), with only small differences between metrics for a single given day of the week. Mean-reversion varies from about 1 to 2 day^{-1}. The simulated data for the time period from January 1, 2006 to February 12, 2006 (5000 paths) show that the real data are close to the simulated data and in no case are the real data outside of the 1% and 99% percentiles for any of the four noise metrics.

APPENDIX A: WIENER PROCESS AND ORNSTEIN–UHLENBECK PROCESS

The Wiener process W_t, often called Brownian motion, is a continuous-time stochastic process characterized by three facts.

(1) $W_0=0$.
(2) W_t is almost surely continuous.
(3) W_t has independent increments with distribution $W_t-W_s \sim N(0,t-s)$ (for $0\leq s<t$).

$N(\mu,\sigma^2)$ denotes the normal distribution with expected value μ and variance σ^2. The condition that it has independent increments means that if $0\leq s_1\leq t_1\leq s_2\leq t_2$ then $W_{t_1}-W_{s_1}$ and $W_{t_2}-W_{s_2}$ are independent random variables.

The unconditional probability density function of a one dimensional Wiener process at a fixed time t is given by

$$f_{W_t}(x;t)=\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}$$

with expectation zero, $E(W_t)=\mu_W=0$, and covariance and correlation, $R(t_1,t_2)=K(t_1,t_2)=\min\{t_1,t_2\}$.

Its derivation follows immediately from the definition that W_t (at a fixed time t) is normally distributed $W_t-W_0=W_t\sim N(0,t)$.

Derivation of the last is (suppose $t_1<t_2$)

$$R(t_1,t_2)=E[(W_{t_1}-E(W_{t_1}))(W_{t_2}-E(W_{t_2}))]=E[W_{t_1}W_{t_2}].$$

Then add and subtract W_{t_1},

$$E[W_{t_1}W_{t_2}]=E[W_{t_1}(W_{t_2}-W_{t_1}+W_{t_1})]$$
$$=E[W_{t_1}(W_{t_2}-W_{t_1})]+E[W_{t_1}]^2.$$

Since $W_{t_1}=W_{t_1}-W_{t_0}$ and $W_{t_2}=W_{t_2}-W_{t_0}$ are independent,

$$E[W_{t_1}(W_{t_2}-W_{t_1})]=E[W_{t_1}]E[W_{t_2}-W_{t_1}]=0.$$

Because of that we have $R(t_1,t_2)=E[W_{t_1}]^2=t_1$.

The Ornstein–Uhlenbeck process (r_t), also known as the mean-reverting process, is a stochastic process. [It is the continuous-time analog of the discrete-time AR(1) process,

TABLE V. Volatility (in dB) and speed of reversion (in day^{-1}) for level L_{evening}.

	Volatility		Speed of reversion		Mean value	
Day	Method 1	Method 2	Method 1	Method 2	Volatility	Speed reversion
Monday	1.166	0.666	2.584	0.924	0.916	1.754
Tuesday	1.073	0.613	2.690	1.068	0.843	1.879
Wednesday	1.154	0.707	4.143	1.016	0.930	2.579
Thursday	1.157	0.651	2.830	0.941	0.904	1.885
Friday	1.415	0.700	2.506	0.918	1.057	1.712
Saturday	1.407	0.745	2.855	0.942	1.076	1.899
Sunday	1.247	0.744	1.839	0.841	0.996	1.340
Mean	1.232	0.690	2.778	0.950	0.961	1.864

TABLE VI. Volatility (in dB) and speed of reversion (in day^{-1}) for level L_{night}.

	Volatility		Speed of reversion		Mean value	
Day	Method 1	Method 2	Method 1	Method 2	Volatility	Speed reversion
Monday	1.184	0.545	1.859	0.851	0.864	1.355
Tuesday	0.853	0.537	1.306	0.736	0.695	1.021
Wednesday	0.971	0.620	2.335	0.903	0.795	1.619
Thursday	1.180	0.596	1.453	0.766	0.888	1.109
Friday	0.941	0.648	1.761	0.828	0.795	1.295
Saturday	0.886	0.627	1.765	0.829	0.756	1.297
Sunday	1.149	0.628	1.446	0.764	0.888	1.105
Mean	1.023	0.600	1.704	0.811	0.812	1.257

and in contrast to the Wiener process admits a stationary probability distribution.] It is given by the stochastic differential equation

$$dr_t = -\theta(r_t - \mu)dt + \sigma dW_t,$$

where θ, μ, and σ are parameters and W_t denotes the Wiener process.

The equation is solved by variation of parameters. Apply Ito's lemma to the function $f(r_t,t)=r_t e^{\theta t}$ to get

$$df(r_t,t) = \theta r_t e^{\theta t}dt + e^{\theta t}dr_t = e^{\theta t}\theta\mu dt + \sigma e^{\theta t}dW_t.$$

Integrating from 0 to t, we obtain $r_t e^{\theta t}=r_0+\int_0^t e^{\theta s}\theta\mu ds+\int_0^t \sigma e^{\theta s}dW_s$ whereupon we see $r_t=r_0e^{-\theta t}+\mu(1-e^{-\theta t})+\int_0^t \sigma e^{\theta(s-t)}dW_s$. Thus, the first moment is given by (assuming that r_0 is a constant)

$$E(r_t) = r_0 e^{-\theta t} + \mu(1-e^{-\theta t}).$$

Denote $s\wedge t=\min(s,t)$, we can use the Ito isometry to calculate the covariance function by

$$\begin{aligned}\operatorname{cov}(r_s,r_t) &= E[(r_s - E[r_s])(r_t - E[r_t])]\\ &= E\left[\int_0^s \sigma e^{\theta(u-s)}dW_u \int_0^t \sigma e^{\theta(v-t)}dW_v\right]\\ &= \sigma^2 e^{-\theta(s+t)}E\left[\int_0^s e^{\theta u}dW_u \int_0^t e^{\theta v}dW_v\right]\\ &= \frac{\sigma^2}{2\theta}e^{-\theta(s+t)}(e^{2\theta(s\wedge t)}-1).\end{aligned}$$

It is also possible (and often convenient) to represent r_t (unconditionally) as a scaled time-transformed Wiener process

$$r_t = \mu + \frac{\sigma}{\sqrt{2\theta}}W(e^{2\theta t})e^{-\theta t}$$

or conditionally (given r_0) as

$$r_t = r_0 e^{-\theta t} + \mu(1-e^{-\theta t}) + \frac{\sigma}{\sqrt{2\theta}}W(e^{2\theta t}-1)e^{-\theta t}.$$

TABLE VII. Results for residuals test of L_{evening} noise level. In brackets p-value. (AR 1–2: autoregressive test order 2; ARCH 1–1: autoregressive heteroskedasticity test; Normality: test for normality distribution; Hetero and Hetero-x: heteroskedasticity tests; and Reset: regression specification test. These tests are calculated with commercial software PCGIVE.)

	Tests					
	AR 1–2	ARCH 1–1	Normality	Hetero	Hetero-x	Reset
Monday	0.0423 [0.9586]	1.6597 [0.1988]	59.315 [0.0000]	0.0723 [0.9302]	0.0723 [0.9302]	2.961 [0.0865]
Tuesday	0.82138 [0.4410]	0.83036 [0.3630]	13.330 [0.0013]	0.83323 [0.4358]	0.83323 [0.4358]	0.72473 [0.3954]
Wednesday	0.088847 [0.9150]	16.901 [0.0001]	8.3255 [0.0156]	3.4178 [0.0343]	3.4178 [0.0343]	0.17719 [0.6741]
Thursday	0.96666 [0.3817]	0.012489 [0.9111]	7.9405 [0.0189]	0.53652 [0.5854]	0.53652 [0.5854]	0.44086 [0.5073]
Friday	1.8709 [0.1561]	0.20562 [0.6506]	35.862 [0.0000]	0.41558 [0.6604]	0.41558 [0.6604]	0.19662 [0.6578]
Saturday	1.1397 [0.3215]	0.022553 [0.8807]	45.966 [0.0000]	0.41016 [0.6640]	0.41016 [0.6640]	0.042171 [0.8375]
Sunday	1.0298 [0.4011]	0.095590 [0.7574]	39.788 [0.0000]	0.89178 [0.4112]	0.89178 [0.4112]	0.22564 [0.6352]

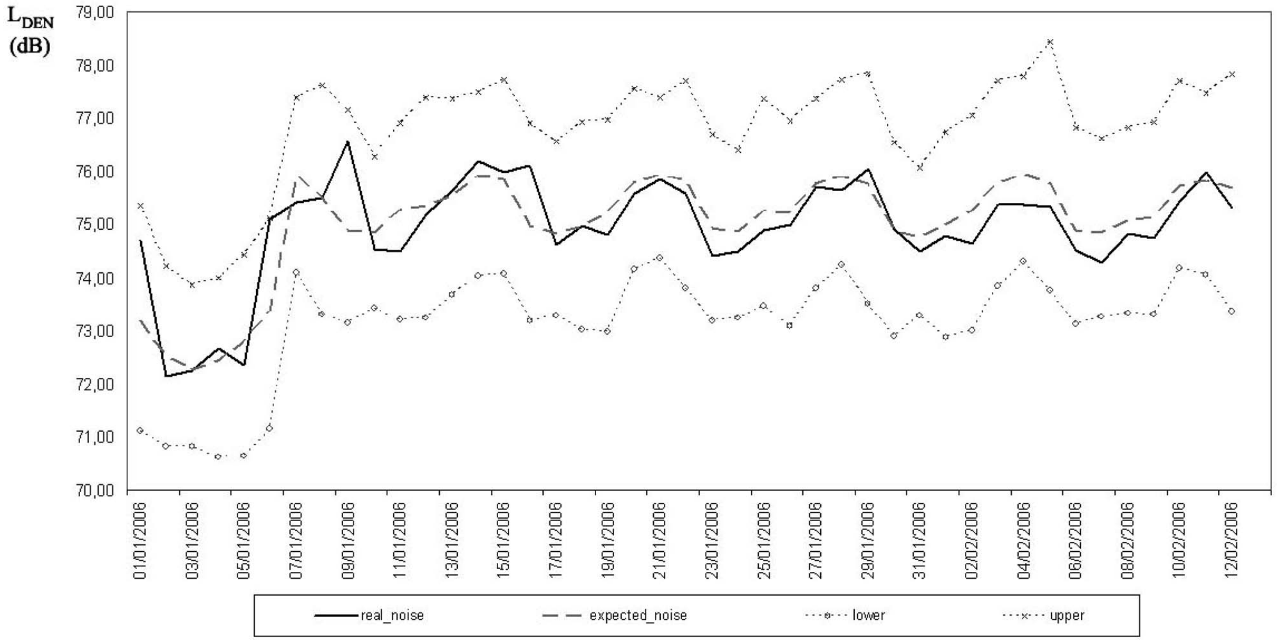


FIG. 4. Simulation and confidence intervals for noise level L_{den}.

APPENDIX B: CALCULATION PROCEDURE

We describe here the procedures followed to obtain the results given in the tables in the body of the paper (the same procedure for each noise level separately).

We have a data vector that contains the daily noise level for 5 years, D_t ($t = 1, 2, 3, \ldots, 1827$ data).

With the help of a computer program we calculate the parameter vector $\xi = [b_0, b_1, b_2, b_3, \beta_1, \beta_2, \ldots, \beta_n]$ using the method of least-squares [Eq. (7)] and then we calculate the values of parameters of Eq. (1), using Eq. (8). (With this procedure we have obtained the deterministic model, see Tables I and II.)

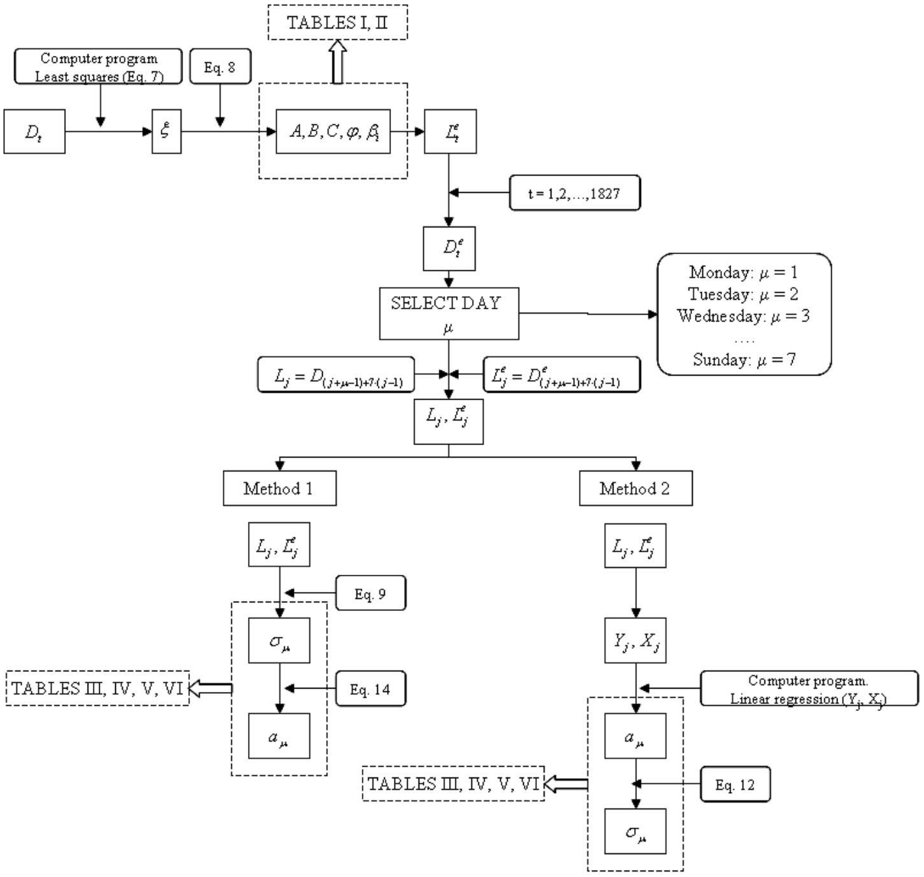


FIG. 5. Scheme of calculation procedure.

We then calculate the volatility and mean-reversion parameters. With the deterministic model and giving values $t=1,2,\ldots,1827$ to L_t^e we determine the data series D_t^e ($t=1,2,3,\ldots,1827$) that contains the estimated deterministic data.

At this point we will determine, for each day, μ, separately, the data series L_j and L_j^e,

$$L_j = D_{(j+\mu-1)+7\times(j-1)}, \quad L_j^e = D^e_{(j+\mu-1)+7\times(j-1)}.$$

Using method 1, and the data series L_j and L_j^e by Eq. (9) we obtain the volatility σ_μ, and with this value in Eq. (14) we obtain the mean-reversion parameter a_μ.

Using method 2, we calculate two data variables

$$Y_j = (L_j - L_j^e) - (L_{j-1} - L_{j-1}^e) \quad \text{and} \quad X_j = -(L_{j-1} - L_{j-1}^e).$$

Using computer programs we calculate the linear regression between variables (Y_j, X_j) that gives the mean-reversion parameter of Eq. (11). This regression is the equation

$$Y_j = a_\mu X_j + u_j \rightarrow (u_j = \sigma_j \varepsilon_j).$$

Finally we use the value a_μ in Eq. (12) to obtain the volatility σ_μ for method 2.

This procedure can also be observed more clearly in the scheme showed in Fig. 5.

[1] C. Steele, "A critical review of some traffic noise prediction models," Appl. Acoust. **62**, 271–287 (2001).

[2] Anon., "Handbook of acoustic noise control," WADC Technical Report No. 52–204, Wright Air Development Center, Air Research and Development Command, United States Air Force (Wright-Patterson Air Force Base), Ohio, 1953.

[3] A. F. Nickson, "Can community reaction to increased traffic noise be forecast?," in Proceedings of the Fifth International Congress on Acoustics (1965).

[4] C. Lamure, "Niveaux de bruit au voisinage des autoroutes (Noise levels in the vicinity of highways)," in Proceedings of the Fifth International Congress on Acoustics (1965).

[5] D. R. Johnson and E. G. Saunders, "The evaluation of noise from freely flowing road traffic," J. Sound Vib. **7**, 287–309 (1968).

[6] M. E. Delany, D. G. Harland, R. A. Hood, and W. E. Scholes, "The prediction of noise levels L_{10} due to road traffic," J. Sound Vib. **48**, 305–325 (1976).

[7] M. Koyasu, "Method of prediction and control of road traffic noise in Japan," Internoise 78, San Francisco, CA (1978).

[8] H. Tachibana and M. Sasaki, "ASJ prediction methods of road traffic noise," Internoise 94, Yokohama, Japan (1994).

[9] K. Takagi and K. Yamamoto, "Calculation methods for road traffic noise propagation proposed by ASJ," Internoise 94, Yokohama, Japan (1994).

[10] T. Suksaard, P. Sukasem, S. M. Tabucanon, I. Aoi, K. Shirai, and H. Tanaka, "Road traffic noise prediction model in Thailand," Appl. Acoust. **58**, 123–130 (1999).

[11] A. Calixto, F. B. Diniz, and P. H. T. Zannin, "The statistical modeling of road traffic noise in an urban setting," Cities **20**, 23–29 (2003).

[12] K. W. Yeow, N. Popplewell, and J. F. W. Mackay, "Method of predicting L_{eq} created by urban traffic," J. Sound Vib. **53**, 103–109 (1977).

[13] D. Skarlatos and P. Drakatos, "On selecting the minimum observation time for determining the L_{eq} of a random noise with a given level of confidence," J. Sound Vib. **152**, 141–148 (1992).

[14] E. Manatakis and D. Skarlatos, "A statistical model for evaluation and prediction of the noise exposure in a construction equipment area," Appl. Acoust. **63**, 759–773 (2002).

[15] S. Yamaguchi, T. Saeki, and Y. Kato, "A fundamental consideration on estimating L_{eq} of specific noise under the existence of background noise," Appl. Acoust. **55**, 165–180 (1998).

[16] F. A. Farrelly and G. Brambilla, "Determination of uncertainty in environmental noise measurements by bootstrap method," J. Sound Vib. **268**, 167–175 (2003).

[17] P. D. Schomer and R. E. Devor, "Temporal sampling requirements for estimation of long-term average sound levels in the vicinity of airports," J. Acoust. Soc. Am. **69**, 713–719 (1981).

[18] R. E. Devor, P. D. Schomer, W. A. Kline, and R. D. Neathamer, "Development of temporal sampling strategies for monitoring noise," J. Acoust. Soc. Am. **66**, 763–771 (1979).

[19] K. Kumar and V. K. Jain, "Autoregressive integrated moving averages (ARIMA) modelling of a traffic noise time series," Appl. Acoust. **58**, 283–294 (1999).

[20] L. C. Fothergill, "The variation of environmental noise outside six dwellings between three seasons," Appl. Acoust. **10**, 191–200 (1977).

[21] J. B. Large, "Effects of weather conditions on airport noise prediction," Internoise 86, Cambridge, MA (1986).

[22] E. A. Björk, "Community noise in different seasons in Kuopio, Finland," Appl. Acoust. **42**, 137–150 (1994).

[23] J. Romeu, S. Jimenez, T. Pamies, and M. Genesca, "L_{DEN} assessment methodology for acoustic maps: Simulation or measurements?," Internoise 2003, Seogwipo, Korea (2003).

[24] B. Øksendal, *Stochastic Differential Equations*, 6th ed. (Springer, New York, 2003).

[25] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of Brownian motion," Phys. Rev. **36**, 823–841 (1930).

[26] D. T. Gillespie, "Exact numerical simulation of the Ornstein-Uhlenbek process and its integral," Phys. Rev. E **54**, 2084–2091 (1996).

[27] F. Dornier and M. Queruel, "Caution to the wind," Energy Power and Risk Management, Weather Risk Report, **8**, 30–32 (2000).

[28] M. H. A. Davis, "Pricing weather derivatives by marginal value," Quant. Finance **1**, 305–308 (2001).

[29] D. C. Brody, J. Syroka, and M. Zervos, "Dynamical pricing of weather derivatives," Quant. Finance **2**, 189–198 (2002).

[30] P. Alaton, B. Djehiche, and D. Stillberger, "On modelling and pricing weather derivatives," Appl. Math. Finance **9**, 1–20 (2002).

[31] F. E. Benth, "On arbitrage-free pricing of weather derivatives based on fractional Brownian Motion," Appl. Math. Finance **10**, 303–324 (2003).

[32] I. V. Basawa and B. L. S. Prasaka Rao, *Statistical Inference for Stochastic Processes* (Academic, New York, 1980), pp. 212–213.

[33] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. (Springer, New York, 1990).

[34] B. M. Bibby and M. Sørensen, "Martingale estimation functions for discretely observed diffusion processes," Bernoulli **1**, 17–39 (1995).

Some spatial and temporal effects on the speech privacy of meeting rooms

John S. Bradley,[a] Marina Apfel, and Bradford N. Gover
National Research Council, 1200 Montreal Road, Ottawa K1A 0R6, Canada

(Received 16 December 2008; revised 19 February 2009; accepted 19 February 2009)

This paper reports on initial experiments concerning how key spatial and temporal effects in rooms influence the speech privacy provided by enclosed rooms. The first part of the work demonstrates that for the same signal-to-noise ratio, the intelligibility of speech and the threshold of intelligibility are significantly different for transmission between real rooms than in the previous results in approximately free-field conditions [B. N. Gover and J. S. Bradley, J. Acoust. Soc. Am. **116**, 3480–3490 (2004)]. The second part investigates the influence of aspects of the spatial and temporal components of sound fields in typical rooms, to explain these differences for transmission between real rooms. These components included the separate effects of early-arriving and later-arriving reflected speech sounds. They also included the effects of spatially separated speech and noise sources as well as more diffuse noise representative of typical meeting rooms. In realistic combinations these effects are of practical importance and can change privacy criteria by 5 dB or more. Ignoring them could lead to costly over-design of the sound insulation required to achieve adequate speech privacy. [DOI: 10.1121/1.3097771]

PACS number(s): 43.55.Hy [NX] Pages: 3038–3051

I. INTRODUCTION

Most previous work has assumed that speech privacy is only influenced by the signal-to-noise ratio of the speech and the concurrent ambient noise. The primary importance of signal-to-noise ratios (SNRs), as determinants of speech privacy, was established by the pioneering work of Cavanaugh *et al.*[1] on speech privacy of enclosed offices, which related privacy ratings to articulation index values. A more recent study[2] found a uniformly weighted signal-to-noise ratio measure ($\mathrm{SNR}_{\mathrm{uni32}}$) to be a good predictor of ratings of both the audibility and intelligibility of speech sounds from an adjacent room. Although it is well known that the intelligibility of speech can be reduced by reverberation[3] and increased by the spatial separation of speech and noise sources,[4] these effects have not previously been considered in studies of the speech privacy of enclosed rooms.

This new work is an initial investigation of the key spatial and temporal effects in rooms on the audibility and intelligibility of speech in noise for conditions representative of typical rooms where speech privacy is required. The new work includes two studies. The first more exploratory study (Sec. III) demonstrates that the intelligibility of speech and the threshold of intelligibility are significantly different for propagation between real rooms than in the previous results for simulated conditions in approximately free-field conditions.[2] The second part of the work (Sec. IV) investigates the magnitude of the effects of each aspect of the spatial and temporal components of sound fields in typical rooms. These included the separate effects of early-arriving and later-arriving reflected sounds. They also included the effects of spatially separated speech and noise sources as well as more diffuse noise representative of typical meeting rooms.

There have been many previous studies related to understanding the temporal and spatial effects of room acoustics on our ability to understand speech heard in combination with competing sounds. It is often considered that reduced speech intelligibility corresponds to increased speech privacy, but in some situations higher degrees of privacy might require speech to be inaudible. Classical room acoustics studies have long identified optimum reverberation times for maximizing the intelligibility of speech.[3] Earlier work to explain spatial effects was concerned with explaining our ability to understand speech in the presence of other interfering speech sounds, the so-called *cocktail party effect*.[5] There have been at least two reviews of the many studies related to the cocktail party effect.[6,7]

Interfering sounds can mask the target speech sounds (that we wish to hear) and reduce the intelligibility of the speech. The intelligibility of speech is first a SNR issue and the work of French and Steinberg,[8] which led to the articulation index, can explain the monaural signal-to-noise effects on speech intelligibility. Masking is influenced by both monaural and binaural effects. Even for monaural listening, head shadow effects can influence the intelligibility of speech as a function of the relative directions of the target speech and interfering sounds.

It is usually possible to better understand the target speech mixed with interfering sounds by listening binaurally. The benefit of listening binaurally rather than monaurally is referred to as a *binaural advantage*. Many studies have tried to explain the cause of binaural advantages and have shown interaural level, time, and phase differences to be important.[9] These interaural differences vary with the direction of the sound source relative to the head of the listener and hence

[a] Author to whom correspondence should be addressed. Electronic mail: john.bradley@nrc-cnrc.gc.ca

can help us to discriminate among spatially separated sound sources.

Although many of the earlier studies focused on the effect of interfering speech sounds on the target speech, interfering noises can lead to larger reductions in intelligibility. For example, with a single interfering talker, it is possible for listeners to hear the target speech in the gaps of the interfering speech. This is not possible with more or less constant noises such as ventilation type noise common in buildings. When the interfering speech is made up of a number of talkers, the masking effect of the speech tends to be similar to that of noise with a similar spectrum and level.[10]

The masking effect of an interfering sound is greatly influenced by the direction of arrival of the masking sound relative to that of the target speech. Experiments in free-field conditions have shown that separating the target speech and masking sound by as little as 10° is detectable, and that a 20° difference leads to quite significant increases in the intelligibility of the target speech.[4] Systematic studies have reported the resulting increased intelligibility as a function of the angle of separating the speech and noise sources.[6,11] This reduction in the masking effect of the interfering noise by spatially separating the noise and the speech sources is referred to as a *spatial release from masking*.

Studies of spatial effects have mostly been conducted in free-field conditions and have not often included the effects of reverberation. Where reverberation has been included, it has been shown to reduce the magnitude of the spatial release from masking,[12] indicating that in typical rooms with reverberant sound, listeners are less able to benefit from a spatial release from masking when the target speech source and the interfering sound source are spatially separated. Plomp[13] systematically investigated the combined effects of varied reverberation time with varied separation of the speech and interfering noise sources. Although based on subjective ratings of the intelligibility rather than on intelligibility test scores, the study showed a gradual decrease in the spatial release from masking as reverberation time was increased.

Most of the studies to date have focused on understanding individual parts of the overall issue of spatial and temporal effects of room acoustics and have most often been carried out in free-field conditions. Only a few studies have included the effects of room reverberation and usually the term reverberation has been used loosely to include all types of reflected sound. Most often the interfering signal has been speech[14,15] and not typical room noises such as that from ventilation systems.

II. COMMON EXPERIMENTAL PROCEDURES FOR EXPERIMENTS

This paper describes two different studies to investigate the effects of room acoustics on the speech privacy of meeting rooms. The first investigation was intended to explore how closely the results of previous tests in approximately free-field conditions[2] could be replicated in more realistic acoustical conditions. In these validation tests subjects listened to speech transmitted through real walls from an adjacent room. The second series of tests was carried out in simulated sound fields to determine the effects of various details of the spatial and temporal characteristics of speech and noise sounds on the intelligibility of speech for speech privacy situations. Although the experimental setups and goals of each investigation were different, many aspects of the experimental procedure were the same.

Experimental conditions were characterized in terms of uniformly weighted (over speech frequencies) SNRs as defined by the following:

$$\mathrm{SNR}_{\mathrm{uni32}} = \frac{1}{16}\sum_{f=160}^{5000} \{L_s(f) - L_n(f)\}_{-32} \quad (\mathrm{dB}), \qquad (1)$$

where $L_s(f)$ and $L_n(f)$ are the 1/3-octave band speech and noise levels at the listener's position. The −32 indicates that differences in the brackets are clipped so that they can never be less than −32 dB. $\mathrm{SNR}_{\mathrm{uni32}}$ values were previously shown[2] to be the most successful compromise for predicting subjective judgments of both the audibility and the intelligibility of speech when rating the speech privacy of meeting rooms.

Conditions were subjectively evaluated using speech tests to determine the audibility and intelligibility of speech for each test condition following procedures similar to those in the initial study.[2] Recordings of the Harvard sentences[16] were used as the speech test material. They were high quality digital recordings spoken by a male talker in an anechoic room. The Harvard sentences are phonetically balanced and of low predictability, which is necessary for conditions of low intelligibility for which guessing could distort the scores in some other types of speech intelligibility tests.

The test protocol consisted of first turning on the noise signal and then a few seconds later playing one sentence. After the noise had stopped, the subject could tell the experimenter what they had heard, using a microphone to communicate with the researcher who was outside the test room. The subjects first did a practice test consisting of ten sentences that included $\mathrm{SNR}_{\mathrm{uni32}}$ values distributed over the complete range included in the full test.

Because in some cases no screening of subjects for hearing sensitivity was possible, some may have had some hearing loss. In addition, others may not have had English as their first language. Therefore, the data analyses are based on the results of the listeners with the ten highest intelligibility scores over all conditions of the test. Using only the best ten listeners was expected to exclude listeners with less sensitive hearing or others who may have not been sufficiently fluent in English.

Subjects were all adults and were employees of the National Research Council who volunteered to participate after being contacted by electronic mail. They did not receive any compensation for their participation. The tests were approved by the Ethics Review Board of the National Research Council (Protocol No. 2006-06) and each subject signed a consent form after all of the details of the experiment were explained to them.

TABLE I. Construction details of the three walls used in the two-room validation tests, (RC, resilient channels; STC, sound transmission class).

Wall No.	Face No. 1 gypsum board (mm)	Stud type	Cavity insulation	RC	Face No. 2 gypsum board (mm)	STC
1	2×16	92 mm steel	90 mm mineral fiber	No	2×16	56
2	13	90 mm wood	90 mm glass fiber	Yes	2×13	53
3	13	90 mm wood	90 mm glass fiber	Yes	13	46

III. TWO-ROOM VALIDATION TEST

A. Procedure

The initial tests[2] had included a wide range of speech and noise levels but were carried out in approximately free-field conditions with spatially separated speech and noise sources. The test subjects listened to speech sounds, modified to represent transmission through one of several walls, from a loudspeaker system in front of them. At the same time they heard ambient noises from a second loudspeaker system located above them. These conditions were intended to represent a worst-case condition (i.e., minimum speech privacy) in which it would be most easy to understand speech. The new two-room validation tests[17] were intended to explore how the results might differ for more typical room acoustics conditions.

In the new validation tests, speech was radiated into one room and listeners heard it while located close to the common wall in an adjacent room. The recorded test sentences were reproduced using a dodecahedron loudspeaker located approximately 2 m from the center of the test wall in a large ($250\ m^3$) reverberation chamber. The speech sounds were naturally transmitted through the test wall into the adjacent room ($140\ m^3$) where the listener was located. The constructions of the three walls are described in Table I and they had sound transmission class (STC) values of 46, 53, and 56.

Sound absorbing material was added to both rooms to reduce the reverberation times (averaged over frequencies from 160 to 5000 Hz) in the receiving room to 0.64 s and in the source room to 0.80 s. With the added sound absorbing material present, the listeners heard speech sounds in realistic conditions representative of typical meeting rooms and heard speech sounds that had been transmitted through real walls.

The integrated 1/3-octave band levels of each test sentence were obtained as room average levels in the source room. The attenuation of speech sounds was determined using a broadband white noise test signal, as the difference between the average level in the source room and the level at a point 0.25 m from the test wall in the receiving room (similar to the procedure of ASTM E2638-08).[18] The average levels of each test sentence in the source room and the measured attenuations were used to determine the speech levels in the receiving room at the location of the listener 0.25 m from the test wall. This was a more accurate estimate of the speech levels at the receiver position than would be possible by directly measuring the, often low level, transmitted speech sounds.

Ambient noise was added to the receiving room from a single loudspeaker located across the room from the listener. Random noise, which was equalized to have an approximately −5 dB/octave spectrum shape representative of typical ventilation noise,[14,15] was used. The ambient noise levels were measured using a single microphone at the location of the listener's head. Three different ambient noise levels were used: 24, 29, and 34 dBA. These were selected so that at the listener's position, speech sounds varied from barely audible to completely intelligible.

From the transmitted speech levels and the ambient noise levels at the location of the listener's head, SNR_{uni32} values were calculated. For each of three test walls, speech and noise levels were adjusted to give the most useful range of SNR_{uni32} values. The intention was that the SNR_{uni32} values should be evenly distributed over conditions from, not audible to most listeners, to quite intelligible to most listeners. An even distribution of SNR_{uni32} values from just below −25 dB to a maximum of about −3 dB was used.

The test protocol consisted of first turning on the noise signal and then a few seconds later playing one sentence. A few seconds after the end of the sentence, the noise stopped. In this experiment the start and stop of the noise was marked by a chime sound so that subjects did not miss the start of the often very low levels of noise. In a few cases there was no sentence present to minimize the probability of subjects always guessing that they did hear some speech. After the noise had stopped, the subject told the experimenter what they had heard using a microphone to communicate with the researcher who was outside the test rooms. The subject first said whether any speech sounds were audible, and then if some speech was audible said whether they could hear the cadence or rhythm of the speech. Finally if they had heard the cadence, they repeated the words that they had understood back to the experimenter.

After all subjects had listened to all sentences, it was possible to calculate the fraction of the subjects, for each test sentence, who were able to hear any speech sounds, and determine an estimate of the audibility threshold as the SNR_{uni32} value for which 50% could just hear some speech sound. Similarly, from the number of subjects who were able to hear the cadence of the speech, an estimate of the thresh-

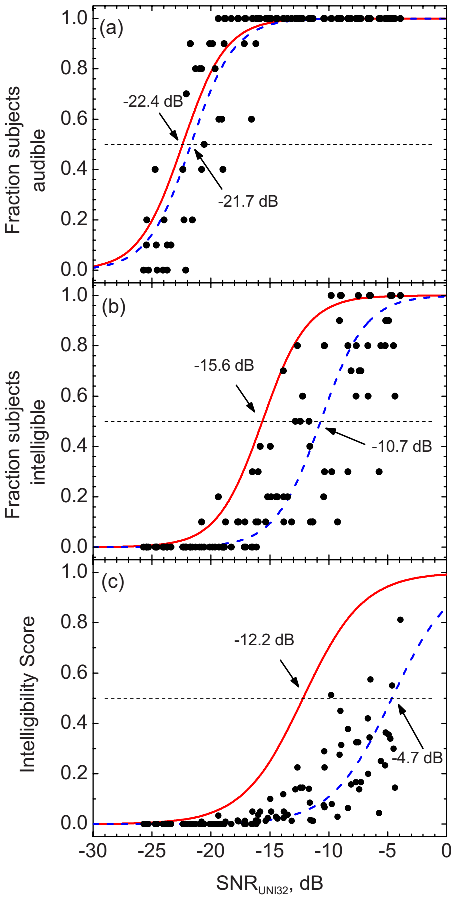


FIG. 1. (Color online) Comparison of results of initial study in approximately free-field conditions (solid line) (Ref. 2) with the results of a two-room validation study (data points and dashed line): (a) fraction of subjects finding some speech audible, (b) fraction of subjects able to understand at least one word, and (c) speech intelligibility scores, all plotted versus values of the uniformly weighted signal-to-noise ratio (SNR$_{\text{uni32}}$). The decibel values and arrows on each plot indicate the various threshold values [(a) and (b)] and speech reception threshold (c).

old of the cadence was determined. The intelligibility threshold was determined as the point at which 50% of the subjects could just understand at least one word of each test sentence and the speech intelligibility score was determined as the fraction of the words correctly understood.

B. Results

The scores were first plotted versus SNR$_{\text{uni32}}$ values separately for each wall tested. However, it was not possible to detect systematic differences among the results for the different walls and the data for all three wall tests were combined.

The upper part of Fig. 1 plots the fraction of listeners indicating that they heard some speech sounds versus SNR$_{\text{uni32}}$ values for the combined data from all three walls. Each point indicates the fraction of the ten best listeners who heard speech sound for that particular condition. The solid line in this plot is the best-fit regression line obtained in the previous work.[2] The dashed line is the same form of curve but shifted to the right to better fit the new measured data. The point at which 50% of the subjects found the speech to be just audible was described as the threshold of audibility of the speech and corresponded to a SNR$_{\text{uni32}}$ of −22.4 dB in the initial study. The best-fit lines were Boltzmann equations given by

TABLE II. Regression coefficients for the Boltzmann best-fit equations in Fig. 1. "Initial" identifies results from the previously published (Ref. 2) initial study and "validation" results form the new two-room validation study (Ref. 17).

Response	Expt.	dx	x_0
Threshold of audibility	Initial	1.8053	−22.41
	Validation	1.8053	−21.70
Threshold of cadence	Initial	1.4037	−20.05
	Validation	1.4037	−17.50
Threshold of intelligibility	Initial	1.8379	−15.64
	Validation	1.8379	−10.70
Speech intelligibility score	Initial	2.5259	−12.19
	Validation	2.5259	−4.65

$$y = \frac{(A_1 - A_2)}{1 + e^{(x-x_0)/dx}} + A_2, \tag{2}$$

where y is the fraction responding or intelligibility score, x is the corresponding SNR, x_0 is the SNR for a y value of 0.5, dx is related to the slope of the midportion of the curve, A_1 is the minimum y value=0.0, and A_2 is the maximum y value =1.0.

The regression fits to the new data were obtained by shifting the corresponding regression line from the initial study[2] to the right an amount that minimized the rms deviation about the regression line for the new data. Thus the dx parameter was not changed and only the x_0 value was changed corresponding to shifting the regression line horizontally.

Table II lists the x_0 and dx parameters for each of the regression lines in Fig. 1. For the audibility scores, the shift between the new results and the previous results for the threshold of audibility was only 0.7 dB (i.e., −22.4 to −21.7 dB).

To examine the significance of the differences between the initial study results and the new results, the standard deviations of the measured values about the new best-fit line in terms of SNR$_{\text{uni32}}$ values were calculated. This cannot be done for scores of either 0 or 1 for which SNR$_{\text{uni32}}$ values are not defined for this type of equation. However, the standard deviation of the data points about the new best-fit line was calculated for all other points to give an indication of the relevance of the shift between the two best-fit lines. For the results in Fig. 1, the standard deviation of the remaining data points (excluding points with scores of 0 or 1) about the new best-fit line was ±2.2 dB in terms of SNR$_{\text{uni32}}$ values. This is much larger than the 0.7 dB shift, and indicates that within the limits of the data, the new audibility data replicate and hence validate the previous results with respect to the audibility of transmitted speech sounds.

Figure 1(b) plots the fraction of listeners who understood at least one word versus SNR$_{\text{uni32}}$ values. For these data the new best-fit line that is shifted 4.9 dB to the right minimizes the vertical deviations of the data about the line. This is a quite large shift and almost all of the data points are to the right of the previous best-fit line (the solid line in this figure). The standard deviation of the data about the new best-fit line in terms of SNR_{uni32} values was ±3.0 dB. The shift in the new best-fit line relative to the initial best-fit line is much larger than the scatter of the data about the new best-fit line. For these data, listeners were much less likely to understand at least one word for a given SNR_{uni32} value than in the previous work. These results do not replicate the previous work and indicate a greater degree of speech privacy for a given SNR_{uni32} value than did the previous results. According to the new results, the threshold of intelligibility (the point where 50% of the subjects could just understand at least one word) corresponds to a SNR_{uni32} value of −11 dB rather than the −16 dB value from the initial work.

The fractions of the best ten subjects indicating that they heard the cadence of the speech sounds versus SNR_{uni32} values were also determined.[17] For these data the new regression line was shifted 2.5 dB to the right to best fit the new data and the threshold of cadence was shifted from −20.0 to −17.5 dB. This shift was intermediate to that for the audibility threshold and that for the intelligibility threshold.

The speech intelligibility scores are plotted versus SNR_{uni32} values in Fig. 1(c). The results are quite similar to those for the speech intelligibility threshold (SIT) in Fig. 1(b) in that a quite large shift is required to align the new best-fit regression line with the measured speech intelligibility scores. In this case the new best-fit line was shifted 7.5 dB to the right relative to the previous best-fit line shown in Fig. 1(c). The standard deviation about the new best-fit line in Fig. 1(c) in terms of SNR_{uni32} values was ±2.9 dB. Again the new results cannot be said to replicate the old results. The new data indicate much lower speech intelligibility scores at a given SNR_{uni32} value.

C. Discussion of results of validation tests

The main purpose of these listening tests was to compare the results of the initial study with the new results obtained in a more realistically valid acoustical environment. The new results presented here do partially agree with the previous results, but also indicate some significant differences.

There were not significant differences between the audibility threshold results from the previous work and the new results presented here. These new results validate the previous threshold of audibility of speech sounds and confirmed it to be a SNR_{uni32} value of −22 dB.

On the other hand, the mean trends for the threshold of intelligibility and for the speech intelligibility scores indicate differences of 5–7 dB relative to the previous studies. These are large and important differences that could lead to requiring walls with STC values 5–7 points larger than necessary. Such a large difference could lead to significant additional construction costs and it is essential to understand the cause of these differences.

There are many possible sources of error that could influence the results. For example, there could be errors in the measurement of speech and noise levels. However, this is unlikely to be the cause of the differences, since the same SNR values were used for both the audibility threshold and for the intelligibility threshold results. Since there is good agreement for the threshold of audibility results, it is reasonable to assume that the speech and noise level measurements were correct in both studies and that the subjects were capable of hearing the speech sounds.

It is likely that the differences in intelligibility results are due to factors that would affect the understanding of speech and not the simple perception of the presence of speech sound as in the audibility test. Perhaps the most obvious factor is reverberation. It is well known that the intelligibility of speech is influenced by both the SNR and the reverberation time of the listening space. However, the effects of reverberation on speech intelligibility at the very low SNRs that are of concern in speech privacy issues are not well defined. In our new two-room validation experiment, the subjects listened to just audible speech modified by the reverberation of both the source and the receiving rooms, but no measure of the effects of reverberation was included in determining the SNR_{uni32} values.

One other possible contributing factor was the differences in the spatial characteristics of the speech and noise sounds to which the listeners were exposed, which are well known to influence the intelligibility of speech.[6,7] In the initial work[2] subjects listened in approximately free-field conditions. The speech sounds arrived from a loudspeaker system directly in front of the seated subject and the noise sounds arrived from another loudspeaker system directly overhead. Such conditions would maximize a listener's ability to understand speech.

In the new tests, speech sounds were radiated into one room with an average reverberation time of 0.80 s (averaged over the frequencies 160–5000 Hz); they traveled through a real wall and from the wall a further 0.25 m to the listener's ear. Other speech sounds would reflect about the rooms and arrive a little later at the listener's ears. The noise source was located at the opposite side of the room to the listener and hence the direct speech and noise sounds were spatially separated. However, the noise source was in a room with an average reverberation time over speech frequencies of 0.64 s and the listener would hear reflected noise sounds from many directions.

Although the conditions in the new experiment were more realistic and more representative of listening in real rooms, it is not known how the differences in added reflected speech and noise sounds would each affect speech intelligibility scores. Small amounts of reverberant sound do not usually have large effects on the intelligibility of speech. However, it is possible that we are more sensitive to the negative effects of reverberation for the very low signal-to-noise conditions in these experiments. Our knowledge of the effects of the spatial separation of speech and noise sources is almost entirely based on the perception of only the direct sounds in free-field conditions.[6,7] It is quite possible that the addition of reflected sound to the speech and noise signals considerably modifies these effects.

TABLE III. Horizontal and vertical angles of the loudspeakers relative to the listener's head. Angle 0,0 is straight ahead of the listener's head.

	Loudspeaker	Horizontal angle (deg)	Vertical angle (deg)
1	Center	0	0
2	Left	−32	0
3	Right	+32	0
4	Center high	0	25
5	Left high	−37	28
6	Right high	+37	28
7	Left rear	−115	0
8	Right rear	+115	0

IV. INVESTIGATIONS OF SPATIAL AND TEMPORAL EFFECTS

A. Procedure

A second series of four experiments[19] was carried out to better understand the magnitude of particular details of the spatial and temporal differences in the sound fields of the initial tests and the new two-room validations tests. The experiments were mostly carried out in simulated sound fields in an anechoic room using an eight-channel simulation system. The eight loudspeakers of the system were positioned around the subject at angles listed in Table III and illustrated in Fig. 2. Five of the loudspeakers (1, 2, 3, 7, and 8) were in the horizontal plane of the listener's ears and the other three were raised up above this plane. The listener sat facing loudspeaker No. 1, which reproduced the direct sound.

The signals to each loudspeaker were processed by four Yamaha DME32 digital signal-processing units connected together to function as one large unit. Speech and noise signals were separately processed and mixed together for each loudspeaker. The loudspeakers were Tannoy model 800A units with concentric drivers so that all frequencies were radiated from the same location. In some cases all loudspeakers radiated speech and noise signals and in other cases only selected loudspeakers were used depending on the purpose of each test condition. For the noise signals, large delays were introduced between the signals to the different loudspeakers so that they arrived at the listener incoherently and when all eight loudspeakers were used conditions were perceived as very diffuse.

Because early-arriving sounds are perceived differently than later-arriving sounds,[20] early- and later-arriving speech sounds were varied independently. Each loudspeaker could reproduce simulations of four early-arriving sounds for a total of 32 early-arriving sounds. The first early-arriving sound arrived from loudspeaker No. 1 (see Table III) to simulate the direct sound. The other early-arriving sounds arrived within a 50 ms interval after the arrival of the simulated direct sound and were intended to simulate various early-arriving reflections. Digital reverberator components in the DME32 units were used to simulate the many later-arriving reflections that would occur in a room.

One of the new experiments was carried out in the same conditions as the initial study[2] so that the same geometry of speech and noise sources could be included. This is a sound isolated, quiet (background level of 13 dBA), and acoustically dead space. In this space, speech sounds were reproduced by loudspeakers approximately 2 m in front of the listener and located behind a curtain. A second set of loudspeakers in the ceiling void above the subject was used to produce simulated ventilation noise. In some cases the loudspeakers in front of the subject were used to reproduce both speech and noise sounds. A Yamaha DME32 unit was used to control the sounds to each of the loudspeaker systems. Each set of loudspeakers consisted of two Paradigm Compact Monitors and a Paradigm PW sub-woofer with a response corrected to be flat ±1 dB from 60 to 12 000 Hz at the listener's position.

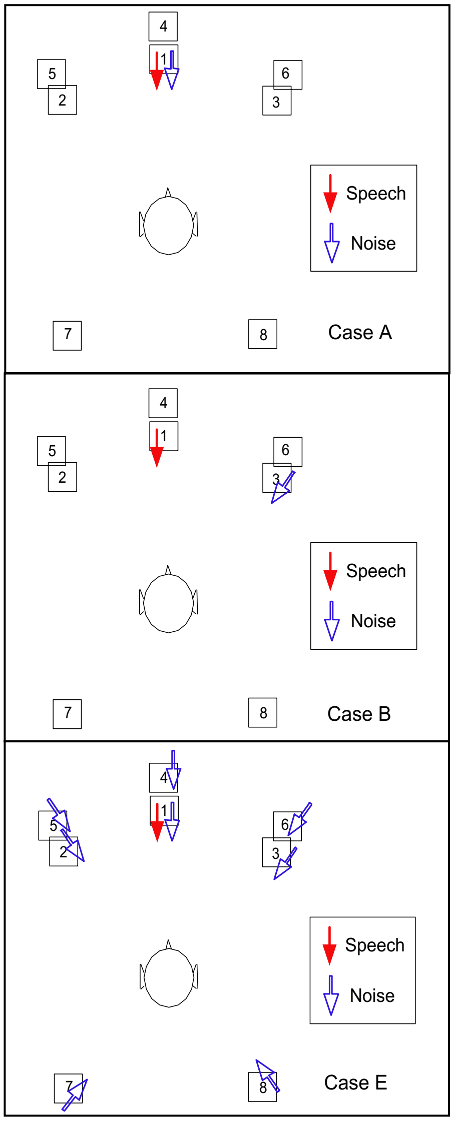


FIG. 2. (Color online) Plan of loudspeaker locations and descriptions of simulated sound fields: Case A: speech and noise only from the same loudspeaker (No. 1) directly in front of the listener, case B: speech and noise only from two separate loudspeakers (Nos. 1 and 3), and case E: speech from ahead only and noise from all loudspeakers.

In some cases speech sounds were filtered by the DME32 unit to represent transmission through a wall. A wall consisting of 16 mm gypsum board on both sides of lightweight 90 mm steel studs and with glass fiber material in the

TABLE IV. Descriptions of cases A, B, and E to demonstrate simple spatial release from masking and cases C, D, and F to demonstrate the effects of added early-arriving reflections of speech sounds.

Case	Speech	Noise	Wall
A	Direct from No. 1	Direct from No. 1	None
B	Direct from No. 1	Direct from No. 3	None
C	Direct+early reflections	Direct from No. 3	None
D	Direct+early+T_{60}=1	Direct from No. 3	None
E	Direct from No. 1	Diffuse from all	None
F	Direct+early reflections	Diffuse from all	None
G	Direct+early+T_{60}=1	Diffuse from all	None

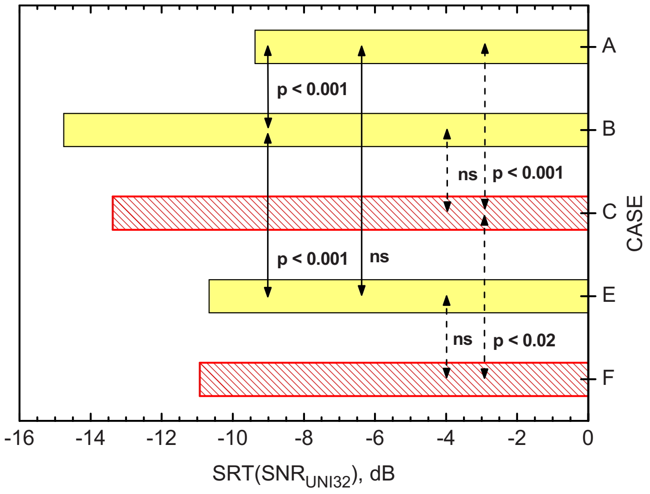


FIG. 3. (Color online) SRT values for cases A, B, C, E, and F with configurations illustrated in Fig. 2 and described in Table IV. Arrows indicate the two cases for which significance of the difference is given. Solid filled bars: simple spatial release from masking cases. Hatched bars: cases with added early-arriving reflections.

cavity was simulated. This construction would correspond to a STC rating of 47 and is typical of many interior office walls.

For these experiments ambient noise with an approximately −5 dB per octave spectrum shape was again used with an overall level of 45 dBA. This is often referred to as "neutral" sounding and is representative of typical indoor noise spectra.[14,15]

To focus on the main differences between the two previous studies, in these tests only speech intelligibility scores were obtained. A number of spatially different combinations of speech and noise sources were compared as well as differences in added reflected sounds. These are later described with the results of each comparison. For each configuration, tests were carried out at two different SNRs. To make it possible to combine the results for two different SNRs all speech intelligibility scores were converted to speech reception threshold (SRT) values. The SRT is the SNR for which the mean intelligibility score is 50%. This provides more accurate descriptors of the intelligibility scores for each configuration by basing the results on a larger number of responses because the results of two SNRs could be combined.

SRT values were calculated using the same Boltzmann equations described previously. When plotting speech intelligibility scores versus SNR$_{\text{uni32}}$ values, the SRT value is the x_0 value in Eq. (2). Using the same dx values as in the previous studies (Table II), the SRT can be calculated using the Boltzmann equation and the mean intelligibility scores with the corresponding SNR$_{\text{uni32}}$ values from each test. The results were obtained in four separate experiments, which are described in Table VII in the Appendix along with the SRT values for each configuration tested.

B. Simple spatial release from masking

The first tests demonstrated simple spatial release from masking effects to confirm the validity of the new results as well as to serve as a reference case with which subsequent results can be compared. Configurations A, B, and E are described in Fig. 2 and Table IV. Figure 2 shows a top down view of the subject and loudspeakers with solid arrows indicating sources of speech sounds and open arrows for sources of simulated ventilation noise. As illustrated in Fig. 2, speech and noise were produced by only loudspeaker No. 1 directly in front of the subject in case A. In case B the speech and noise sources were separated in the horizontal plane so that the speech came from only loudspeaker No. 1 (directly in front of the subject) and the noise came from only loudspeaker No. 3 (32° to the right of straight ahead). In case E, the speech again came from only loudspeaker No. 1 but the noise was radiated from all eight loudspeakers. Because of the large delays between the noise signals to each loudspeaker, the subject experienced a very diffuse noise sound.

The SRT values for these three cases are compared in the filled bars of Fig. 3. (SRT values and details of all cases are summarized in Table in the Appendix). Of the three, case A has the highest SRT value (−9.37 dB) indicating the lower intelligibility scores that result when the speech and noise come from exactly the same source location. Case B has a much lower SRT value (−14.75 dB), indicating a spatial release from masking when the sources of the speech and the noise are spatially separated. However, when the noise comes from all directions, as in case E, the SRT (−10.67 dB) is similar to that for case A.

A oneway analysis of variance for the experiment No. 1 results, which included these three conditions, indicated significant variations in SRT values (F=25.15, p<0.001). A post hoc Bonferroni test of the individual differences indicated that cases A and B were significantly different (p<0.001) as were cases B and E (p<0.001). However, the difference between the SRT values for cases A and E was not statistically significant.

When the noise source was separated from the speech source by 32° in the horizontal plane the SRT decreased by 5.4 dB, which agrees with previously published results.[6,11] The comparisons with case E results suggest that a diffuse masking noise arriving from all directions leads to results that are quite similar to the case of coincident speech and noise sources (case A) and in this experiment cases A and E were not significantly different. That is, in real rooms with at least somewhat diffuse noise, and where the listener is in the reverberant field of the talker, we would expect less spatial release from masking for separated speech and noise sources.

C. Effects of early-arriving reflections of speech

Early-arriving reflections were radiated from all eight loudspeakers, so that they arrived within about 50 ms after the direct sound and decreased realistically in amplitude with increasing time. Such early-arriving reflections have been shown to increase speech intelligibility equivalent to increasing the level of the direct sound by the increase in energy added by the early-arriving reflections.[20] In these experiments the speech levels were adjusted to be the same for both with and without early-arriving reflections cases. Case C included direct speech and early reflections (ERs) of speech from all loudspeakers combined with simulated ventilation noise from only loudspeaker No. 3. Case F was similar to case C but included simulated ventilation noise arriving diffusely from all eight loudspeakers (as in case E).

Figure 3 compares the SRT values for cases C and F (hatched bars) with those configurations previously described (solid filled bars). Case C is essentially the result of adding ERs to the conditions of case B without any change in the overall speech level. Figure 3 indicates a small increase in SRT value for case C relative to case B. Case F is essentially the result of adding early-arriving reflections of speech sounds to case E. The results in Fig. 3 indicate a small decrease in the SRT value for case F relative to case E.

A oneway analysis of variance on the experiment No. 1 conditions showed a statistically significant pattern of variations in SRT values ($F=25.25$, $p<0.001$) but a post hoc Bonferroni test showed that the SRT values for cases B and C (with added ERs) were not significantly different. Similarly the SRT values for cases E and F were not significantly different.

Although not statistically significant, adding ERs, while maintaining the same overall speech level, did increase the SRT value a little for case C compared to case B. This may have been because the speech energy in the ERs included some speech from the same direction as the noise, which might tend to increase the SRT a little for this case. When ERs were added to configurations in which the noise came from all eight loudspeakers, the added early-arriving speech reflections decreased the SRT slightly but not significantly (cases E and F). In both cases adding ERs of the speech energy was largely equivalent to increasing the level of the direct speech sound as was expected. ERs did not significantly affect spatial unmasking effects in these simulations.

D. Added effects of reverberant speech

The digital reverberators in the DME32 units were used to add a reverberant decay of the speech sounds after the early-arriving reflections, but with the overall speech level not varied. In the configurations presented here, the reverberant sound had a mid-frequency reverberation time of 1.0 s. Figure 4 compares the resulting SRT values with those for the previously described cases.

Case D was constructed by adding a reverberant decay to the case C configuration, which had included direct speech, early-arriving reflections of the speech sounds, and noise from only loudspeaker No. 3. When reverberant speech was added to case C, the SRT was increased from -13.78 to -10.27 dB and this difference was statistically significant (oneway $F=25.25$, $p<0.001$, post hoc Bonferroni $p<0.001$).

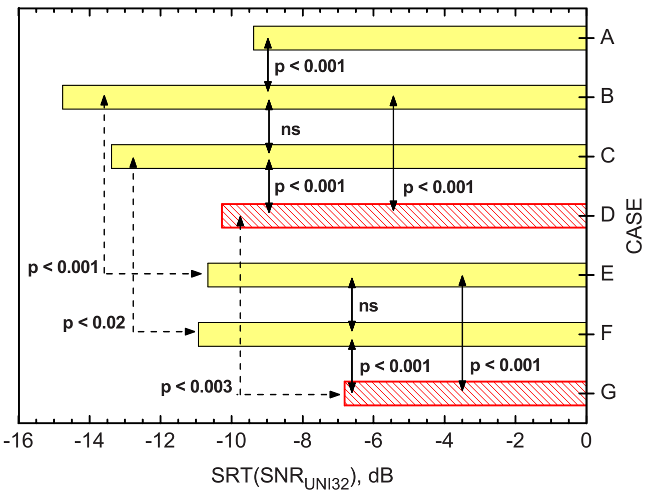


FIG. 4. (Color online) SRT values for cases A–G with configurations illustrated in Fig. 2. Solid filled bars are repeated from previous figure. See Table IV for descriptions of cases.

Similarly, for the cases with diffuse noise from all loudspeakers, adding reverberant sound increased the SRT from -10.93 dB for case F to -6.82 dB for case G. This difference was statistically significant (oneway $F=25.25$, $p<0.001$, post hoc Bonferroni $p<0.001$).

While adding ERs of the speech did not significantly change the SRT (for constant speech level), adding reverberant speech increased the SRT by 3–4 dB independent of the spatial differences in the simulated ambient noise. The addition of reverberant speech adds to the masking of the speech independently from that of adding diffuse noise.

E. Effects of varied reverberation time

The effect of reverberant speech was further investigated by comparing conditions with reverberation times of 0.5, 1.0, and 2.0 s. All cases included direct speech and early-arriving reflections with added reverberant speech with one of three different reverberation times. These comparisons were also repeated with and without modifications to the spectrum of the speech sound to simulate the effect of propagation through the wall.

Case P corresponded to no wall, and speech with added ERs. Speech and noise sounds were reproduced by all eight loudspeakers. This base case without reverberant sound was compared with cases Q, R, and S, which had added reverberant speech with reverberation times of 0.5, 1.0, and 2.0 s, respectively. The resulting mean SRT values for each case are given in Fig. 5 (see Table V).

For the cases without a simulated wall, adding reverberant speech with a 0.5 s reverberation time (case Q) to case P (which had only ERs of the speech) only increased the SRT by a small amount and the difference was not statistically significant. However, adding more reverberant speech corresponding to a 1.0 s reverberation time (case R) and a 2.0 s reverberation time (case S) produced larger and statistically

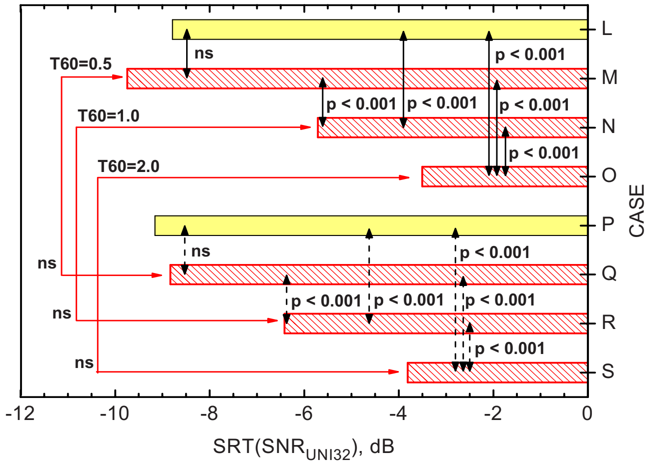


FIG. 5. (Color online) SRT values for cases L–S, all with speech and noise sounds from all eight loudspeakers. Bars with hatched lines show cases with varied reverberation time. See Table V for descriptions of cases.

significant increases in SRT values (oneway on the experiment No. 4 data: $F=94.15$, $p<0.001$, post hoc Bonferroni, $p<0.001$).

Case L was similar to case P but with the inclusion of filtering to simulate transmission through a wall for the speech sounds. The SRT value from case L is compared with those for cases M, N, and O with reverberation times of 0.5, 1.0, and 2.0 s, respectively, in Fig. 5. All of these cases included simulated transmission through a wall for the speech sounds.

When a simulated wall was included, the results of cases L, M, N, and O showed a similar progression of changes in the SRT values. Adding only reverberant speech with a 0.5 s reverberation time (case M) led to a non-significant change. However, adding reverberant speech with a 1.0 or a 2.0 s reverberation time each led to large and statistically significant increases in SRT values (oneway on the experiment No. 3 data: $F=31.51$, $p<0.001$, post hoc Bonferroni $p<0.001$).

For these cases with reverberation times of 0.5, 1.0, and 2.0 s, the corresponding with-wall and without-wall cases were not significantly different (independent t-test) (indicated by the solid U-shaped arrows in Fig. 5).

In order to better understand the changes in SRT values caused by increasing the reverberant speech sound, the mean SRT values are plotted versus the logarithm of the reverberation time in Fig. 6. The data for the cases with only ERs are included with a reverberation time of 0.5 s because these results were not statistically different than the results for a 0.5 s reverberation time. On this plot the SRT values are seen to increase linearly as the logarithm of the reverberation time increased. The results for the with and without a simulated wall cases led to very similar results. The two regression equations and an average for all data are

TABLE V. Descriptions of the effects of added reverberation: cases P, Q, R, and S without a wall and cases L, M, N, and O with a wall.

Case	Speech	Noise	Wall
L	Direct+early reflections	Diffuse from all	Wall
M	Direct+early+$T_{60}=0.5$	Diffuse from all	Wall
N	Direct+early+$T_{60}=1$	Diffuse from all	Wall
O	Direct+early+$T_{60}=2$	Diffuse from all	Wall
P	Direct+early reflections	Diffuse from all	None
Q	Direct+early+$T_{60}=0.5$	Diffuse from all	None
R	Direct+early+$T_{60}=1$	Diffuse from all	None
S	Direct+early+$T_{60}=2$	Diffuse from all	None

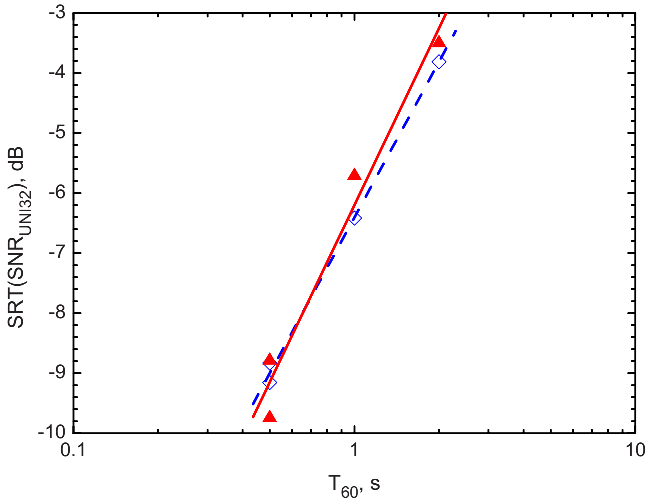


FIG. 6. (Color online) Mean SRT values plotted versus the logarithm of the reverberation time (T_{60}) of the simulated speech sounds. Dashed line and open symbols: no wall; solid line and filled symbols: speech transmission through a wall.

$$\mathrm{SRT} = 8.602 \log_{10}(T_{60}) - 6.407 \quad (\text{no wall}), \tag{3}$$

$$\mathrm{SRT} = 9.774 \log_{10}(T_{60}) - 6.201 \quad (\text{with wall}), \tag{4}$$

$$\mathrm{SRT} = 9.187 \log_{10}(T_{60}) - 6.304 \quad (\text{all data}). \tag{5}$$

These can be used to predict the effects of reverberation in meeting rooms on the SRT of the transmitted speech.

F. Effects of horizontal and vertical separations of speech and noise sources

Figure 3 showed the results of separating single speech and noise source by 32° in the horizontal plane. Because the initial study[2] included vertically separated speech and noise sources, such cases were repeated here both with and without a simulated wall. The SRT values for the horizontally and the vertically separated conditions are compared in Fig. 7. In all cases separating speech and noise sources led to a highly significant decrease in SRT values ($p<0.001$). A comparison of the SRT values for cases A and B indicated a 5.4 dB shift in SRT values when the speech and noise sources were separated by 32° horizontally.

Case I was essentially a repeat of case A except in a completely different experiment and test facility. In both cases speech and noise sounds came from only the loudspeaker directly in front of the listener. The difference between the SRT values of cases A and I was very small and was not statistically significant. This makes it possible to compare the effects of horizontal separation of speech and noise sources (cases A and B) with vertical separation (cases H and I). In case H the speech and noise sources were separated by 90° in the vertical plane but the spatial release from

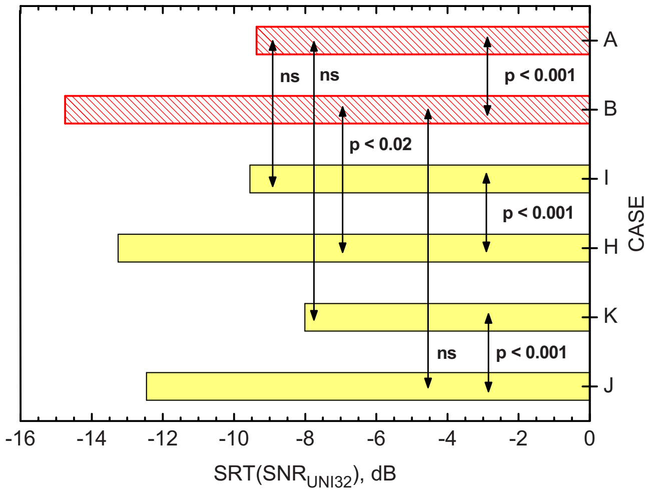


FIG. 7. (Color online) SRT values for cases A, B, I, H, K, and J with separations of speech and noise in the horizontal plane (hatched bars) and in the vertical plane (solid filled bars).

masking (decrease in SRT value) was a little less than when speech and noise sources were separated by 32° in the horizontal plane (cases A and B).

Comparing SRT values for cases K and J indicates a similar spatial release from masking when the speech and noise were separated by 90° in the vertical plane and when speech sounds were in both cases modified to represent transmission through a wall. The SRT values tended to be slightly higher than for the corresponding cases without a simulated wall. However, the differences were not statistically significant. (It is interesting to note that the SRT value for case J was −12.46 dB, which was very similar to the value of −12.19 obtained previously.[2])

G. Effects of single, diffuse, and semi-diffuse noise sources

The difference between the masking effects of ambient noise from a single noise source, which was spatially separated from the speech source, and speech with a diffuse noise from all eight loudspeakers represents two extremes. It is likely that in real rooms intermediate cases are found for which the noise might be described as "semi-diffuse." Such semi-diffuse conditions were produced by radiating the simulated ambient noise predominantly from three adjacent loudspeakers. Because one case included a cluster of noise sources from the rear side of the listener, a single rear-side noise source was also tested as a reference case. These new noise source conditions were compared with the previously described conditions that included either a single noise source or diffuse noise from all eight loudspeakers.

Figure 8 compares SRT values for varied noise masking configurations. The upper five configurations (D, T, V, U, and G) on the graph are for conditions that did not include a simulated wall. Conditions, which included a simulated wall, are in the lower three bars of the graph (W, X, and N). The speech signal in all conditions included direct sound, ERs, and reverberant speech with a 1 s reverberation time to represent conditions in a real room.

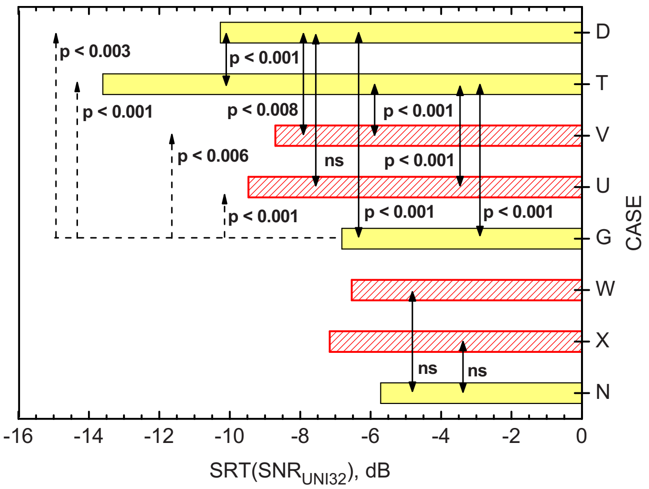


FIG. 8. (Color online) SRT values for cases D, G, N, T, U, V, W, and X with varied masking noise configurations. Cases D, T, V, U, and G do not include a simulated wall; cases W, X, and N include a simulated wall. (Bars with hatched lines are semi-diffuse cases.)

Comparing SRT cases D and T shows that the single noise source from the rear side (case T with noise from only loudspeaker No. 8) led to a larger spatial release from masking (i.e., lower SRT) than for a single noise source from the front side (case D) and the difference was highly significant (independent t-test, $p<0.001$). SRT case V for a semi-diffuse noise from the front side has a higher SRT than the single noise source from the front side (i.e., case D) (independent t-test, $p<0.008$). Similarly, a semi-diffuse noise source from the rear side (SRT case U) had a higher SRT than the single noise source from the rear side (SRT case T) (independent t-test, $p<0.001$). However, SRT case G, with diffuse noise from all eight loudspeakers, had the highest SRT of all the cases without a simulated wall.

The semi-diffuse conditions (cases U and V) were intermediate to the single separated noise source conditions (cases D and T) and the all eight loudspeaker noise source condition (case G). Noise sources to the rear side (cases T and U) led to lower SRT values than the corresponding noise sources from the front side (cases D and V). As shown in Fig. 8 all of the differences due to changes in either the direction or diffuseness of the noise were statistically significant.

The lower three bars of Fig. 8 (configurations W, X, and N) include results for similar conditions except that the speech sounds were filtered to simulate transmission through the wall. These results show a little higher SRT values than the corresponding cases without walls. Although the corresponding conditions seemed to have higher SRT values when a simulated wall was included, none of the differences between the with- and without-wall cases were statistically significant. That is, cases with semi-diffuse noise from the front side (cases V and W) were not significantly different and neither were the two semi-diffuse noise from the rear-side conditions (cases U and X). Similarly, the two cases with all eight loudspeaker radiating noise (cases G and N) were not significantly different. From these results one must conclude

TABLE VI. SITs in terms of $\mathrm{SNR}_{\mathrm{un132}}$ values in decibels. "ER," direct sound and early-arriving reflections; "T_{60}=0.5," direct sound early-arriving reflections and reverberant sound with a 0.5 s reverberation time.

SRT case	Wall	Speech	Noise	SIT (dB)	Expt.
P	No wall	ER	All	−12.62	4
Q	No wall	T_{60}=0.5	All	−11.89	4
R	No wall	T_{60}=1	All	−10.35	4
S	No wall	T_{60}=2	All	−7.61	4

that there is no proof of an effect of transmission though a wall when conditions are described in terms of $\mathrm{SNR}_{\mathrm{uni32}}$ values.

H. Changes in speech intelligibility thresholds

The main focus of this work was on speech intelligibility scores expressed in terms of SRT values. However, previously established criteria for acceptable speech privacy were in terms of the SIT values.[2] The SIT is the SNR for which 50% of a panel of listeners can just understand at least one word of a test sentence. For approximately free-field conditions, the threshold of intelligibility was found to correspond to a $\mathrm{SNR}_{\mathrm{uni32}}$ value of −16 dB.[2] However, in the two-room validation tests [see Fig. 1(b)], the threshold of intelligibility was increased by 4.9 dB to a $\mathrm{SNR}_{\mathrm{uni32}}$ value of approximately −11 dB.

To determine threshold of intelligibility values requires data for conditions with a significant number of responses with low intelligibility scores extending down to zero. In most of the new tests such conditions were deliberately avoided to provide data mostly in the range 10%–90% intelligibility scores. However, one series of conditions with varied reverberation times and with ambient noise coming from all eight loudspeakers did include a significant number of low intelligibility cases from which SITs could be determined. These data were for natural speech without simulated transmission through a wall and were used to determine new estimates of the effects of reverberation and diffuse noise on the SIT criteria.

Values of the threshold of intelligibility were calculated in a manner similar to the calculation of SRT values described in Sec. IV A of this paper. The fraction of the listeners indicating at least one word was understood for each test configuration was considered in terms of plots of these values versus $\mathrm{SNR}_{\mathrm{uni32}}$ values. A Boltzmann equation was fitted to the data using the same dx value as previously obtained for SITs (Ref. 2) corresponding to a value of 1.8739 (see Table II). SIT values were calculated using Eq. (2) and using $dx=1.8739$. The fraction of subjects understanding at least one word was y and x_0 was the threshold of intelligibility. As before, x was the $\mathrm{SNR}_{\mathrm{uni32}}$ value corresponding to the y value. The resulting SIT values are given in Table VI. The other information in Table VI is repeated from the description of configurations in Table VIIVII in the Appendix.

The calculated SITs are plotted versus the logarithm of T_{60} values in Fig. 9. The SIT values are approximately linearly related to the logarithm of the reverberation time similar to the plot of SRT values versus T_{60} in Fig. 6. As in Fig. 6, the case with only ERs was plotted as having a T_{60} value of 0.5 s.

Figure 9 suggests that for a typical meeting room, the SIT would be approximately a $\mathrm{SNR}_{\mathrm{uni32}}$ value of −11 dB, corresponding to a T_{60} value of 0.75 s. A little lower or higher values are indicated for lower or higher reverberation times. The criterion for the SIT should be raised from −16 dB for free-field conditions with spatially separated speech and noise sources,[2] to −11 dB for conditions where there is reverberant speech and diffuse ambient noise as found in most meeting rooms. This agrees well with the result from the two-room validation study as shown in Fig. 1(b). Small adjustments for differences in reverberation times could be made if needed but are usually not justified.

V. DISCUSSION

The new results clearly demonstrate that temporal and spatial effects influence speech privacy in that they can significantly change the intelligibility of speech and the threshold of intelligibility. Both the results in Sec. IV H and the results of the two-room validation study indicate that shifts in the threshold of intelligibility of 5 dB or more are possible. Ignoring these effects could lead to a 5 dB over-design of the sound insulation of the meeting room.

The SRT (expressed in terms of $\mathrm{SNR}_{\mathrm{uni32}}$ values) for the various conditions similarly indicate that speech intelligibility scores can be significantly affected by spatial and temporal effects of room acoustics. These results can be used to explain the difference between the initial results in approximately free-field conditions[2] and the new two-room validation study results. In the initial study, subjects heard speech

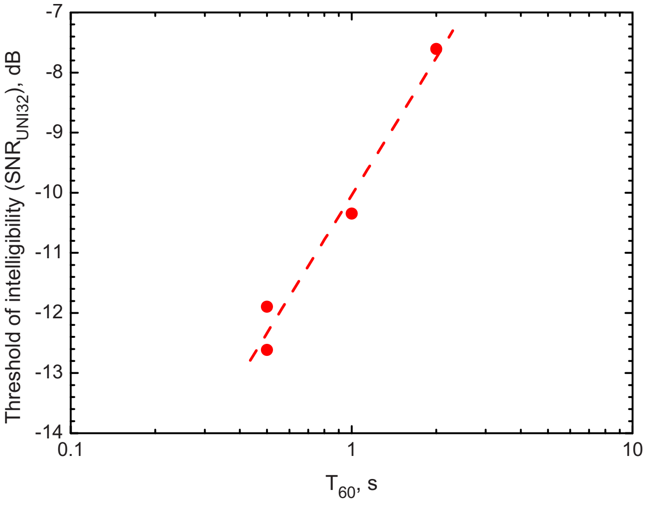


FIG. 9. (Color online) SIT versus T_{60} for unmodified speech (i.e., no simulated wall) and diffuse noise.

sounds from only directly in front of them and simulated ventilation noise from only above them. Going from this (case J) to the co-located speech and noise configuration (case K) decreased the SRT by 4.45 dB. Going from a co-located speech and noise configuration to a condition with diffuse noise would increase SRT by a further 1.30 dB. Combining these two differences would give a total SRT change of 3.15 dB. Finally adding on the effects of room reverberation from Fig. 9 leads to a further increase in SRT of 3.03 dB to give a total expected increase in SRT of 6.18 dB between the conditions of the initial study[2] and those of the two-room validation tests. This is reasonably close to the 7.5 dB difference in SRT values seen in Fig. 1(c).

Clearly these results can be used to estimate the effects for other conditions including varied sound diffusion of the rooms, varied reverberation times, and the effects of various spatial separations of speech and noise sources in spaces with high sound absorption. In spaces that are highly absorptive, listeners will benefit from spatial separation of speech and noise sources and speech privacy will be reduced. Strong early-arriving reflections of the speech sound will increase the effective speech level because the early-arriving reflections are not perceived as spatially or temporally separated from the direct sound and are equivalent to increasing the level of the direct sound. However, reverberant sound will minimize any spatial release from masking and lead to decreased intelligibility and hence increased speech privacy.

It is difficult to precisely compare the new results with those from the many previous studies in the literature because of the methodological differences among the various investigations. For example, subjective ratings of conditions have frequently been used rather than speech intelligibility scores.[11,13] Kollmeier and Wesselkamp[21] showed that the results of these two approaches are correlated but there are differences in the magnitude of the effects and the form of the trends with varying SNR can also be different. A number of studies have used such subjective ratings in an iterative procedure to determine SRT values. In their tests, the subject heard the same speech material repeatedly and decided when it appeared to be just intelligible. This is quite different than listening to each test sentence only once and counting the fraction of words correctly understood, as in the current work.

No previous work has examined the separate effects of early-arriving reflections of speech sounds on the various spatial effects. Descriptions of room acoustics conditions and reverberation are often not very detailed and conditions with as little as a 0.4 s reverberation time have been tested as a reverberant extreme.[12] In previous studies interfering sounds have most frequently been speech and much of the work was focused on explaining the cocktail party effect. Where the interfering sound has been noise, it has most often been noise with a speech spectrum shape. At least one study used white noise[22] but none have used noise representative of typical indoor ambient noises.

The initial test results comparing SRT cases A and B confirmed the expected spatial release from masking when the speech and noise sources were horizontally separated by 32° in free-field conditions. The 5.4 dB difference in SRT values for these two cases is of similar magnitude to results in several previous studies.[4,11,13] In Bronkhorst's review[6] he indicated that values between about 4 and 6 dB occur for this angular separation. The decreased SRT for the rear-side noise source (case T) is also supported by the results in Bronkhorst's summary[6] that indicate maximum spatial release from masking when the interfering source is at an angle of 110°–120° from straight ahead. A 90° vertical separation of the speech and noise sources (speech from in front of the listener and noise from overhead, as in cases H and I) had a 4.7 dB spatial release from masking, a little smaller than the 5.4 dB difference for the 32° horizontal separation. No previous measurements of the effect of a vertical separation of speech and noise sources were found.

There are very few previous results that can be compared with the diffuse interfering noise in the present study where the noise came incoherently from all eight loudspeakers as in case E of the present work. This resulted in a SRT only 1.3 dB lower than the case of coincident speech and noise sources (case A). That is, with diffuse noise, there is very little spatial release from masking. Some previous work has showed that the magnitude of the spatial release from masking decreases as the number of spatially separated noise sources increases.[11] In a similar manner, the semi-diffuse cases in the present work show that the spatial release from masking is significantly reduced when there were three spatially close masking sources compared to a single noise source.

Adding ERs to the speech sounds did not significantly change measured SRT values either with a horizontally separated speech and noise or with noise from all loudspeakers. This extends our understanding of the beneficial effects of early-arriving reflections on the intelligibility of speech[20] and it can be said that early-arriving reflections of speech sounds do not reduce our ability to benefit from spatially separated speech and noise sources. In rooms, the addition of early-arriving reflections will increase the effective SNR and enhance the intelligibility of speech.

Adding reverberant speech sound does degrade the intelligibility of speech in noise. The addition of reverberant speech with a 1 s reverberation time had about the same magnitude of increase in SRT as did adding diffuse noise to the case B results with neither reverberation nor diffuse noise. That is, although they are completely independent effects, adding diffuse noise or adding reverberant speech in these cases led to about the same 4 dB increase in SRT values. When both diffuse noise and reverberant speech were included (case G), then the SRT was increased by about 8 dB or approximately the sum of the individual effects.

The effect of adding reverberant speech increased linearly with the logarithm of the reverberation time above a reverberation time of about 0.5 s. The addition of reverberant sound with a 0.5 s reverberation time did not significantly change the measured SRT relative to the case with only early-arriving reflections added to the speech sound. It is only for more reverberant conditions that the negative effects of reverberation became significant. In these experiments adding ERs and reverberation to the speech was accomplished while maintaining a constant overall speech level at

TABLE VII. SRT values and descriptions of conditions for 24 test configurations. Column "Expt." indicates to which of the four experiments each case belonged (each experiment used different subjects). Column "Wall" indicates whether a simulated wall was included or not. Column "Speech" indicates the composition of the speech signal: "Direct," direct sound only; "ER," direct sound and early-arriving reflections, "$T_{60}=0.5$," direct sound early-arriving reflections and reverberant sound with $T_{60}=0.5$ s; "$T_{60}=1$," direct sound early-arriving reflections and reverberant sound with $T_{60}=1.0$ s; "$T_{60}=2$," direct sound early-arriving reflections and reverberant sound with $T_{60}=2.0$ s. Column "Noise" indicates the composition of the noise signal: "Front," from only loudspeaker No. 1 directly in front of subject; "Front side," from only loudspeaker No. 3; "All," uncorrelated noise from all eight loudspeakers; "Ceiling," from only immediately overhead; "Rear side," from only loudspeaker No. 8; "Front-side diffuse," predominantly from loudspeaker Nos. 1, 3, and 6; and "Rear-side diffuse," predominantly from loudspeaker Nos. 6, 8.

SRT case	Expt.	Wall	Speech	Noise	SRT (dB)
A	1	No wall	Direct	Front	−9.37
B	1	No wall	Direct	Front side	−14.75
C	1	No wall	ER	Front side	−13.38
D	1	No wall	$T_{60}=1$	Front side	−10.27
E	1	No wall	Direct	All	−10.67
F	1	No wall	ER	All	−10.93
G	1	No wall	$T_{60}=1$	All	−6.82
H	2	No wall	Direct	Ceiling	−13.26
I	2	No wall	Direct	Front	−9.55
J	2	Wall	Direct	Ceiling	−12.46
K	2	Wall	Direct	Front	−8.01
L	3	Wall	ER	All	−8.79
M	3	Wall	$T_{60}=0.5$	All	−9.74
N	3	Wall	$T_{60}=1$	All	−5.71
O	3	Wall	$T_{60}=2$	All	−3.50
P	4	No wall	ER	All	−9.16
Q	4	No wall	$T_{60}=0.5$	All	−8.83
R	4	No wall	$T_{60}=1$	All	−6.42
S	4	No wall	$T_{60}=2$	All	−3.81
T	4	No wall	$T_{60}=1$	Rear side	−13.60
U	3	No wall	$T_{60}=1$	Rear-side diffuse	−9.47
V	3	No wall	$T_{60}=1$	Front-side diffuse	−8.71
W	3	Wall	$T_{60}=1$	Front-side diffuse	−6.54
X	3	Wall	$T_{60}=1$	Rear-side diffuse	−7.16

the position of the listener. In a real room, the addition of reflected sound would increase the overall level of the speech, which could further modify the intelligibility of speech depending on the relative amounts of early-arriving and later-arriving speech sounds.

In some cases the speech sounds were modified to simulate propagation through a wall and cases were compared both with and without the effect of a simulated wall. When this was done for cases with varied reverberation time and also for varied noise diffusion, there were no significant additional effects of adding a simulated wall. That is, the results apply equally well to natural speech as they do to speech filtered by propagation through a wall. Similarly in the validation study, there were no differences for the results using three different walls. However, this was only true when results were considered in terms of uniformly weighted signal-to-noise ratio ($\mathrm{SNR}_{\mathrm{uni32}}$) values and not for SNR measures with other frequency weightings.[19] The uniform frequency weighting of the $\mathrm{SNR}_{\mathrm{uni32}}$ measure seems to make it an ideal measure for assessing speech privacy conditions.

VI. CONCLUSIONS

Although speech privacy may be primarily a signal-to-noise issue, it is also significantly influenced by the spatial and temporal characteristics of sound in rooms. The two-room validation study results and the subsequent investigations of the various combinations of speech and noise showed that SRTs can be changed by 5 dB or more by the differences in the spatial relationships between speech and noise sources as well as the effects of reflected sound in rooms. As speech privacy is most often considered to correspond to reduced speech intelligibility, these effects are practically important and ignoring them could lead to expensive over-design of the sound insulation of meeting rooms, to provide the required speech privacy against eavesdroppers.

These new results have made it possible to revise previous estimates of the threshold of intelligibility from a $\mathrm{SNR}_{\mathrm{uni32}}$ value of −16 to a value of −11 dB to be more representative of conditions in typical meetings rooms. The new results also make it possible to estimate further modifications to speech privacy criteria to better suit particular combinations of room reverberation and spatial separations of speech and noise sources.

These results do not suggest that the audibility of speech in noise is influenced by the same spatial and temporal room acoustics effects. That is, the threshold of audibility of speech transmitted from an adjacent room is confirmed to correspond to a $\mathrm{SNR}_{\mathrm{uni32}}$ value of -22 dB as previously determined.[2]

ACKNOWLEDGMENTS

Most of the original experimental work and initial analyses was carried out by M.A. for her diploma thesis at the Fachhochschule für Technik in Stuttgart, Germany. Funding for this work was provided by Public Works and Government Services Canada (PWGSC) and the Royal Canadian Mounted Police (RCMP).

APPENDIX: SUMMARY OF THE TEST CONFIGURATIONS

The measured SRT values and details of the 24 measurement conditions are summarized in Table VII. The data were obtained from four different experiments that each used different subjects. The configurations included in each experiment are indicated in col. 2 of the table.

[1] W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, and B. G. Watters, “Speech privacy in buildings,” J. Acoust. Soc. Am. **34**, 475–492 (1962).

[2] B. N. Gover and J. S. Bradley, “Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms,” J. Acoust. Soc. Am. **116**, 3480–3490 (2004).

[3] V. O. Knudsen and C. M. Harris, *Acoustical Designing in Architecture* (Acoustical Society of America, 1980).

[4] D. D. Dirks and R. H. Wilson, “The effect of spatially separated sound sources on speech intelligibility,” J. Speech Hear. Res. **12**, 5–38 (1969).

[5] E. C. Cherry, “Some experiments on the recognition of speech with one and two ears,” J. Acoust. Soc. Am. **25**, 975–979 (1953).

[6] A. W. Bronkhorst, “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” Acust. Acta Acust. **86**, 117–128 (2000).

[7] M. Ebata, “Spatial unmasking and attention related to the cocktail party problem,” Acoust. Sci. & Tech. **24**, 208–219 (2003).

[8] N. R. French and J. C. Steinberg, “Factors governing the intelligibility of speech sounds,” J. Acoust. Soc. Am. **19**, 90–119 (1947).

[9] A. W. Bronkhorst and R. Plomp, “The effect of head-induced interaural time and level differences on speech intelligibility in noise,” J. Acoust. Soc. Am. **83**, 1508–1516 (1988).

[10] A. W. Bronkhorst and R. Plomp, “Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,” J. Acoust. Soc. Am. **92**, 3132–3139 (1992).

[11] J. Peissig and B. Kollmeier, “Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,” J. Acoust. Soc. Am. **101**, 1660–1670 (1997).

[12] J. Koehnke and J. M. Besing, “A procedure for testing speech intelligibility in a virtual listening environment,” Ear Hear. **17**, 211–217 (1996).

[13] R. Plomp, “Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound (speech or noise),” Acustica **34**, 200–211 (1976).

[14] D. F. Hoth, “Room noise spectra at subscribers’ telephone locations,” J. Acoust. Soc. Am. **12**, 449–504 (1941).

[15] W. E. Blazier, “Revised noise criteria for application in the acoustical design and rating of HVAC systems,” Noise Control Eng. **16**, 64–73 (1981).

[16] “IEEE recommended practice for speech quality measurements,” IEEE Trans. Audio Electroacoust. **17**, 227–246 (1969).

[17] J. S. Bradley and B. N. Gover, “Validation of architectural speech security results,” IRC/NRC Research Report No. RR-221, National Research Council, Ottawa, Canada, March 2006.

[18] ASTM Standard E2638-08, “Standard test method for objective measurement of the speech privacy of closed rooms,” ASTM International, West Conshohocken, PA, 2008.

[19] J. S. Bradley, M. Apfel, and B. N. Gover, “Spatial and temporal effects of room acoustics on the speech privacy of meeting rooms,” IRC/NRC Research Report No. RR-265, National Research Council, Ottawa, Canada, October 2008.

[20] J. S. Bradley, H. Sato, and M. Picard, “On the importance of early reflections for speech in rooms,” J. Acoust. Soc. Am. **113**, 3233–3244 (2003).

[21] B. Kollmeier and M. Wesselkamp, “Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment,” J. Acoust. Soc. Am. **102**, 2412–2421 (1997).

[22] H. Levitt and L. R. Rabiner, “Predicting binaural gain in intelligibility and release from masking for speech,” J. Acoust. Soc. Am. **42**, 601–608 (1967).

Determining biosonar images using sparse representations

Bertrand Fontaine[a] and Herbert Peremans
Active Perception Laboratory, Universiteit Antwerpen, 13 Prinsstraat, 2000 Antwerpen, Belgium

(Received 4 November 2008; revised 23 February 2009; accepted 24 February 2009)

Echolocating bats are thought to be able to create an image of their environment by emitting pulses and analyzing the reflected echoes. In this paper, the theory of sparse representations and its more recent further development into compressed sensing are applied to this biosonar image formation task. Considering the target image representation as sparse allows formulation of this inverse problem as a convex optimization problem for which well defined and efficient solution methods have been established. The resulting technique, referred to as $\mathcal{L}1$-minimization, is applied to simulated data to analyze its performance relative to delay accuracy and delay resolution experiments. This method performs comparably to the coherent receiver for the delay accuracy experiments, is quite robust to noise, and can reconstruct complex target impulse responses as generated by many closely spaced reflectors with different reflection strengths. This same technique, in addition to reconstructing biosonar target images, can be used to simultaneously localize these complex targets by interpreting location cues induced by the bat's head related transfer function. Finally, a tentative explanation is proposed for specific bat behavioral experiments in terms of the properties of target images as reconstructed by the $\mathcal{L}1$-minimization method.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3101485]

PACS number(s): 43.60.Ek, 43.80.Lb, 43.60.Pt [EJS] Pages: 3052–3059

I. INTRODUCTION

Bats emitting frequency modulated (FM) pulses can locate and recognize prey even in the presence of dense clutter. It has been argued in Simmons *et al.*, 1995 that bats do this by creating a detailed time-domain representation of the complex acoustic scene in front of them. Starting with the spectrogram correlation and transformation receiver (SCAT) model (Saillant *et al.*, 1993) a number of processing schemes (Matsuo *et al.*, 2001; Boonman and Ostwald, 2007) have been proposed to explain how "glint" based representations (Simmons and Chen, 1989) of such a target's biosonar image might be constructed from the received echo signals by bats. Glints are particularly strong echoes caused by favorably oriented reflecting surfaces on the target.

However, as noted in Wiegrebe, 2008, biosonar images expressed in terms of glints are not true target impulse responses, i.e., the sum of all reflections from a target when ensonified with a Dirac impulse, as they leave out the echo amplitude information, which can vary considerably for echoes from complex natural targets. While improved versions of this approach have been described (Matsuo *et al.*, 2004) that allow more accurate approximations of the target impulse response itself to be reconstructed, these methods, such as the SCAT model, break down when confronted with complex echo signals consisting of multiple (>3) very closely spaced reflections (Peremans and Hallam, 1998).

All these methods have in common that they perceive a biosonar image as expressed in terms of the target's profile along a range or time axis. An alternative biosonar image proposal (Wiegrebe, 2008), the "sonar image buffer," relies instead on a time-frequency representation as generated by the peripheral auditory system further processed by a sonar analysis model inspired by human pitch extraction to explain the results of various behavioral experiments. Nevertheless, it is noted in Grunwald *et al.*, 2004 that a true target impulse response representation might improve the fit between the models' predictions and certain experimental results in the classification of natural textures.

In this paper a new approach based on sparse representations (Chen *et al.*, 1998) is proposed to derive a true target impulse response from the echo signals received by bats. Introducing the sparsity assumption, i.e., the property that the target impulse response can be written as a linear superposition of a few well-chosen signals from a pre-specified dictionary, allows reformulation of this inverse filter problem as a convex optimization problem for which solutions can be efficiently generated using well understood algorithms (Kim *et al.*, 2007). It should be noted that we make no claim as to the biological plausibility of the specific algorithm used here. Instead, the sparsity assumption is seen as an important part of a "computational level" theory of biosonar as described in terms of Marr's theory of complex information processing systems (Marr, 1982). Indeed, the experimental results described below show that constraining the target impulse response to be sparse is a powerful way to regularize the ill-posed problem of finding a target's biosonar image given the received echo signal. In addition, the sparsity assumption proposed here is more general than the minimum phase and amplitude spectrum continuity assumptions that are introduced in Matsuo and Yano, 2004 for similar purposes. Furthermore, by extending sparse representations with compressed sensing techniques (Donoho, 2006), it is shown that one can achieve performance levels very similar to those shown by bats in behavioral experiments using biologically

[a] Author to whom correspondence should be addressed. Electronic mail: bertrand.fontaine@ua.ac.be

plausible measurement rates.

II. SPARSE REPRESENTATION OF ECHO SIGNALS

As a first approximation (Altes, 1976) the signal received by a bat can be considered as a linear superposition of echoes, each one a scaled and delayed version of the emitted call. If it is assumed that the phase of the emitted pulse is not changed upon reflection, a target image R can be entirely characterized by a sum of scaled and time-shifted Dirac impulses

$$R(t) = \sum A_i \delta(t - t_i), \qquad (1)$$

where A_i denotes the amplitude of the echo reflected by reflector i and $t_i = 2r_i / v_s$ with r_i denoting the range of reflector i and v_s the speed of sound. As the reflection coefficient is assumed positive (Matsuo *et al.*, 2004; Matsuo and Yano, 2004), a non-negativity constraint can be added: $R(t) \geq 0$, $\forall t$.

The measured echo $S(t)$ can be written as the convolution of the target impulse response and the emitted call $C(t)$. Hence, introducing sampled versions of the signals of interest, i.e., the column vector $S = [S(0T_s), S(1T_s), \ldots, S(nT_s)]^T$ denoting the received signal with n the total number of samples and T_s the sample period, and analogously for the vectors C and R denoting the emitted call and the target image respectively, yields

$$S = \mathbf{D}R, \qquad (2)$$

where $\mathbf{D}$ is a $n \times n$ matrix whose kth column contains the sampled version of the emitted call C delayed by k samples

$$\mathbf{D} = \begin{bmatrix} C(0T_s) & 0 & \cdots & 0 \\ C(1T_s) & C(0T_s) & \cdots & 0 \\ C(2T_s) & C(1T_s) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ C(mT_s) & C((m-1)T_s) & \cdots & 0 \\ 0 & C(mT_s) & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & C(0T_s) \end{bmatrix}. \qquad (3)$$

The matrix $\mathbf{D}$ is called the dictionary (Chen *et al.*, 1998) to build S from R.

For example (see Fig. 1), if the received signal contains two echoes with amplitudes 1 and 0.3, respectively, arriving at times kT_s and pT_s, the "simplest" way to build S from R using $\mathbf{D}$ is to set the kth and the pth samples of R equal to 1.0 and 0.3 and 0 for all the other ones. Because the number of echoes in S and thus the number of non-zero entries in R are small with respect to the number of samples n, R is said to be a sparse vector representation of S in the dictionary $\mathbf{D}$.

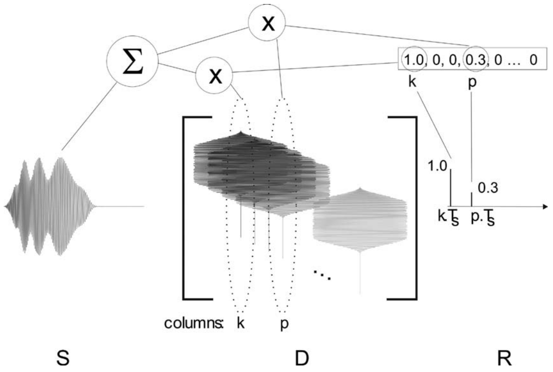


FIG. 1. The sparse representation R together with the dictionary $\mathbf{D}$ allows reconstruction of the received signal S.

Applying this approach to the interpretation of complex target echoes leads to an inverse problem, i.e., to estimate the representation R in a dictionary $\mathbf{D}$ using the measurements S. Assuming R to be sparse, this inverse problem can be solved by finding from among all possible target images the vector R that contains the minimum number of non-zero entries, i.e., finding the solution of the following 0-norm minimization problem:

$$\min \|R\|_0 \text{ subject to } S = \mathbf{D}R, \quad R \geqslant 0, \qquad (4)$$

with the 0-norm $\| \, \|_0$ counting the number of non-zero entries in a vector. This is a nondeterministic polynomial-time (NP)-hard problem since there is no way to solve it except testing all possible k-collections of columns of $\mathbf{D}$ with $k = 1, \ldots, n$, and looking for the smallest k-collection that synthesizes S. However, it has been shown (Candes and Romberg, 2006; Fuchs, 2004) that under appropriate conditions related to the dictionary $\mathbf{D}$ and the sparsity of the solution vector the 0-norm minimization problem can be restated as a 1-norm minimization problem with the same unique solution as the original problem. Hence, the problem is reformulated as

$$\min \|R\|_1 \text{ subject to } S = \mathbf{D}R, \quad R \geqslant 0. \qquad (5)$$

As the 1-norm $\|\mathrm{R}\|_1 = \Sigma_{i=1}^{n} |R_i|$ is a convex function the reformulated problem can be solved by well defined techniques such as linear programming. In other words, the intractable problem (4) can be solved efficiently using Eq. (5) provided the solution vector R is sufficiently sparse.

To make the analysis more realistic, white noise η is added to the measurements

$$S = \mathbf{D}R + \eta. \qquad (6)$$

Taking the noise into account, the inverse problem can be reformulated one more time (Fuchs, 2004) and stated it in its final form as

$$\min\left[\tfrac{1}{2}\|S - \mathbf{D}R\|_2^2 + \gamma \|R\|_1\right], \quad R \geqslant 0 \qquad (7)$$

with $\gamma > 0$. To solve this convex quadratic optimization problem, a specialized interior-point method, hereafter referred to as the $\mathcal{L}1$-minimization method, is used. A detailed explanation of this solution method is out of the scope of this paper (Kim *et al.*, 2007) but suffice it to note that it is also very efficient. The code used to solve the $\mathcal{L}1$-minimization was adapted from Koh *et al.* (2007).

In this final formulation, i.e., Eq. (7), the exact representation of S will not be found but a vector R with a small 1-norm resulting in a measurement vector that resembles S closely. The 1-norm penalty function puts relatively large weight on small residuals and so tends to produce residuals,

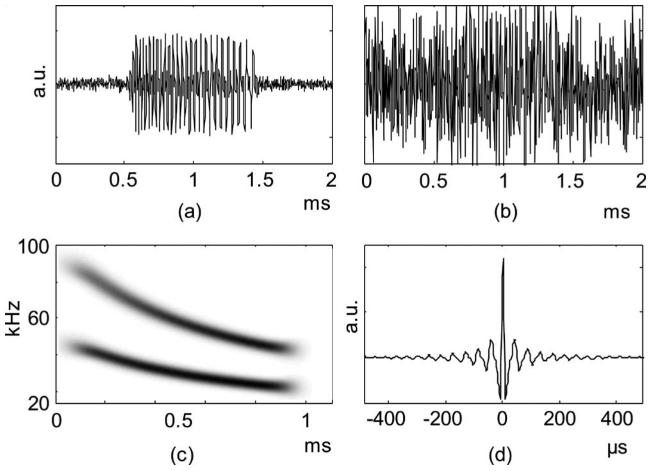


FIG. 2. Time pressure plots of the emitted call for (a) SNR=45 dB and (b) SNR=25 dB; (c) spectrogram and (d) autocorrelation function of the emitted call.

which are very small or even exactly zero. Hence, the main motivation of the 1-norm penalty term is that its use typically yields a sparse vector R. In contrast, the solution of the corresponding 2-norm minimization typically has all components non-zero. Note that the weight γ, which is the only free parameter, balances the influence of the noise with the sparsity of the solution vector. γ should be small for high signal to noise ratio (SNR) conditions and should increase along with the noise. A simple heuristic (Allgower and Georg, 1993) provides an upper bound for γ

$$\gamma \leqslant \gamma_{\max} = \|2D^T S\|_\infty. \tag{8}$$

Choosing γ as a fixed fraction of this bound, e.g., $\gamma = 0.1\gamma_{\max}$ in the experiments presented below, results in the correct noise dependent behavior.

III. SIMULATION RESULTS

A. Methods

As explained in the Sec. II (see Fig. 1) the columns of the dictionary $\mathbf{D}$ contain time-shifted versions of the emitted pulse. The properties of the emission used below are loosely based on those of the calls emitted by so-called FM-bats such as the big brown bat. It is a 1-ms duration, FM, sinusoidal signal with a fundamental component whose frequency sweeps down from 50 to 20 kHz and a first harmonic sweeping down from 100 to 40 kHz [Fig. 2(c)].

The amount of noise present in the measurements is characterized (Menne and Hackbarth, 1986) by the SNR defined as

$$\mathrm{SNR(dB)} = 10 \log_{10}\left(2\frac{E}{N_0}\right), \tag{9}$$

with E the energy of the echo signal (measured in Joule) and N_0 the spectral density of the noise (measured in Joule). Two examples of SNRs, one at 45 dB and the other at 25 dB, are plotted in Figs. 2(a) and 2(b). In bat echolocation processing, 5 dB SNR is considered low, 25 dB as moderate, and 45 dB as high (Neretti *et al.*, 2003).

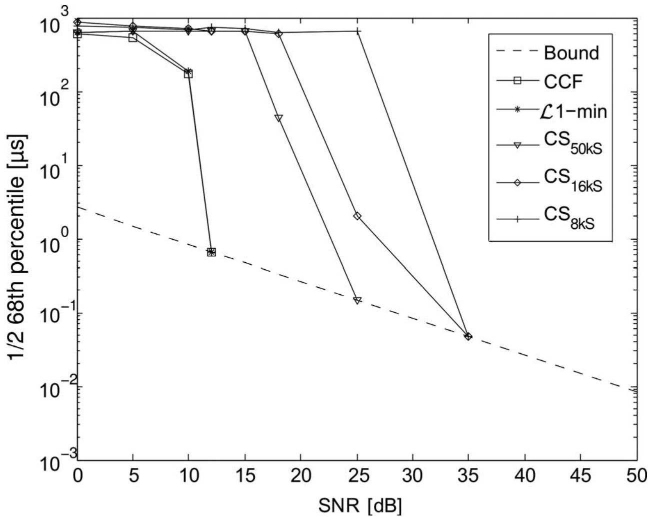


FIG. 3. Estimation of the range accuracy at different SNRs for the CCF receiver, ℒ1-minimization, and compressed sensing (CS) at three measurement rates (see discussion). The theoretical bound is also drawn.

B. Echo delay accuracy

Bats can estimate the distance to a target from the time delay between the received echo and the corresponding emitted call. Behavioral experiments have shown that the big brown bat can distinguish differences in echo delay as small as 50–100 μs (1–2 cm of range) in a two-choice discrimination test (Moss and Schnitzler, 1995) or even below 0.1 μs in jitter experiments (Simmons *et al.*, 1990). This performance measure is called the echo delay accuracy, i.e., the uncertainty in the estimate of the arrival time of the echoes from a single reflecting point. To reproduce this measurement situation in simulation the signal is seen as consisting of two components: one at the beginning of the signal, which is the picked up emission, and a delayed replica of this emitted pulse. We then analyze the uncertainty on the delay estimate as a function of SNR.

From radar theory it is known that a matched filter, i.e., a filter that computes the cross correlation function (CCF) between the noise-free emission pulse and the noisy echo, results in a minimal variance echo delay estimate (Van Trees, 2001). For relatively high SNR the uncertainty on the position of the peak in the CCF is given by (Menne and Hackbarth, 1986)

$$\sigma = (2\pi Bd)^{-1}, \tag{10}$$

where B is the non-centralized rms-bandwidth of the emitted pulse and $d = \sqrt{2E/N_0}$. In our simulations the peaks of the reconstructed target images are located in a priori windows (width=1000 μs) around the expected time of reflection. The delay is the difference between the estimated arrival time of the echo and the known time of emission.

The results of testing both methods when confronted with a target image consisting of a single Dirac impulse are shown in Fig. 3. Because of its reduced sensitivity to outliers the echo delay error in these experiments (200 trials) is characterized, as is customary, by estimating the 68th percentile P_{68} instead of the standard deviation σ. Note that the true values of these variables are related to $\sigma = P_{68}/2$ for a Gauss-

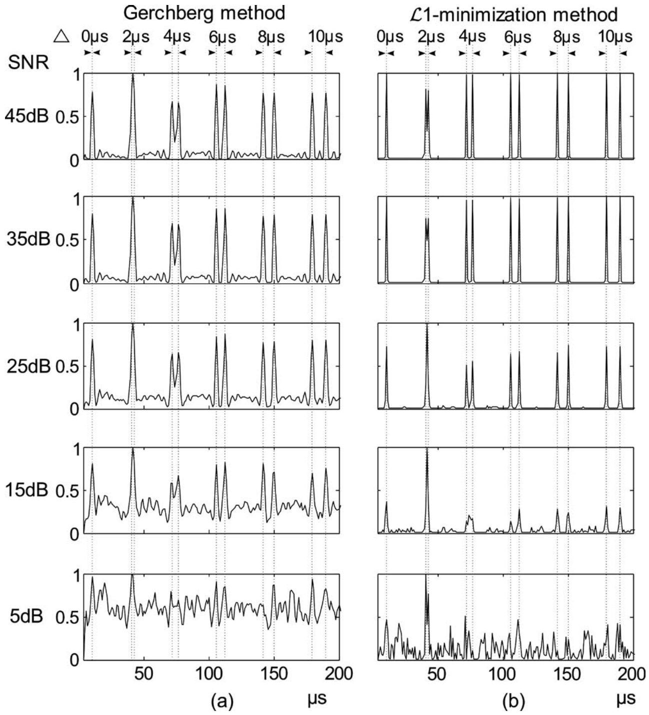


FIG. 4. Mean reconstructed target images (50 trials) for the (a) Gerchberg method and (b) $\mathcal{L}$1-minimization method at five SNR conditions. The true target image consists of 11 echoes at 10, 40, 42, 72, 76, 106, 112, 142, 150, 180, and 190 μs (vertical lines). The echoes all have the same amplitude.

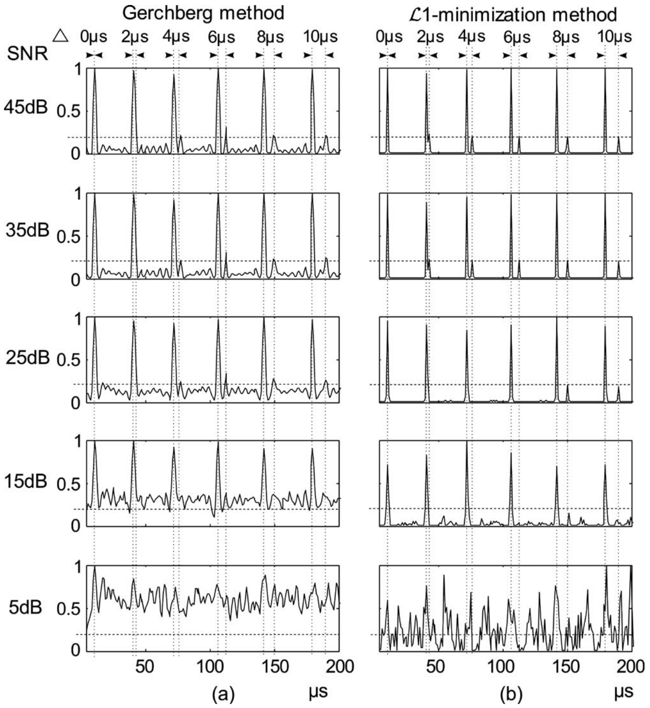


FIG. 5. Mean reconstructed target images (50 trials) for the (a) Gerchberg method and (b) $\mathcal{L}$1-minimization method at five SNR conditions. The true target image consists of 11 echoes at 10, 40, 42, 72, 76, 106, 112, 142, 150, 180, and 190 μs (vertical lines). The echoes have alternating amplitudes equal to 1 and 0.2.

ian random variable. These results clearly show that the $\mathcal{L}$1-minimization method's sensitivity to noise is equal to that of the CCF receiver. Both methods' performance is still significantly better than that measured in bats (Menne and Hackbarth, 1986).

C. Echo delay resolution

In this experimental setting the received signal consists of two closely spaced echoes in addition to the picked up emission. Hence, the true target image consists of two scaled Dirac impulses separated by a small delay. While the CCF receiver in combination with a peak detector is optimal (minimum variance criterion) for a signal containing isolated, i.e., well separated echoes, it becomes suboptimal for signals consisting of closely spaced echoes. Indeed, to extract a target image from the output of the CCF receiver in this case a suboptimal thresholding mechanism that can distinguish between peaks corresponding with true echoes and peaks corresponding with sidelobes has to be used (Peremans *et al.*, 1993). Hence, since the echo delay resolution task can be viewed as a target image reconstruction task, the Gerchberg method is chosen, a well established (Gerchberg, 1974; Papoulis, 1975) inverse filter technique, to compare the $\mathcal{L}$1-minimization approach with instead.

Figure 4 shows the mean target image reconstructions calculated from received signals consisting of pairs of equal amplitude echoes at delays=0, 2, 4, 6, 8, and 10 μs (vertical lines) for various SNR conditions. Figure 5 shows the mean target image reconstructions for signals consisting of pairs of unequal amplitude echoes. Note that while the echoes have been grouped in pairs of echoes separated by increasing delays both reconstruction methods have actually calculated a single target image containing all 11 echoes as this had no effect on either method's performance.

As can be seen by comparing the mean reconstructed target images (Fig. 4) with the individual results of all the 50 experiments that were conducted for each experimental condition shown in Fig. 6, the biosonar images reconstructed by both methods are very robust for high and medium SNR conditions, i.e., SNR$\geq$25 dB, but loose coherence quickly for lower SNR conditions, i.e., SNR$\leq$15 dB.

From these experiments it can concluded that the main advantage of the $\mathcal{L}$1-minimization method compared to the Gerchberg method is the significantly improved sparsity of the reconstructed target images resulting in the detection of far less spurious, i.e., non-existent, glints. Furthermore, the results in Fig. 5 show that the sparse nature of the target images reconstructed by the $\mathcal{L}$1-minimization method also allows a larger dynamic range in the strengths of the individual echoes the target image is comprised of. Indeed, the much reduced background activity makes it easier to recognize the peaks corresponding with true impulses in the target impulse response. Finally, for small echo separations ($<$4 μs), the Gerchberg method, independent of the SNR condition, is no longer capable of correct target image reconstruction. Given high SNR conditions, the $\mathcal{L}$1-minimization method on the other hand can correctly, i.e., both amplitude and position, resolve the individual Dirac impulses making up the target impulse response at the smallest delays.

To illustrate how this capacity of the $\mathcal{L}$1-minimization method for resolving closely spaced echoes translates into superior reconstruction performance when confronted with

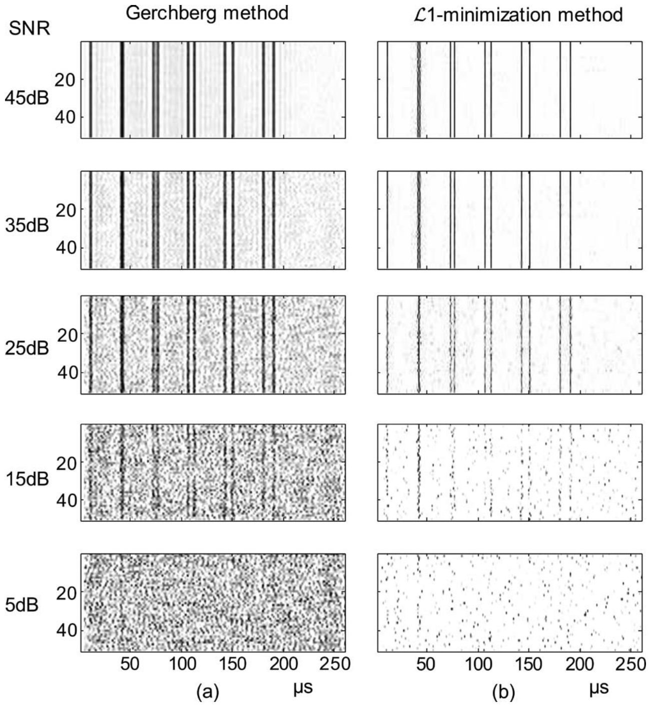


FIG. 6. Reconstructed images (50 trials) for the (a) Gerchberg method and (b) $\mathcal{L}$1-minimization method at five SNR conditions. The true target image consists of 11 echoes at 10, 40, 42, 72, 76, 106, 112, 142, 150, 180, and 190 μs. The echoes all have the same amplitude.

complicated targets Fig. 7 shows the reconstruction results for a received signal consisting of six closely spaced, i.e., within a time window of 30 μs, echoes with variable amplitudes, i.e., ranging from 0.2 to 1.0.

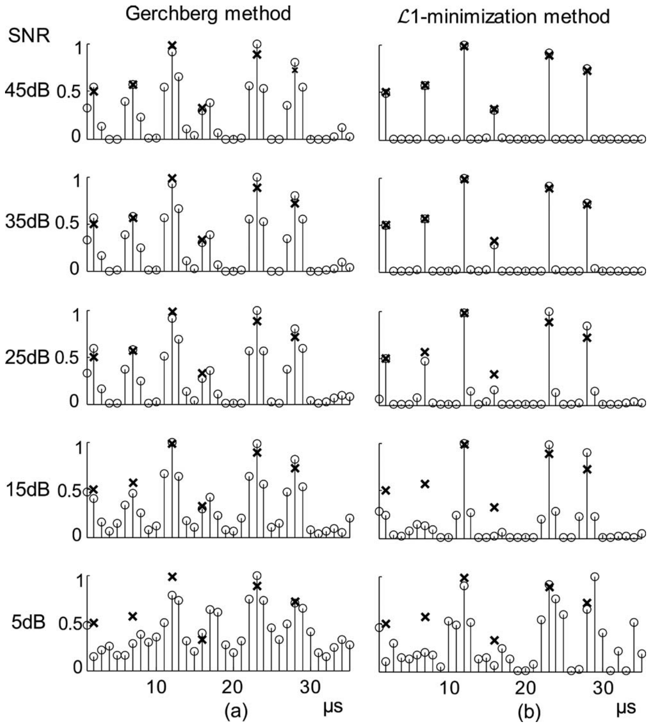


FIG. 7. Mean reconstructed target images (50 trials) for the (a) Gerchberg method and (b) $\mathcal{L}$1-minimization method at five SNR conditions. The true target image consists of six closely spaced echoes with different strengths (true target impulse response indicated by crosses).

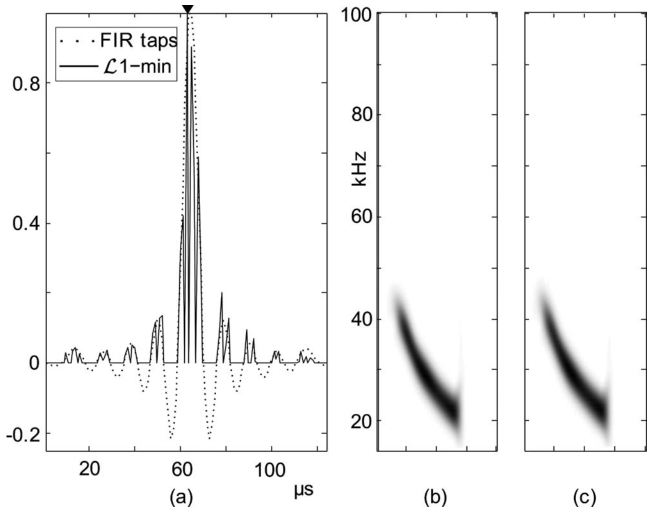


FIG. 8. (a) Estimated target image for a single echo (true position indicated by filled triangle) containing only the fundamental component and the true low-pass filter impulse response. (b) Spectrogram of the emitted call convolved with the estimated target image. (c) Spectrogram of received echo signal.

D. Filtered echoes

In reality, a received echo signal, even if containing only a single echo, is not a scaled and delayed copy of the emitted call, i.e., the target image is not a delayed Dirac impulse. Even if only minimal filtering occurs during emission, reflection, and reception, the air itself will always act as a low-pass filter (Kinsler *et al.*, 1999). Hence, to illustrate the sensitivity of $\mathcal{L}$1-minimization to unmodeled filtering, additional experiments are presented with a filtered version of the emitted call. The dictionary **D** is kept unaltered, i.e., based on a 1-ms duration FM sinusoidal signal with a fundamental component sweeping down from 50 to 20 kHz and a first harmonic sweeping down from 100 to 40 kHz. The received echo contains only the fundamental component, i.e., the echo is a low-pass filtered version of the emitted call. As the received signal consists of a single echo only, the estimated target image should consist of a single delayed Dirac impulse if $\mathcal{L}$1-minimization were perfectly insensitive to such unmodeled filtering. However, as can be seen in Fig. 8(a) for the low-pass filtered echo, the resulting target image consists of multiple Dirac impulses distributed around the true delay. A comparison of the received signal and the convolution of the estimated target image with the emitted call [Fig. 2(c)], as shown in Figs. 8(b) and 8(c), indicates that the applied filter operation is indeed well approximated by the set of weighted and shifted Dirac impulses in the reconstructed target image. It should be noted that the (lack of) sparsity of the reconstructed target image is strongly affected by the choice of directory **D** as the $\mathcal{L}$1-minimization method is constrained to approximate the observed echo filtering in terms of a weighted (positive valued) sum of dictionary elements. This effect is clearly illustrated by superimposing the true filter impulse response on the reconstructed target impulse response as shown in Fig. 8(a). Hence, if the dictionary were to contain appropriately filtered (Altes, 1976) versions of the emitted call—an example of such a situation will be shown

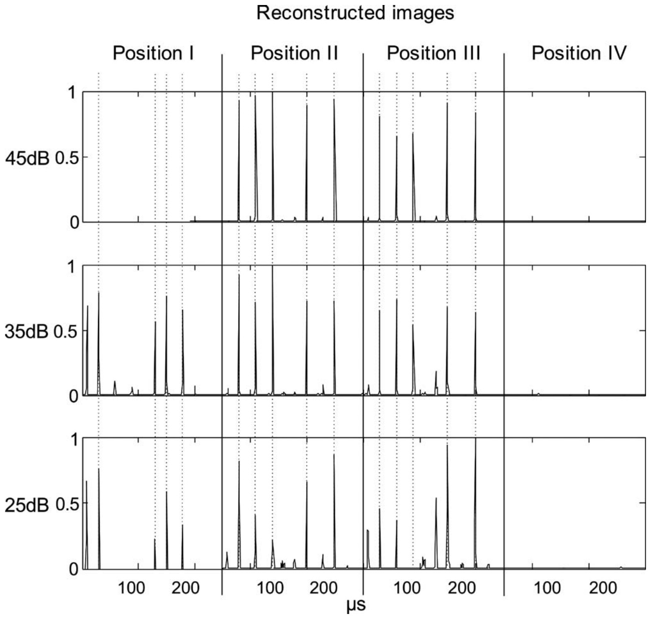


FIG. 9. Mean (50 trials) of the reconstructed images ($\mathcal{L}1$-minimization method) at three SNR conditions. The three true target images consist of five echoes each at the positions indicated by the vertical lines. The echoes all have the same amplitude.

in Sec. IV—the target image can be reconstructed in a sparser way.

IV. DISCUSSION

From spatial hearing theory it is known that upon arrival at the listener's head the received acoustic signals are filtered by the head related transfer function (HRTF) (Blauert, 1997). In particular, for FM-bats, this HRTF-induced filtering is considered to provide important spatial cues (Aytekin *et al.*, 2004) that help the bat localize its target. Hence, from the experiments with filtered echoes described above, it can be concluded that applying the $\mathcal{L}1$-minimization approach to bat echolocation requires a dictionary that takes the HRTF into account. To illustrate the fact that such an approach allows one to simultaneously determine the target biosonar image and the target's position the results are shown for an experimental setup consisting of three targets at positions $[\text{azimuth},\text{elevation}]=([-20°,-20°],[0°,20°],[0°,0°])$ with the target biosonar images given by equal amplitude Dirac impulses at [10,30,130,150,180], [30,60,90,150,200], and $[30,60,90,150,200]$ μs, respectively. Note that the second and third biosonar images are identical. Hence, in this case, the HRTF-induced filtering is a necessary condition to allow disambiguating the echoes belonging to the second and the third target (Asari *et al.*, 2006).

The new dictionary $\mathbf{D}'$ is built up by concatenating matrices corresponding to the dictionary $\mathbf{D}$ used above. Each matrix D_i is filtered by the HRTF corresponding to position i.

$$\mathbf{D}' = |\mathbf{D}_1| \cdots |\mathbf{D}_i|\mathbf{D}_k|. \qquad (11)$$

First, the results shown in Fig. 9 indicate that for high SNR conditions the $\mathcal{L}1$-minimization method is still capable of target impulse response reconstruction, i.e., Dirac impulses positioned accurately with approximately correct amplitudes, despite the presence of extensive HRTF-induced filtering, provided an appropriate dictionary is used. Furthermore, these results also show that target image reconstruction can be combined with reliable target localization based on HRTF-induced spectral cues. To illustrate this more clearly the dictionary used in the experiment was extended with a fourth filtered copy of the original dictionary $\mathbf{D}$ corresponding with a fourth possible target position $[\text{azimuth},\text{elevation}]=([-35°,-35°])$. As no target was present at this location the $\mathcal{L}1$-minimization method reconstructed an empty target image (position IV in Fig. 9) at this position, as it should.

The standard approach to compress signals is to first acquire the signals at full Nyquist rate, next map them on a well-chosen set of basis vectors, and then finally retain and encode the largest coefficients only. The theory of compressed sensing (Donoho, 2006) shows that a number of inefficiencies associated with this approach, i.e., collecting and transforming large amounts of data and then throwing away most of them during compression, can be avoided. In particular, it is shown that an unknown signal that is compressible by a known transform can be measured at fewer than the nominal number of sample points and yet be accurately reconstructed. Indeed, if the measurement samples are appropriately chosen linear combinations of the transform coefficients an approximate reconstruction of the signal can be obtained by solving for the transform coefficients consistent with the measured data that have the minimal $\mathcal{L}1$-norm, i.e., the sparse representation of the signal.

For the experimental results shown so far all data were collected and processed at a sample rate of 1 Msamples/sc. To illustrate the power of the $\mathcal{L}1$-minimization method in the context of compressed sensing the reconstructed biosonar target images are shown in Fig. 10 when the sample rate at which the received signal is acquired is lowered from 100 ksamples/s to only 3 ksamples/sc. In these experiments the columns of the new compressed sensing dictionary $\mathbf{D}_{\text{CS}}$ are set equal to correspondingly down-sampled versions of the emission (Baraniuk and Steeghs, 2007) used in the experiments described in Secs. I–III. Note that the call duration is lengthened to 5 ms to ensure enough sample even at low measurement rates. From these results it is concluded that for high SNR conditions the correct biosonar target images can be reconstructed at a precision much higher than that of the measurements. Indeed, a Dirac impulse can be localized in time with greater precision than the Δt between time samples. This illustrates the $\mathcal{L}1$-minimization method's potential for hyper-acuity performance. Furthermore, it can be shown (Baraniuk and Steeghs, 2007) that the vocalizations, i.e., FM-chirps, used by bats are instrumental in making such compressed sensing possible. Hence, the present authors conjecture that the use of FM-chirps in combination with the possibility to sparsely represent natural biosonar target images in an appropriately chosen basis provides a means to bridge the gap that exists between the apparent availability to bats of high resolution biosonar target images (microseconds scale) and the much slower physiologically plausible timescales (hundreds of microseconds up to milliseconds scale). Indeed, from the results presented in Fig. 2, it can be con-

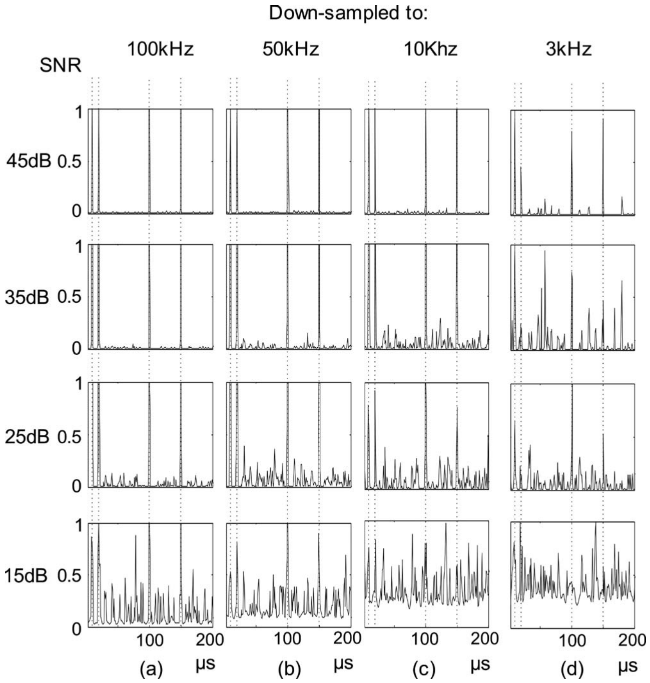


FIG. 10. Mean (50 trials) of the reconstructed images ($\mathcal{L}1$-minimization method) at four SNR conditions. Measured data are sampled at (a) 100, (b) 50, (c) 10, and (d) 3 ksamples/s.

cluded that the $\mathcal{L}1$-minimization method gives rise to a significantly higher accuracy than shown by bats in echo delay experiments (Menne and Hackbarth, 1986). However, these results were derived for a biologically implausible measurement rate of 1 Msamples/s. Figure 2 also shows how using the $\mathcal{L}1$-minimization method in a compressed sensing setting results in decreasing echo delay accuracy for decreasing measurement rates, i.e., sample rates of 50 ksamples/s down to 8 ksamples/s. To make the comparison with the bat behavioral experiments more explicit the delay difference at 75% correct choices as a function of SNR for different measurement rates is shown in Fig. 11. Taking 60 μs as average bat performance figure for SNR between 30 and 40 dB (Menne and Hackbarth, 1986; Wiegrebe, 2008) it can be seen that comparable performances are reached for measurement rates between 16 and 8 ksample/s. Hence, the delay accuracy results can be brought in line with performance levels shown by bats in behavioral experiments if the measurement rates are made more physiologically plausible.

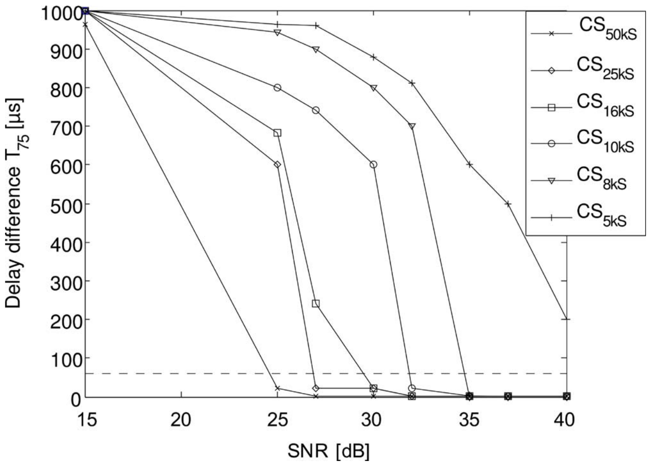


FIG. 11. (Color online) Delay difference at 75% correct choice for the simulated range as a function of the SNR. The full lines correspond with the results for CS with measurement rates of 50, 25, 15, 10, 8, and 5 ksamples/s. The dashed line corresponds to the delay difference measured for bats in behavioral experiments (60 μs).

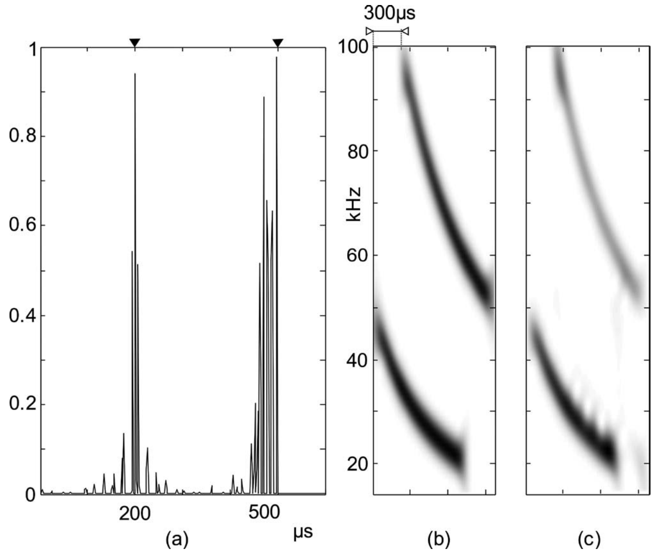


FIG. 12. (a) Estimated target image for a single echo (positions of harmonic components indicated by filled triangles) containing a harmonic component shifted over 300 μs with respect to the fundamental component. (b) Spectrogram of received echo signal. (c) Spectrogram of the emitted call convolved with the estimated target image.

As illustrated in Sec. III (see Fig. 8) a loss of sparsity occurs from the use of a simple dictionary as the extra filtering, in addition to the filter formed by the target impulse response, performed on the echo signals is approximated by taking weighted sums of dictionary elements. We conjecture that the resulting loss of sparsity of such a reconstructed biosonar target image might explain the behavioral experiments described in Stamper *et al.*, 2008 that presented FM-bats with echoes consisting of time-shifted harmonic components. In the simulation of these experiments the harmonic component of the FM-chrip is shifted with respect to the fundamental component over 300 μs [see Fig. 12(b)] and reconstructed the biosonar target image with the standard dictionary **D**. The result shown in Fig. 12(a) indicates that, as is the case for a filtered echo, spurious impulses are introduced into the reconstructed biosonar target image. In particular, in addition to a large impulse at the position of the fundamental component of the echo, a large extra impulse is present at the position of the shifted harmonic component as well as many smaller spurious impulses at intermediate positions. We note that the shape of this reconstructed biosonar target image, i.e., two large peaks at the positions of the fundamental and the harmonic components against a background of smaller peaks, shows an intriguing resemblance to the performance curve of the bats during these behavioral experiments.

From the evidence presented above the conclusion is drawn that the use of sparse representations for biosonar target reconstruction is quite powerful. To determine whether bat echolocation is indeed based on such sparse representation and compressed sensing techniques requires both specific behavioral experiments to test this hypothesis as well as more detailed implementation proposals that take into account what is known about the physiology of the bat hearing system.

ACKNOWLEDGMENT

This work was funded through the CILIA project (EU-FET).

Allgower, E., and Georg, K. (**1993**). "Continuation and path following," Acta Numerica **2**, 1–64.

Altes, R. (**1976**). "Sonar for generalized target description and its similarity to animal echolocation systems," J. Acoust. Soc. Am. **59**, 97–105.

Asari, H., Pearlmutter, B., and Zador, A. (**2006**). "Sparse representations for the cocktail party problem," J. Neurosci. **26**, 7747–7490.

Aytekin, M., Grassi, E., Sahota, M., and Moss, C. (**2004**). "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," J. Acoust. Soc. Am. **116**, 3594–3605.

Baraniuk, R., and Steeghs, P. (**2007**). "Compressive radar imaging," Proceedings of IEEE Radar Conference, pp. 128–133.

Blauert, J. (**1997**). *Partial Hearing* (MIT, Cambridge, MA).

Boonman, A., and Ostwald, J. (**2007**). "A modeling approach to explain pulse design in bats," Biol. Cybern. **97**, 159–172.

Candes, E., and Romberg, J. (**2006**). "Quantitative robust uncertainty principles and optimally sparse decompositions," Found Comput. Math. **6**, 227–254.

Chen, S., Donoho, D., and Saunders, M. (**1998**). "Atomic decomposition by basis pursuit," SIAM J. Sci. Comput. (USA) **20**, 33–61.

Donoho, D. (**2006**). "Compressed sensing," IEEE Trans. Inf. Theory **52**, 5406–5425.

Fuchs, J.-J. (**2004**). "On sparse representations in arbitrary redundant bases," IEEE Trans. Inf. Theory **60**, 1341–1344.

Gerchberg, R. (**1974**). "Super-resolution through error energy reduction," Opt. Acta **21**, 709–720.

Grunwald, J.-E., Schörnich, S., and Wiegrebe, L. (**2004**). "Classification of natural textures in echolocation," Proc. Natl. Acad. Sci. U.S.A. **101**, 5670–5674.

Kim, S.-J., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. (**2007**). "A method for large-scale l1-regularized least squares," IEEE J. Sel. Top. Sign. Process. **1**, 606–617.

Kinsler, L., Frey, A., Coppens, A., and Sanders, J. (**1999**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).

Koh, K., Kim, S. J., and Boyd, S. (**2007**). "l1_ls: Simple Matlab solver for l1-regularized least squares problems," URL: http://www.stanford.edu/~boyd/l1_ls/ (Last viewed 11/4/2008).

Marr, D. (**1982**). *Vision* (Freeman, New York).

Matsuo, I., Kunugiyama, K., and Yano, M. (**2004**). "An echolocation model for range discrimination of multiple closely spaced objects: Transformation of spectrogram into the reflected intensity distribution," J. Acoust. Soc. Am. **115**, 920–928.

Matsuo, I., Tani, J., and Yano, M. (**2001**). "A model of echolocation of multiple targets in 3d space from a single emission," J. Acoust. Soc. Am. **110**, 607–624.

Matsuo, I., and Yano, M. (**2004**). "An echolocation model for the restoration of an acoustic image from a single-emission echo," J. Acoust. Soc. Am. **116**, 3782–3788.

Menne, D., and Hackbarth, H. (**1986**). "Accuracy of distance measurement in bat eptesicus fuscus: Theoretical aspects and computer simulations," J. Acoust. Soc. Am. **79**, 386–397.

Moss, C., and Schnitzler, H.-U. (**1995**). "Behavioral studies of auditory information processing," in *Hearing by bats*, edited by A. Popper and R. Fay (Springer-Verlag, New York), pp. 87–145.

Neretti, N., Sanderson, M., Intrator, N., and Simmons, J. (**2003**). "Time-frequency model for echo-delay resolution in wideband biosonar," J. Acoust. Soc. Am. **113**, 2137–2145.

Papoulis, A. (**1975**). "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **22**, 735–742.

Peremans, H., and Hallam, J. (**1998**). "The spectrogram correlation and transformation receiver, revisited," J. Acoust. Soc. Am. **104**, 1101–1110.

Peremans, H., Audenaert, K., and Van Campenhout, J. (**1993**). "A high-resolution sensor based on tri-aural perception," IEEE Trans. Rob. Autom. **9**, 36–48.

Saillant, P., Simmons, J., Dear, S., and McMullen, T. (**1993**). "A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver," J. Acoust. Soc. Am. **94**, 2691–2712.

Simmons, J., and Chen, L. (**1989**). "The acoustic basis for target discrimination by fm echolocating bats," J. Acoust. Soc. Am. **86**, 1333–1350.

Simmons, J., Ferragamo, M. J., Moss, C., Stevenson, S. B., and Altes, R. (**1990**). "Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: The shape of target images in echolocation," J. Comp. Physiol. [A] **95**, 589–616.

Simmons, J., Saillant, P., Wotton, J., Haresign, T., Ferragamo, M., and Moss, C. (**1995**). "Composition of biosonar images for target recognition by echolocating bats," Neural Networks **8**, 1239–1261.

Stamper, S., Bates, M., Benedicto, D., and Simmons, J. A. (**2008**). "Degradation of fm-bat echo delay acuity from misaligned harmonics," Second International Conference on Acoustic Communication by Animals, Corvalis, OR, pp. 259–260.

Van Trees, H. (**2001**). *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York).

Wiegrebe, L. (**2008**). "An autocorrelation model of bat sonar," Biol. Cybern. **98**, 587–595.

Scattering effect on the sound focused personal audio system

Ji-Ho Chang,[a] Jin-Young Park, and Yang-Hann Kim
Department of Mechanical Engineering, Center for Noise and Vibration Control, Korea Advanced Institute of Science and Technology, Science Town, Daejeon 305-701, Korea

(Received 2 December 2008; revised 19 February 2009; accepted 22 February 2009)

Recently, a personal audio system has been studied that uses an array of loudspeakers to localize sound to only the area around a user. To realize this system, beamforming or acoustic contrast control has been applied on the assumption that sources radiate sound in a free-field. This means that not only reflection by walls, but also the scattering effect by the user's head is neglected. Reflection by walls is negligible because personal devices are usually used in a short distance so that direct sound is dominant over reverberant sound. However, the scattering effect by the user's head has a considerable effect on the focused sound field. For example, the region where sound energy is not focused becomes louder when a user is actually in the focused region due to the scattered sound by the user's head in the focused region. In this paper, the scattering effect is shown computationally on the simple assumption that the user's head is a rigid sphere. Then, an improving control method, which overcomes this effect, is proposed. The method is shown to outperform the previous method in terms of lowering the sound level in the side regions when a user is in the bright zone.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3101453]

PACS number(s): 43.60.Fg, 43.38.Hz [TDM] Pages: 3060–3066

I. INTRODUCTION

Recently, a sound focused personal audio system[1] has been studied. It localizes sound to only the area around a user without producing annoying noise to others nearby, making uncomfortable earphones or headsets unnecessary. This system could potentially impact audio industries because personal devices such as desktops, laptop computers, portable multimedia players, and cellular phones are widely used.

To test whether or not the system is practically applicable, a linear array of nine loudspeakers attached to a 17 in. monitor was selected in a previous study.[1] Then, acoustic contrast control[2] was applied to maximize the acoustic potential energy ratio between the bright and the dark zones. The bright zone is defined as a zone where we want to enhance the sound level, and the dark zone is defined as a zone we want to make quiet.[2] In the selected example,[1] the bright zone was the front region of the loudspeaker array and the dark zone was the side region. Then, experiments were performed in the free-field condition where reflection from walls and scattering due to the user were not considered. The energy ratio between the bright and dark zones is regarded as acoustic contrast.[2] The results[1] showed more than the nearly 20 dB of acoustic contrast between the bright and the dark zones in the frequency range of 800 Hz–5 kHz, and the contrast increased with respect to frequency, as expected; in the higher frequency, we have higher directivity. The contrast control was shown to be superior to the uncontrolled case and the time reversal array.[1]

However, in practice, the performance of the system is worse than expected. When a user enters the bright zone, the focused sound waves are scattered by the user's head, and then the dark zone becomes louder. It means that this degrades the advantage of the proposed personal audio system. Therefore, we aim to overcome this problem by taking the scattering effect into account. In other words, the objective of this study is to improve the performance of the system so that it does not produce noise in the dark zone in spite of the scattering effect.

Diffraction or scattering effect due to the user's head and upper body has been studied by many of researchers relating to the head related transfer function.[3,4] However, these studies are mainly interested in the change in the sound, which reaches the user's ears, while the present paper is interested in the change in sound level in the outer regions. On the other hand, Jones and Elliott[5,6] investigated the personal audio system that is applicable in adjacent seats in an aircraft or road vehicle, and they used dummy heads to consider the diffraction or scattering effect due to the user's head and upper body. However, they did not deal with the diffraction or scattering effect itself.

The paper is organized as follows. In Sec. II, the problem is defined and in Sec. III, we show how sound fields are changed by the scattering effect. Then in Sec. IV, the optimal solution under the scattering condition is obtained.

II. PROBLEM DEFINITION

A. Basic assumptions

For simplicity, we assume that there is no reflection from walls and the surface of the monitor. This is practically acceptable if the user uses the device in a shorter distance than the radius of reverberation so that direct sound is dominant compared with reverberant sound.

Figure 1 illustrates the physical configuration of the system that we selected: the loudspeaker array and the rigid

[a] Author to whom correspondence should be addressed. Electronic mail: chang.jiho@gmail.com

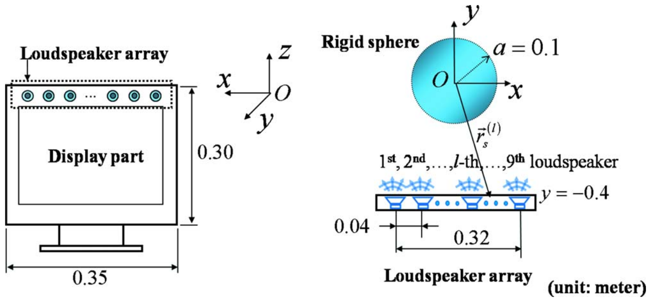


FIG. 1. (Color online) The loudspeaker array and the rigid sphere; considering the scale of the 17 in. monitor display (the width is 0.35 m and the height is 0.30 m). We regard the aperture of the loudspeaker array to be 0.32 m (left). We used nine control sources with equal spacing, 0.04 m, between adjacent sources. The user's head is regarded as a rigid sphere whose radius a is 0.1 m, and the distance between the loudspeaker array and the rigid sphere is regarded as 0.4 m considering the normal distance from the user to the monitor (right).

sphere. As used in the previous study,[1] considering the geometrical scale of a 17 in. monitor display (the width is 0.35 m and the height is 0.30 m), we regard the aperture of the loudspeaker array to be 0.32 m. We used nine control sources with equal spacing, 0.04 m, between adjacent sources.

The user's head is regarded as a rigid sphere whose radius a is 0.1 m, meaning that the details of the head and the upper body are neglected. This is acceptable in this study by the following reasons: First, the region of interest is a horizontal plane including the location of the array of loudspeakers, and then the reflection by the upper body on the plane is negligible compared with the user's head. Second, this study is interested in the sound level change in the dark zone, and the effect of head detail to the level change is negligible because the wavelength in the frequency range of interest can be regarded to be larger than the head details. The distance between the loudspeaker array and the center of the rigid sphere is regarded as 0.4 m considering the normal distance from the user to the monitor. Let us assume that the origin of the coordinate corresponds with the center of the rigid sphere. The position of the lth loudspeaker $\boldsymbol{r}_s^{(l)}$ is determined as

$$\begin{aligned}\boldsymbol{r}_s^{(l)} &= (x_s^{(l)}, y_s^{(l)}, 0)\\ &= (-0.16 + 0.04(l-1), -0.4, 0) \quad (\text{meter}), \quad l\\ &= 1, 2, \ldots, 9. \qquad (1)\end{aligned}$$

The frequency range is from 800 Hz to 5 kHz, which is determined by the characteristics of the loudspeakers that were used in the previous study; the coherence function between the input voltage into the loudspeaker and the measured pressure at a point in front of the loudspeaker is nearly 1 in the frequency range, meaning that the loudspeaker has a linear response with respect to the input signal.

B. The incident and scattered fields

When an incident wave meets an obstacle, it is scattered. The scattered sound field is normally defined as the difference between the changed field and the incident field. The changed field is often called the total field because it is the sum of the incident and scattered fields.

On the surface of the rigid sphere that indicates the user's head, the particle velocity perpendicular to the surface is zero. Then, the total field by the lth loudspeaker, $p_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)$, satisfies the following boundary condition:

$$\frac{\partial}{\partial r}[p_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)]_{r=a} = 0, \quad l = 1, 2, \ldots, 9. \qquad (2)$$

By applying this boundary condition, each total field by each loudspeaker is obtained. In the Appendix, we provide details on the total fields in the case that loudspeakers are regarded as monopoles.

III. THE CHANGE IN SOUND FIELDS BY THE SCATTERING EFFECT

When there is no scatterer, the sound field generated by all loudspeakers is expressed by

$$p(\boldsymbol{r},\omega) = \sum_{l=1}^{9} p_{\text{inc}}^{(l)}(\boldsymbol{r},\omega) = \sum_{l=1}^{9} q^{(l)}(\omega) h_{\text{inc}}^{(l)}(\boldsymbol{r},\omega), \qquad (3)$$

where $q^{(l)}(\omega)$ is the amplitude of the input signal fed into the lth loudspeaker, $p_{\text{inc}}^{(l)}(\boldsymbol{r},\omega)$ is the incident field generated by the lth loudspeaker, and $h_{\text{inc}}^{(l)}(\boldsymbol{r},\omega)$ is the transfer function from the input signal $q^{(l)}(\omega)$ to the incident sound field $p_{\text{inc}}^{(l)}(\boldsymbol{r},\omega)$. On the other hand, the sound field created by all loudspeakers with the rigid sphere is

$$p(\boldsymbol{r},\omega) = \sum_{l=1}^{9} p_{\text{tot}}^{(l)}(\boldsymbol{r},\omega) = \sum_{l=1}^{9} q^{(l)}(\omega) h_{\text{tot}}^{(l)}(\boldsymbol{r},\omega), \qquad (4)$$

where $h_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)$ is the transfer function from the input signal $q^{(l)}(\omega)$ to the total field $p_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)$.

It is noteworthy that substituting $h_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)$ for $h_{\text{inc}}^{(l)}(\boldsymbol{r},\omega)$ with the same solution $q^{(l)}(\omega)$ allows us to see how the sound field created by all loudspeakers is changed by the rigid sphere. The transfer function $h_{\text{inc}}^{(l)}(\boldsymbol{r},\omega)$ and the total field $h_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)$ are calculated or measured, and then a sound field by all loudspeakers depends on the input signals that are fed into the loudspeakers, $q^{(l)}(\omega)$.

A. The solution without the scattering effect

First, let us obtain the optimal solution in which the scattering effect is not considered. Let us denote it by

$$\mathbf{q}_1(\omega) = [q_1^{(1)}(\omega) q_1^{(2)}(\omega) \cdots q_1^{(9)}(\omega)]^T, \qquad (5)$$

where subscript 1 indicates that this solution is obtained without considering the scattering effect, while subscript 2 is used to indicate the solution considering the scattering effect that is explained in Sec. IV.

We obtain the solution by applying acoustic contrast control[2] in the incident fields. We consider a two-dimensional (2D) plane, x-y plane where $z=0$, because we cannot control the sound field around the axis independently by the line array, which has dependency around the axis of the array (x axis). Then, the bright and the dark zones are determined on the 2D plane. The bright zone is determined by a square in front of the array to include the user's head, as illustrated in Fig. 2, and can be expressed by

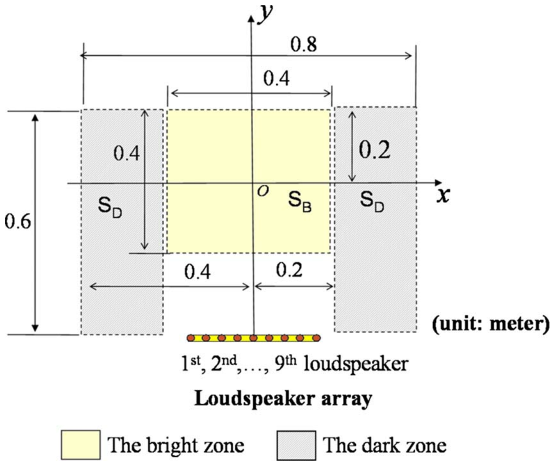


FIG. 2. (Color online) The bright and the dark zones; the bright zone is determined by a square in front of the array to include the user's head, and the dark zone is determined by rectangles at the sides to block sound waves from the array.

$$S_{B,1} = \{(x,y) | -0.2 \le x \le 0.2,\ \ 0.2 \le y \le 0.6\} \quad (\text{unit:meter}). \tag{6}$$

The dark zone is determined by rectangles at the sides to block sound waves from the array and they can be expressed by

$$S_{D,1} = \{(x,y) | -0.4 \le x \le -0.2 \ \text{ or } \ 0.2 \le x \le 0.4,$$
$$0 \le y \le 0.6\} \quad (\text{unit:meter}). \tag{7}$$

Let us determine the measurement spacing Δx and Δy to be 0.02 m, which is smaller than half of the wavelength for the entire frequency range to avoid the spatial sampling problem. As illustrated in Fig. 3, the discrete points can be written as

$$(x_m, y_n) = (-0.41 + 0.02m, \quad -0.01 + 0.02n) \quad (\text{meter}), \tag{8}$$

where $m = 1, 2, \ldots, 40$ and $n = 1, 2, \ldots, 30$.

Let us denote the transfer function from the input signal to the incident sound field at (x_m, y_n) by each loudspeaker as a vector $\mathbf{h}_{\text{inc}}(x_m, y_n, \omega)$,

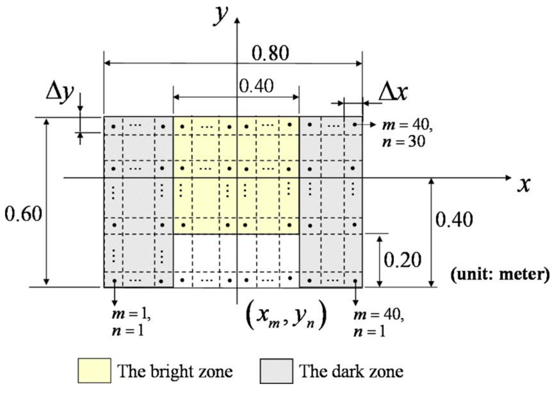


FIG. 3. (Color online) The discretized sound field; an arbitrary position is expressed by $(x_m, y_n) = (-0.41 + 0.02m, \ -0.01 + 0.02n)$ (meter), where $m = 1, 2, \ldots, 40$ and $n = 1, 2, \ldots, 30$.

$$\mathbf{h}_{\text{inc}}(x_m, y_n, \omega) = [h_{\text{inc}}^{(1)}(x_m, y_n, \omega) \ \ h_{\text{inc}}^{(2)}(x_m, y_n, \omega) \ \ \cdots \ \ h_{\text{inc}}^{(9)}(x_m, y_n, \omega)]. \tag{9}$$

Then, the spatial correlation matrices are written in the discrete form

$$\mathbf{R}_{B,1}(\omega) = \frac{1}{0.4^2} \sum_{m=11}^{30} \sum_{n=11}^{30} \mathbf{h}_{\text{inc}}(x_m, y_n, \omega)^H \mathbf{h}_{\text{inc}}(x_m, y_n, \omega) \Delta x \Delta y \quad (\text{unit:meter}), \tag{10}$$

$$\mathbf{R}_{D,1}(\omega) = \frac{1}{0.2 \cdot 0.6} \sum_{m=1}^{10} \sum_{n=1}^{30} \mathbf{h}_{\text{inc}}(x_m, y_n, \omega)^H \mathbf{h}_{\text{inc}}(x_m, y_n, \omega) \Delta x \Delta y + \frac{1}{0.2 \cdot 0.6} \sum_{m=31}^{40} \sum_{n=1}^{30} \mathbf{h}_{\text{inc}}(x_m, y_n, \omega)^H \times \mathbf{h}_{\text{inc}}(x_m, y_n, \omega) \Delta x \Delta y \quad (\text{unit:meter}). \tag{11}$$

The optimal solution is the eigenvector that corresponds to the maximum eigenvalue of the matrix $\mathbf{R}_{D,1}(\omega)^{-1} \mathbf{R}_{B,1}(\omega)$.[2]

B. The change in sound fields in the case of monopoles

By feeding this solution into loudspeakers, we can obtain a sound field where sound energy is focused in the bright zone when there is no scatterer. The field is expressed by

$$p_{\text{inc},1}(x_m, y_n, \omega) = \mathbf{h}_{\text{inc}}(x_m, y_n, \omega) \mathbf{q}_1(\omega). \tag{12}$$

As a simple example, if loudspeakers are regarded as monopoles, $q^{(l)}(\omega)$ is the monopole amplitude and $h_{\text{inc}}^{(l)}(\boldsymbol{r}, \omega)$ is the transfer function from the monopole amplitude to the incident field by the lth loudspeaker. It is expressed by

$$h_{\text{inc}}^{(l)}(\boldsymbol{r}, \omega) = \frac{e^{jkR_s^{(l)}}}{R_s^{(l)}}, \quad R_s^{(l)} = |\boldsymbol{r} - \boldsymbol{r}_s^{(l)}|. \tag{13}$$

As mentioned above, we can obtain the optimal solution set $\mathbf{q}_1(\omega)$ with these incident fields.

The sound fields that are obtained by applying the solution set $\mathbf{q}_1(\omega)$ at three selected frequencies (800 Hz, 3150 Hz, and 5 kHz) are illustrated in Fig. 4. The magnitude of sound pressure is normalized by the pressure at the origin. It is noteworthy that sound energy is well-focused in the bright zone. The acoustic contrasts are 11.7, 23.3, and 33.2 dB at 800 Hz, 3150 Hz, and 5 kHz, respectively. The acoustic contrast increases as the frequency goes up.

The scattering effect by the rigid sphere changes these incident fields into the total fields. The changed sound field is expressed by

$$p_{\text{tot},1}(x_m, y_n, \omega) = \mathbf{h}_{\text{tot}}(x_m, y_n, \omega) \mathbf{q}_1(\omega). \tag{14}$$

The total fields in the case that loudspeakers are regarded as monopoles are provided in the Appendix. It is noteworthy that $\mathbf{h}_{\text{inc}}(x_m, y_n, \omega)$ in Eq. (12) is replaced with $\mathbf{h}_{\text{tot}}(x_m, y_n, \omega)$ in Eq. (14), while the solution set $\mathbf{q}_1(\omega)$ is kept, meaning that the sound fields by loudspeakers are changed due to the rigid sphere when we use the same solution set. The changed sound fields at the three selected frequencies (800 Hz, 3150 Hz, and 5 kHz) are shown in Fig. 5.

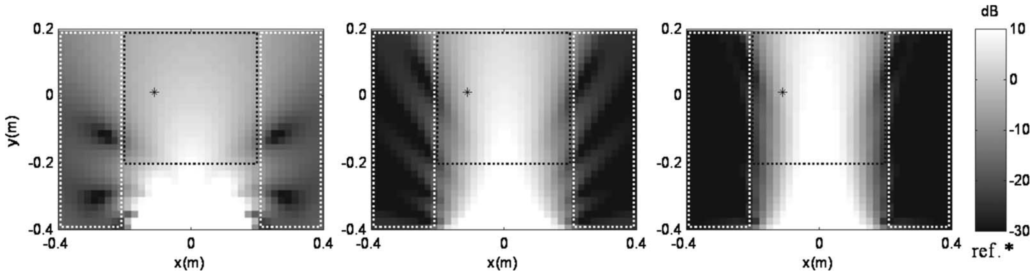


FIG. 4. The focused fields generated by the solution without the scattering effect at 800 Hz (left), 3150 Hz (center), and 5 kHz (right); sound pressure level is normalized by the pressure magnitude at the point $(-0.11,0.01)$, which is marked by *. The square at the center is the bright zone, and the rectangles at the sides are the dark zone. The acoustic contrasts are 11.7, 23.3, and 33.2 dB at 800 Hz, 3150 Hz, and 5 kHz, respectively. The acoustic contrast increases as the frequency goes up.

The sound pressure level is normalized by the pressure magnitude at the point $(-0.11,0.01)$, which is marked by "*." The white circle indicates the rigid sphere. It is noteworthy that the sound pressure level in the dark zone increases. As the frequency increases, the scattering effect becomes intense because of the relative size of the scatterer with respect to wavelength. These results show that scattering effect is not negligible, but has to be considered.

IV. IMPROVEMENT BY CONSIDERING THE SCATTERING EFFECT

A. The solution with the scattering effect

To improve the performance of the system, we need to obtain another solution by considering the scattering effect. Let us denote the solution by

$$\mathbf{q}_2(\omega) = [q_2^{(1)}(\omega) \;\; q_2^{(2)}(\omega) \;\; \cdots \;\; q_2^{(9)}(\omega)]^T. \tag{15}$$

It is noteworthy that we need to obtain the solution by using the transfer functions from the input signal $q^{(l)}(\omega)$ to the total sound field $p_{\text{tot}}^{(l)}(\boldsymbol{r},\omega)$ to consider the scattering effect. In addition, the bright zone needs to be changed because the region where the sound pressure must be enhanced is different from the previous case. In other words, in the previous case, the bright zone was determined as a region that included user's head to allow head movement. However, in this case, the rigid sphere is assumed to model the user's head. Then, the side regions of the sphere are possible positions of ears, and the front and the back regions of the sphere do not include the user's ear. Therefore, the sound pressure levels of the front and the back regions of the sphere do not need to be enhanced, and only the side regions are determined to be the bright zone. As illustrated in Fig. 6, the bright zone is expressed by

$$S_{B,2} = \{(x,y) | -0.2 \le x \le -0.1 \;\text{ or }\; 0.1 \le x \le 0.2, \\ 0.3 \le y \le 0.5\} \quad (\text{unit:meter}), \tag{16}$$

as illustrated in Fig. 6. The dark zone does not need to be changed and is expressed by

$$S_{D,2} = \{(x,y) | -0.4 \le x \le -0.2 \;\text{ or }\; 0.2 \le x \le 0.4, \\ 0 \le y \le 0.6\} \quad (\text{unit:meter}). \tag{17}$$

Then, we construct spatial correlation matrices of the bright and the dark zones, $\mathbf{R}_{B,2}(\omega)$ and $\mathbf{R}_{D,2}(\omega)$ with $\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)$,

$$\mathbf{R}_{B,2}(\omega) = \frac{1}{0.1\cdot 0.2}\sum_{m=11}^{15}\sum_{n=21}^{30}\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)^H\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)\Delta x\Delta y \\ + \frac{1}{0.1\cdot 0.2}\sum_{m=26}^{30}\sum_{n=21}^{30}\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)^H \\ \times\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)\Delta x\Delta y \quad (\text{unit:meter}) \tag{18}$$

and

$$\mathbf{R}_{D,2}(\omega) = \frac{1}{0.2\cdot 0.6}\sum_{m=1}^{10}\sum_{n=1}^{30}\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)^H\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)\Delta x\Delta y \\ + \frac{1}{0.2\cdot 0.6}\sum_{m=31}^{40}\sum_{n=1}^{30}\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)^H \\ \times\mathbf{h}_{\text{tot}}(x_m,y_n,\omega)\Delta x\Delta y \quad (\text{unit:meter}). \tag{19}$$

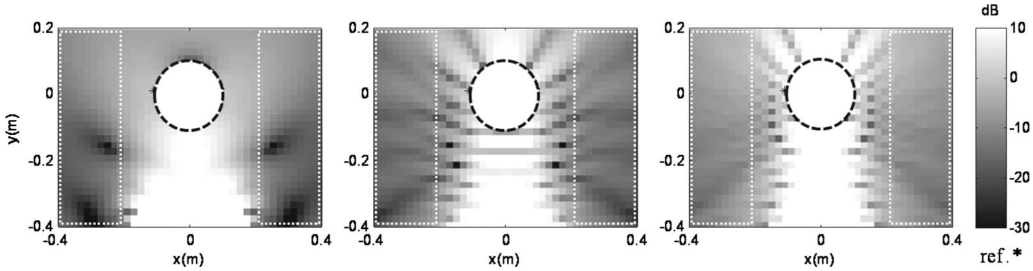


FIG. 5. The total fields by the solution without the scattering effect at 800 Hz (left), 3150 Hz (center), and 5 kHz (right); sound pressure level is normalized by the pressure magnitude at the point $(-0.11,0.01)$, which is marked by *. The white circle indicates the rigid sphere. It is noteworthy that sound pressure level in the dark zone increases due to the scattering effect of the rigid sphere, which is expressed by a white circle at the center.

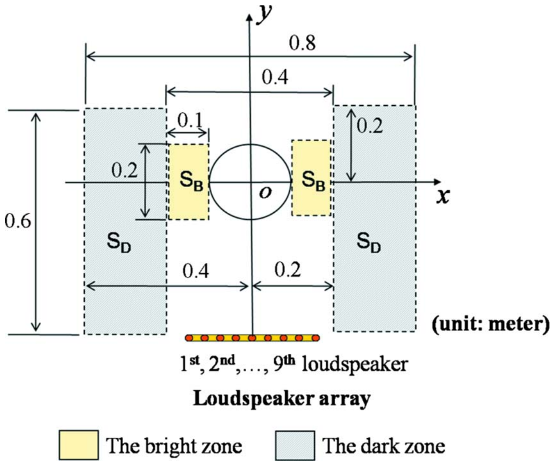


FIG. 6. (Color online) The bright and the dark zones; the bright zone is determined to be rectangles beside the rigid sphere, and the dark zone is two rectangles at the sides to block sound waves from the loudspeaker array.

In the same way, we can obtain the optimal solution $\mathbf{q}_2(\omega)$, which is the eigenvector corresponding to the maximum eigenvalue of the matrix $\mathbf{R}_{D,2}(\omega)^{-1}\mathbf{R}_{B,2}(\omega)$.[2]

By feeding this solution into loudspeakers, we obtain the focused total field. The field is expressed by

$$p_{\mathrm{tot},2}(x_m,y_n,\omega) = \mathbf{h}_{\mathrm{tot}}(x_m,y_n,\omega)\mathbf{q}_2(\omega). \tag{20}$$

Figure 7 illustrates these sound fields at 800 Hz, 3150 Hz, and 5 kHz in the case that loudspeakers are regarded as monopole sources. The solution set $\mathbf{q}_2(\omega)$ is obtained by the eigenvalue problem, as mentioned above, with total fields by monopoles. The magnitude is normalized by pressure at the point $(-0.11,0.01)$, which is marked by $*$. The white circle at the center stands for the rigid sphere, the small rectangles beside the circle are the bright zones, and the two rectangles at the sides are the dark zones. The acoustic contrast is defined as

$$\beta_2(\omega) = \frac{\mathbf{q}_2(\omega)^H\mathbf{R}_{B,2}(\omega)\mathbf{q}_2(\omega)}{\mathbf{q}_2(\omega)^H\mathbf{R}_{D,2}(\omega)\mathbf{q}_2(\omega)}. \tag{21}$$

At 800 Hz, 3150 Hz, and 5 kHz, the contrasts are 8.14, 10.7, and 12.3 dB, respectively.

B. Comparison of two solutions in case of monopoles

For comparison with the results that do not consider the scattering effect, we obtained another solution $\mathbf{q}_1'(\omega)$ under the same configurations illustrated in Fig. 6, by using the transfer functions from the input signal to the incident sound field. That is, the solution $\mathbf{q}_1'(\omega)$ is obtained as the eigenvector correspondent to the maximum eigenvalue of the matrix $\mathbf{R}_{D,1}(\omega)^{-1}\mathbf{R}_{B,1}'(\omega)$, where the $\mathbf{R}_{B,1}'(\omega)$ is defined as

$$\begin{aligned}\mathbf{R}_{B,2}(\omega) &= \frac{1}{0.1\cdot 0.2}\sum_{m=11}^{15}\sum_{n=21}^{30}\mathbf{h}_{\mathrm{inc}}(x_m,y_n,\omega)^H\mathbf{h}_{\mathrm{inc}}(x_m,y_n,\omega)\Delta x\Delta y \\ &+ \frac{1}{0.1\cdot 0.2}\sum_{m=26}^{30}\sum_{n=21}^{30}\mathbf{h}_{\mathrm{inc}}(x_m,y_n,\omega)^H\mathbf{h}_{\mathrm{inc}}(x_m,y_n,\omega)\Delta x\Delta y \\ &(\mathrm{unit{:}meter}).\end{aligned} \tag{22}$$

Figure 8 represents the incident sound fields by the solution $\mathbf{q}_1'(\omega)$ at the three frequencies. The incident fields by the solution $\mathbf{q}_1'(\omega)$ and the previous solution $\mathbf{q}_1(\omega)$ are similar to each other, meaning that $\mathbf{q}_1'(\omega)$ does not have a considerable difference from $\mathbf{q}_1(\omega)$ even if the bright zones are different from each other.

Figure 9 shows the acoustic contrast if the loudspeakers are regarded as monopoles in the three cases of the solutions: $\mathbf{q}_1(\omega)$, $\mathbf{q}_1'(\omega)$, and $\mathbf{q}_2(\omega)$. The contrasts by $\mathbf{q}_1(\omega)$ and $\mathbf{q}_1'(\omega)$ are obtained as follows:

$$\beta_1(\omega) = \frac{\mathbf{q}_1(\omega)^H\mathbf{R}_{B,2}(\omega)\mathbf{q}_1(\omega)}{\mathbf{q}_1(\omega)^H\mathbf{R}_{D,2}(\omega)\mathbf{q}_1(\omega)}, \tag{23}$$

$$\beta_1'(\omega) = \frac{\mathbf{q}_1'(\omega)^H\mathbf{R}_{B,2}(\omega)\mathbf{q}_1'(\omega)}{\mathbf{q}_1'(\omega)^H\mathbf{R}_{D,2}(\omega)\mathbf{q}_1'(\omega)}. \tag{24}$$

The contrast by the solution that considers the scattering effect, $\beta_2(\omega)$, is higher than $\beta_1(\omega)$ and $\beta_1'(\omega)$ at all selected frequencies because the contrast is optimized in this configuration. As the frequency increases, the difference increases as well because the scattering effect becomes intense at higher frequencies. It is noteworthy that the contrast by the solution that considers the scattering effect is enhanced, meaning that the drop in the quality by the scattering effect is improved.

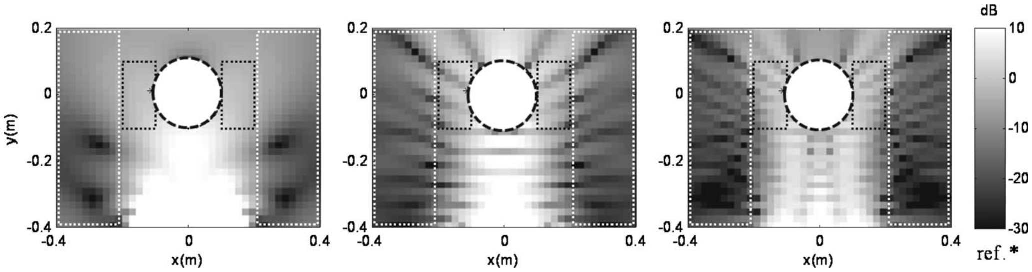


FIG. 7. The total fields by the solution with the scattering effect at 800 Hz (left), 3150 Hz (center), and 5 kHz (right); the magnitude is normalized by pressure at the position $(-0.11,0.01)$, which is marked by $*$. The white circle at the center stands for the rigid sphere, the small rectangles beside the circle are the bright zones, and the two rectangles at the sides are the dark zones. The acoustic contrasts are 8.14, 10.7, and 12.3 dB at 800 Hz, 3150 Hz, and 5 kHz, respectively.

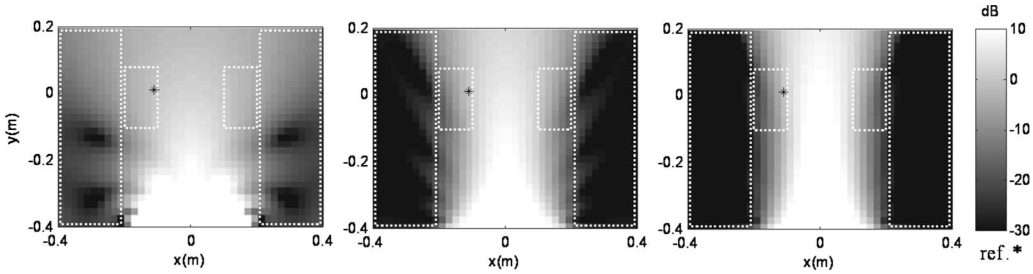


FIG. 8. The focused fields generated by the solution without the scattering effect $\mathbf{q}_1'(\omega)$ at 800 Hz (left), 3150 Hz (center), and 5 kHz (right); sound pressure level is normalized by the pressure magnitude at the point $(-0.11,0.01)$, which is marked by $*$. The bright zone is the rectangles at the center. The results do not have a considerable difference from the fields generated by the $\mathbf{q}_1(\omega)$ (Fig. 4).

V. CONCLUSION

If we apply acoustic contrast control[2] without considering the existence of the user, then we obtain sound fields where sound energy is focused in the bright zone, which is determined to be a square at the center. However, the sound fields are changed by the scattering effect of the user's head, and the dark zone becomes louder. To overcome this problem, we defined the bright zone as the possible positions of the ears in the total field and applied acoustic contrast control. In the case that loudspeakers are regarded as monopoles, we obtained improved results: The acoustic contrast is enhanced, and the sound pressure level in the dark zone is lowered in spite of the scattering effect.

These results shows that scattering effect of the user's head is not trivial in sound focusing technologies that aim to localize sound only on the user (for example, beamforming or time reversal array). It is noteworthy that the contrast control can be used, not only to make a sound beam with the minimum level of side lobes, but also to focus sound only on the user despite the scattering effect.

ACKNOWLEDGMENTS

This research was supported by the National Research Laboratory (NRL) program through the Korea Science and Engineering Foundation funded by the Ministry of Education, Science and Technology (Grant No. R0A-2005-000-10112-0) and the BK21 (Brain Korea 21) project initiated by the Ministry of Education, Science and Technology.

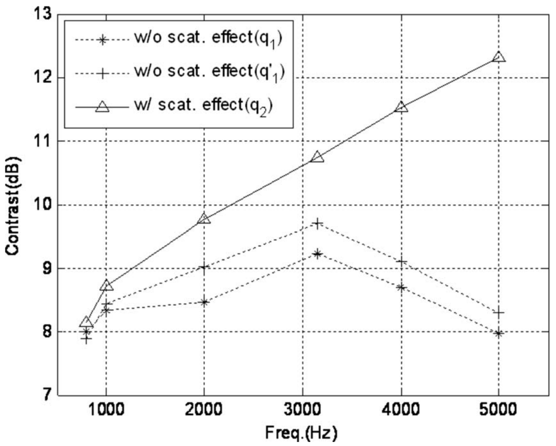


FIG. 9. The increase in the contrast by the solution with the scattering effect; the contrast is enhanced at all selected frequencies (800, 1k, 2k, 3150, 4k, and 5 kHz).

APPENDIX: THE TOTAL FIELD BY A MONOPOLE SOURCE AND A RIGID SPHERE

In order to obtain total fields by a monopole source and a rigid sphere, the boundary condition on the surface of the rigid sphere is applied; on the surface, the particle velocity perpendicular to the surface is zero [Eq. (2)]. In a spherical coordinate, this condition is expressed only with respect to radius r. Then, using the spherical coordinate and expressing sound fields by spherical harmonics[7] allow us to apply this condition easily. Figure 10 illustrates the coordinate that we use and the position of the lth loudspeaker in the spherical coordinate.

If loudspeakers are regarded as monopoles and the monopole amplitude is assumed to be 1, the incident field generated by the lth loudspeaker can be expressed as[8,9]

$$p_{\text{inc}}^{(l)}(\boldsymbol{r},\omega)=\frac{e^{jkR_s^{(l)}}}{R_s^{(l)}}=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(4\pi ik)j_n(kr_<)h_n^{(1)}(kr_>)$$
$$\times Y_n^m(\theta_s^{(l)},\phi_s^{(l)})^* Y_n^m(\theta,\phi),$$
$$R_s^{(l)}=|\boldsymbol{r}-\boldsymbol{r}_s^{(l)}|, \tag{A1}$$

where $i=\sqrt{-1}$, k is the wave number, j_n is the spherical Bessel function, $h_n^{(1)}$ is the first kind of spherical Hankel function, Y_n^m is the spherical harmonics, $r_<$ is the lesser and $r_>$ is the bigger between $(r,r_s^{(l)})$, and $(r_s^{(l)},\phi_s^{(l)},\theta_s^{(l)})$ is the position of the lth loudspeaker.

The scattered wave is an out-going wave that propagates from the inside of the scatterer to the outside. Then, the scattered field $p_{\text{sca}}^{(l)}(r,\theta,\phi,\omega)$ can be expressed as

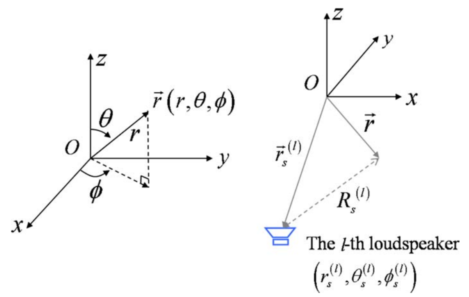


FIG. 10. (Color online) The spherical coordinate (left) and the position of the lth loudspeaker (right) in the spherical coordinate.

$$p_{\text{sca}}^{(l)}(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} B_{mn}^{(l)}(\omega) h_n^{(1)}(kr) Y_n^m(\theta,\phi), \tag{A2}$$

where $B_{mn}^{(l)}(\omega)$ is an unknown coefficient.

Then the total field $p_{\text{tot}}^{(l)}(r,\theta,\phi,\omega)$ is

$$p_{\text{tot}}^{(l)}(r,\theta,\phi,\omega) = p_{\text{inc}}^{(l)}(r,\theta,\phi,\omega) + p_{\text{sca}}^{(l)}(r,\theta,\phi,\omega). \tag{A3}$$

On the surface, the particle velocity is zero and then the boundary condition is expressed as

$$\frac{\partial}{\partial r}[p_{\text{tot}}^{(l)}(r,\theta,\phi,\omega)]_{r=a} = 0. \tag{A4}$$

From this equation, we can get the coefficient $B_{mn}^{(l)}(\omega)$ as

$$B_{mn}^{(l)}(\omega) = -\frac{(ik) j_n'(ka)}{h_n^{(1)\prime}(ka)} h_n^{(1)\prime}(kr_s^{(l)}) Y_n^m(\theta_s^{(l)},\phi_s^{(l)})^*. \tag{A5}$$

[1] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, “A realization of sound focused personal audio system using acoustic contrast control,” J. Acoust. Soc. Am. **125**, 2091–2097 (2009).

[2] J.-W. Choi and Y.-H. Kim, “Generation of an acoustically bright zone with an illuminated region using multiple sources,” J. Acoust. Soc. Am. **111**, 1695–1700 (2002).

[3] F. L. Wightmann and D. J. Kistler, “Monaural sound localization revisited,” J. Acoust. Soc. Am. **101**, 1050–1063 (1997).

[4] V. R. Algazi, C. Avendano, and R. O. Duda, “Elevation localization and head-related transfer function analysis at low frequencies,” J. Acoust. Soc. Am. **109**, 1110–1122 (2001).

[5] S. J. Elliott and M. Jones, “An active headrest for personal audio,” J. Acoust. Soc. Am. **119**, 2702–2709 (2006).

[6] M. Jones and S. J. Elliott, “Personal audio with multiple dark zones,” J. Acoust. Soc. Am. **124**, 3497–3506 (2008).

[7] W. T. Kelvin and P. G. Tait, *Treatise on Natural Philosophy*, Elibron Classics Series Vol. **I** (Adamant Media Corporation, Boston, 2005), pp. 171–218; *Treatise on Natural Philosophy*, Elibron Classics Series Vol. **I** (Cambridge University Press, Cambridge, 1896).

[8] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999), p. 428.

[9] E. G. Williams, *Fourier Acoustics* (Academic, Cambridge, UK, 1999), pp. 224–230.

Enhanced channel estimation and symbol detection for high speed multi-input multi-output underwater acoustic communications

Jun Ling, Tarik Yardibi, Xiang Su, Hao He, and Jian Li[a)]
Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611

(Received 2 August 2008; revised 11 February 2009; accepted 11 February 2009)

The need for achieving higher data rates in underwater acoustic communications leverages the use of multi-input multi-output (MIMO) schemes. In this paper two key issues regarding the design of a MIMO communications system, namely, channel estimation and symbol detection, are addressed. To enhance channel estimation performance, a cyclic approach for designing training sequences and a channel estimation algorithm called the iterative adaptive approach (IAA) are presented. Sparse channel estimates can be obtained by combining IAA with the Bayesian information criterion (BIC). Moreover, the RELAX algorithm can be used to improve the IAA with BIC estimates further. Regarding symbol detection, a minimum mean-squared error based detection scheme, called RELAX-BLAST, which is a combination of vertical Bell Labs layered space-time (V-BLAST) algorithm and the cyclic principle of the RELAX algorithm, is presented and it is shown that RELAX-BLAST outperforms V-BLAST. Both simulated and experimental results are provided to validate the proposed MIMO scheme. RACE'08 experimental results employing a 4×24 MIMO system show that the proposed scheme enjoys an average uncoded bit error rate of 0.38% at a payload data rate of 31.25 kbps and an average coded bit error rate of 0% at a payload data rate of 15.63 kbps. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097467]

PACS number(s): 43.60.Fg, 43.60.Mn, 43.60.Ac [EJS] Pages: 3067–3078

I. INTRODUCTION

Achieving reliable communications with high data rates over underwater acoustic (UWA) channels has long been recognized as a challenging problem owing to the scarce bandwidth available and the double spreading phenomenon, i.e., spreading in both the time (delay spread) and frequency domains (Doppler spread).[1,2] Delay and Doppler spreading are inherent to many practical communication channels[3] but are profoundly amplified in underwater environments.[4–7]

The large spread in delay is the result of multipath propagation and the relatively low velocity of acoustic waves compared to electromagnetic waves.[2] When the delay spread is large, a transmitted symbol may interfere with many of its adjacent symbols at the receiver. This leads to severe intersymbol interference (ISI), which complicates the receiver structure and makes it difficult to extract the desired symbol from the measurements.

Doppler spreading stems from the random relative motion of the transmitters and receivers, and the nonstationarity of the underwater medium. This results in undesired phase shifts, making coherent communications using, for instance, phase-shift keying (PSK),[3] difficult for practical underwater communications. Thus, incoherent strategies, such as frequency-shift keying (FSK),[3] instead drew a lot of interest in the early UWA research.[8,9] Although being immune to double spreading, FSK is much less bandwidth efficient than PSK. After the employment of the phase locked loop (PLL) methodology[10,11] in underwater applications, coherent UWA communications gained popularity.[4,12,13]

While PLL is in general successful in mitigating the effects of Doppler spreading, the delay spread, or equivalently ISI, can be accounted for by either the decision feedback equalizer (DFE)[14,15] or the passive-phase conjugate[16] (PPC) methods. A detailed treatment alongside with performance comparisons of DFE and PPC is presented by Yang.[17,18] In practical UWA systems, the coupling of DFE and PLL has found great success[4,14] and almost became a standard.[6] DFE is a nonlinear equalizer, whose coefficients are updated by an adaptive approach such as the well-known recursive least squares (RLSs) or the least mean square algorithms.[4,19,20] The principle behind PPC is matched filtering, which states that when the channel impulse response (CIR) is convolved with its time-reversed and conjugated version at each receiver and added up, the summation approaches a delta function.[21] This compensates for the channel effects in the received signal. Obviously, the performance of such an approach relies heavily on the accuracy of the CIR estimate, especially when only few receivers exist. Taking one step further beyond the classic coupling of DFE with PLL, Yang[22] presented a hybrid structure combining the advantage of PPC with a single channel DFE and introduced a Doppler shift removal module before feeding the signals to the DFE.[23]

All the aforementioned methods, however, are confined to single-input multiple-output (SIMO) [or single-input single-output (SISO)] UWA communication systems. When multiple transmitters are used, interference between the multiple transmitted signals (besides ISI) degrades the perfor-

[a)]Author to whom correspondence should be addressed. Electronic mail: li@dsp.ufl.edu mance of such methods significantly. This paper focuses on phase coherent communications over multi-input multi-output (MIMO) UWA channels with delay spreading only. We do not consider the effects of Doppler spreading herein. Yet, the methods presented in this paper can easily be extended to deal with Doppler spreading if desired.[24] The main motivation for establishing a MIMO system for underwater communications is the desire for higher data rates. As is well known, MIMO systems enjoy much higher data rates compared with their SIMO counterparts due to exploiting spatial diversity on both the receiver and the transmitter sides. To the best of our knowledge, there is not much work dealing with MIMO UWA communications in the literature. Early attempts[25,26] mainly focus on the design of the equalizer in the receiver end while some recent approaches[27,28] tackle the problem from a coding perspective. The design of a precoder that maps the source data to multiple transmitters in an optimal manner assuming that accurate channel estimates are available was presented by Kilfoyle *et al.*[29,30] In the present paper, we offer a broader view of the MIMO UWA communications problem and address two important issues: (i) estimation of the underwater CIR with delay spread and (ii) detection schemes for recovering the transmitted symbols using the estimated CIR.

In general, the very first task of the receiver is to conduct a training-directed channel estimation.[4,31] To achieve good performance, both well-structured training sequences and a signal processing methodology that can estimate the CIR accurately using the designed training sequences are required. In addition, to address the time-varying nature of the UWA channel, the decision-directed channel estimation is performed regularly using the detected symbols instead of the training symbols.[4,31] Therefore, the channel estimation algorithm should be able to work well both in training- and decision-directed modes.

When designing the training sequences, the delay spread of the UWA channel must be taken into account. For MIMO flat fading channels, i.e., channels without delay spread, Hadamard sequences[32,33] can be used effectively whereas for practical multipath channels, sequences with good auto- and cross-correlation properties instead are required.[34,35] Early research has focused on binary training sequences[34,36] due to practical concerns and simplicity. Later on, the use of polyphase training sequences was proposed, where the possible phase values were confined to a predefined finite set.[35] It is obviously advantageous to allow the phase values to be continuous. Yet, the problem becomes more demanding computationally as the degrees-of-freedom is allowed to increase. The cyclic approach (CA) presented by Li *et al.*[37,38] for probing sequence design enjoys superior performance over the aforementioned methods by allowing continuous phase values while still being computationally tractable. The training sequences designed using the CA methodology possess good auto- and cross-correlation properties as desired for MIMO channel estimation in communications.[37,38]

As mentioned previously, the second phase of channel estimation involves the design of the algorithm that will estimate the CIR using the training sequences (or the previously detected symbols) and highly contaminated measurements, be it either by the ISI and the interference from multiple transmitters or by the unpredictable nature of the underwater medium. Three important sparsity based techniques have been used for underwater channel estimation, namely, the matching pursuit (MP) algorithm, the orthogonal MP (OMP) algorithm,[39] and the least squares MP (LSMP) algorithm.[31,39–44] The main motivation for using MP type of algorithms is that many channels including underwater communication channels[14,45,46] and wireless channels are appropriately modeled as sparse channels consisting of a few dominant delay and Doppler taps.[47] One problem with these methods is that it is difficult to determine the stopping criterion and user intervention might be needed. Moreover, the performance of these methods might degrade significantly depending on the structure of the matrix relating the unknowns to the measurements. For instance, as will be shown in our numerical examples later on, these methods show better performance with CA designed training sequences rather than with arbitrary training sequences, especially when the training length is small. To address these problems, we present a user parameter-free nonparametric iterative adaptive approach[24] (IAA) for estimating the CIR accurately even when the training sequences are arbitrary and short in length. The dominant channel tap estimates of IAA can be used in a Bayesian information criterion (BIC)[48,49] to decide which taps to retain and which ones to discard. This combined method, called IAA with BIC, results in sparse channel estimates. Further improvements in performance can be achieved by initializing the last step of the cyclic and relaxation-based RELAX[50,51] algorithm via the IAA with BIC sparse estimates.

Following the estimation of the CIR is the design of the detection scheme for extracting the payload symbols from the measurements. We use a minimum mean-squared error (MMSE) based filter for signal detection. Two important methods for applying the MMSE filter coefficients to the measurements are the linear combinatorial nulling[52] (LCN) and vertical Bell Labs layered space-time (V-BLAST) algorithms.[53] It is interesting to note that these two approaches resemble the classical periodogram[54,55] and the CLEAN[56] methods used in spectral estimation applications. Being inspired from the improvements of RELAX over the periodogram and CLEAN,[54] we propose the RELAX-BLAST detection algorithm, which is a combination of V-BLAST and the cyclic principle of RELAX as the name suggests, and show that it outperforms V-BLAST.

The rest of this paper is organized as follows. Section II outlines the system configuration and describes the data package structure. Section III formulates the problem of CIR estimation, describes the CA method for training sequence design, and presents the IAA algorithm together with the BIC and RELAX extensions. Next, the symbol detection problem is analyzed in Sec. IV and the MMSE based RELAX-BLAST detection scheme is proposed. Both simulated and experimental results are presented in Sec. V. The sea data were gathered in the rescheduled acoustic communications experiment (RACE'08), which was conducted by the Woods Hole Oceanographic Institution (WHOI) in Narragansett Bay. This paper is concluded in Sec. VI.

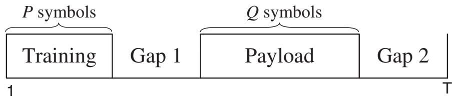


FIG. 1. The structure of a single data package.

The main contribution of the present paper is the thorough investigation of a MIMO UWA communications system by providing a detailed treatment of every step involved from data transmission to symbol detection at the receiver. This is done by presenting approaches for designing well-structured training sequences, a novel channel estimation method and a novel detection scheme. Simulation and experimental results validate the utility of the proposed overall scheme for MIMO underwater communications.

Notation: We denote vectors and matrices by boldface lowercase and boldface uppercase letters, respectively. $\|\cdot\|_2$ denotes the Euclidean norm, $\|\cdot\|_F$ denotes the Frobenius matrix norm, $(\cdot)^T$ denotes the transpose, $(\cdot)^H$ denotes the conjugate transpose, $E(\cdot)$ denotes the expected value, $\mathbf{I}$ denotes the identity matrix of appropriate size, and $\hat{\mathbf{x}}$ denotes the estimate of $\mathbf{x}$.

II. SYSTEM OUTLINE

Consider an $N\times M$ MIMO communications system equipped with N transmit transducers and M receive transducers. The individual data streams of each transmitter are symbol aligned and are sent simultaneously. The data streams of each transmitter consist of successive data packages of the form shown in Fig. 1. The data packages start with a training sequence of length P which is followed by a silent gap, the payload sequence, and another silent gap. During the gap intervals, no signal is transmitted in order to prevent the inner-package ISI (gap 1) between the training and payload symbols and the inter-package ISI (gap 2) between two consecutive packages. The payload sequence, which has length Q ($Q>P$ in general), is the estimation target and each payload symbol is drawn from a quadrature PSK (QPSK) constellation modulated with Gray code.[3] The four constellation points of the QPSK symbols, i.e., $\{e^{j(2n-1)\pi/4}\}_{n=1}^4$, lie on the unit circle. Such a constellation is desirable in practice due to its unit modulus. The same practical constraints require the training symbols to have unit modulus as well but no restriction is imposed on their phase values.

In what follows, our consideration is always confined to one data package of the form given in Fig. 1. Let $x_n(t)$ denote the tth symbol in the package sent by the nth transmitter and let $y_m(t)$ denote the tth symbol in the package received by the mth receiver, where $n=1,\ldots,N$, $m=1,\ldots,M$, $t=1,\ldots,T$, and T is the total symbol length of a single transmitted package. We do not go into the details of the sampling and synchronization procedures and assume that such operations have already been employed and the sampled complex baseband signals are available at the receiver.

Figure 2 shows the $N\times M$ MIMO system structure that we will use throughout the paper. The source bits are encoded, QPSK modulated, interleaved, and demultiplexed for transmission from multiple transducers. A random interleaver is used in order to avoid burst errors, which occur when the channel behaves badly at certain intervals of time.[3] After the signals have been received by the receive array, the processing consists of two steps: estimating the CIR (in training- or decision-directed mode) and detecting the symbols by using the estimated CIR. Once the symbols have been detected, they are multiplexed, deinterleaved, and then fed into a Viterbi decoder to recover the source bits. We now discuss the channel estimation problem.

III. CHANNEL ESTIMATION

In this section, we formulate the problem of channel estimation and describe the CA for training sequence design. We then propose IAA for channel estimation.

A. Problem formulation

1. Training-directed mode

The measurement vector at the mth receiver can be written as

$$\mathbf{y}_m=\sum_{n=1}^{N}\tilde{\mathbf{X}}_n\mathbf{h}_{n,m}+\mathbf{e}_m \tag{1}$$

for $m=1,\ldots,M$, where

$$\mathbf{y}_m=[y_m(1),\ldots,y_m(P+R-1)]^T, \tag{2}$$

$$\mathbf{h}_{n,m}=[h_{n,m}(1),\ldots,h_{n,m}(R)]^T, \tag{3}$$

and $R-1$ is the maximum number of delay taps under consideration. [Note that this corresponds to a $(R-1)\Delta t$ s delay spread, where Δt is the symbol interval.] $\mathbf{h}_{n,m}$ denotes the CIR between the nth transmitter and the mth receiver,

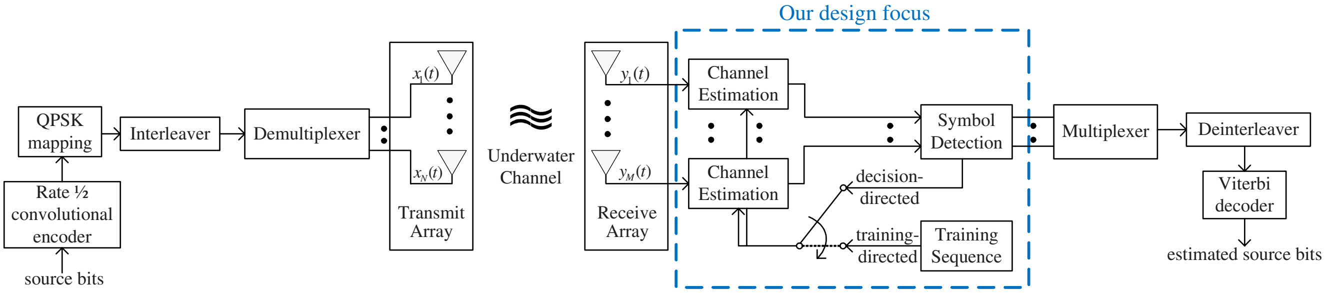


FIG. 2. (Color online) An $N\times M$ MIMO UWA communications system. The blocks inside the dashed rectangle are the focus of our attention in this paper.

$$\tilde{\mathbf{X}}_n = \begin{bmatrix} x_n(1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \\ x_n(P) & & x_n(1) \\ & \ddots & \vdots \\ \mathbf{0} & \cdots & x_n(P) \end{bmatrix}, \tag{4}$$

where $\tilde{\mathbf{X}}_n \in \mathbb{C}^{(P+R-1)\times R}$ contains the nth training sequence and hence is known, and $\mathbf{e}_m$ is the additive noise (thermal or hardware related noise) at the mth receiver. Equation (1) can be rewritten as

$$\mathbf{y}_m = \mathbf{X}\mathbf{h}_m + \mathbf{e}_m, \tag{5}$$

where $\mathbf{X}=[\tilde{\mathbf{X}}_1 \cdots \tilde{\mathbf{X}}_N]$ and $\mathbf{h}_m=[\mathbf{h}_{1,m}^T \cdots \mathbf{h}_{N,m}^T]^T$. The training-directed channel estimation problem then reduces to estimating $\mathbf{h}_m$ from the measurements $\mathbf{y}_m$ and known $\mathbf{X}$. It is assumed that the channel is stationary over the length of $\mathbf{y}_m$. In order to estimate all the channels for the $N\times M$ MIMO system, Eq. (5) has to be solved for $m=1,\dots,M$, i.e., M times. Note that $\mathbf{X}$ does not depend on m.

2. Decision-directed mode

The problem in the decision-directed mode is very similar to that of the training-directed mode except that now the training symbols are replaced with the previously estimated payload symbols. Consequently, Eq. (5) can be expressed as

$$\mathbf{y}_m = \hat{\mathbf{X}}\mathbf{h}_m + \mathbf{e}_m, \tag{6}$$

where $\hat{\mathbf{X}}=[\hat{\mathbf{X}}_1 \cdots \hat{\mathbf{X}}_N]$,

$$\hat{\mathbf{X}}_n = \begin{bmatrix} \hat{x}_n(t_i) & \hat{x}_n(t_i-1) & \dots & \hat{x}_n(t_i-R+1) \\ \hat{x}_n(t_i+1) & \hat{x}_n(t_i) & \dots & \hat{x}_n(t_i-R+2) \\ \vdots & \vdots & & \vdots \\ \hat{x}_n(t_f) & \hat{x}_n(t_f-1) & \dots & \hat{x}_n(t_f-R+1) \end{bmatrix}, \tag{7}$$

$$\mathbf{y}_m = [y_m(t_i), \dots, y_m(t_f)]^T, \tag{8}$$

and where $\hat{x}_n(t_i-R+1)$ and $\hat{x}_n(t_f)$ represent the first and the last previously estimated symbols, respectively, used for updating the channel. The decision-directed channel estimation problem reduces to estimating $\mathbf{h}_m$ from the measurements $\mathbf{y}_m$ and the previously decoded symbols in $\hat{\mathbf{X}}$.

On the one hand, it would be beneficial to keep $L \triangleq t_f - t_i + R$ large for estimating the channel more accurately but on the other hand, for a rapidly varying channel, L must be kept small so that the stationarity assumption of the channel over the length of $\mathbf{y}_m$ holds and so that the channel can be updated more frequently. Therefore, L is a trade-off parameter which should be set according to the experimental conditions.

Note that the channel estimates obtained using the training sequences may become outdated before the first set of payload symbols are estimated due to the gap between the training and payload sequences. However, the length of the gap interval is relatively small and this effect can often be neglected. If the sea is expected to be very nonstationary, a smaller gap interval should be used even though this will increase the ISI between the training and the payload sequences.

B. Training sequence design

We use the algorithm presented by Li *et al.*[37,38] for designing training sequences such that $\mathbf{X}$ in Eq. (5) facilitates the estimation of the CIR. It is desirable to have training symbols with constant modulus, i.e., the training symbols should have the following generic form:

$$x_n(t) = e^{j\phi_n(t)}, \quad t=1,\dots,P, \quad n=1,\dots,N, \tag{9}$$

where $\phi_n(t) \in [0,2\pi)$ represents the phase of the tth training symbol sent by the nth transmitter. Ideally, if $\mathbf{X}^H\mathbf{X}=P\mathbf{I}$ (called the pairwise orthogonality principle), then the channel estimates can be recovered perfectly by matched filtering in the noiseless case. However, pairwise orthogonality is hardly achievable, if not impossible, in practice.[38] Instead, $\epsilon = \|\mathbf{X}^H\mathbf{X} - P\mathbf{I}\|_F^2$ can be made small.

Let $\mathbf{U}$ be an arbitrary semi-unitary matrix (i.e., $\mathbf{U}\mathbf{U}^H=\mathbf{I}$). Then,

$$\epsilon = \|\mathbf{X}^H\mathbf{X} - (\sqrt{P}\mathbf{U})(\sqrt{P}\mathbf{U}^H)\|_F^2. \tag{10}$$

Minimizing ϵ can then be formulated in the following related (but not equivalent) way:[38]

$$\{\phi_n(t)\} = \underset{\{\phi_n(t)\},\mathbf{U}^H}{\operatorname{argmin}} \|\mathbf{X} - \sqrt{P}\mathbf{U}^H\|_F^2, \quad \text{subject to } \mathbf{U}\mathbf{U}^H = \mathbf{I}. \tag{11}$$

This optimization problem can be solved efficiently by using the CA method[38,57] which guarantees that the cost function does not increase as the iterations proceed. In the CA method, $\mathbf{U}$ is assumed given when estimating $\{\phi_n(t)\}$ and vice versa. This way, the optimization problem is solved iteratively by dividing it into simpler sub-problems.

When $\mathbf{U}^H$ is fixed, the solution to Eq. (11) has the generic form

$$\phi = \arg\left(\sum_{r=1}^{R} z_r\right), \tag{12}$$

where $\{z_r\}_{r=1}^R$ are given numbers. For example, when the update target is $\phi_1(1)$, z_r represents the (r,r)th diagonal entry of $\sqrt{P}\mathbf{U}^H$.

Given the phases $\phi_n(t)$, the solution to Eq. (11) is given by $\mathbf{U}^H = \bar{\mathbf{U}}\tilde{\mathbf{U}}^H$,[38,58] where

$$\sqrt{P}\mathbf{X} = \bar{\mathbf{U}}\mathbf{\Gamma}\tilde{\mathbf{U}}^H \tag{13}$$

is the singular value decomposition of $\sqrt{P}\mathbf{X}$ ($\bar{\mathbf{U}}$ and $\tilde{\mathbf{U}}^H$ are unitary matrices and $\mathbf{\Gamma}$ is a diagonal matrix with the singular values of $\sqrt{P}\mathbf{X}$ on its diagonal).

The CA algorithm is terminated when the difference of the cost function [defined in Eq. (11)] between two successive iterations drops below a certain threshold. For the CA algorithm to show good performance, it is recommended that $P \gg R$ and $NR < P+R-1$.[37,38] In practice, N is determined from the system configuration while R is selected depending on the experimental conditions and is expected to be the smallest value that can capture the prominent channel features. It seems as if a large P value is preferable for satisfying the two inequalities. However, there are two problems associated with increasing P. First, the accuracy of the initial channel estimation depends on the assumption that the channel is stationary. For a fixed symbol rate, larger P means longer transmission time which means the stationarity assumption is more likely to be violated. Second, larger P means larger overhead and hence lower net data rate. Fortunately, though, the two inequalities can in general be satisfied in practice by selecting the parameters appropriately. Note that the CA method has been used to design the training sequences in Sec. V of this paper.

C. The channel estimation algorithm: IAA

The channel estimation problem at each receiver has the generic form given by [see Eqs. (5) and (6)]

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{e}, \tag{14}$$

where we have omitted the index m and replaced $\mathbf{X}$ (for the training-directed mode) or $\hat{\mathbf{X}}$ (for the decision-directed mode) by $\mathbf{S}$ for notational simplicity. Note that the number of elements in $\mathbf{y}$, namely, d_y, is also different for the two modes. The problem is then to estimate $\mathbf{h}$, which has NR unknowns, given $\mathbf{y}$ and $\mathbf{S}$. In the following, we present the IAA algorithm[24] to solve this problem. IAA makes no assumptions on the statistical properties of the additive noise $\mathbf{e}$. Note that since $\mathbf{h}$ contains the CIR of all N transmitters, IAA will estimate them jointly.

1. IAA

Many existing weighted least squares (WLSs) based channel estimation methodologies require the tuning of one or more user parameters and their assumptions on the CIR are in general not valid in the underwater scenario.[59,60] To account for these problems, we present a user parameter-free iterative WLS based channel estimation technique, called IAA.[24] IAA is an adaptive and nonparametric algorithm, and it does not make any explicit assumptions on the CIR. Let $\mathbf{P}$ be an $NR \times NR$ diagonal matrix whose diagonal contains the squared absolute value of each channel tap, i.e.,

$$P_r = |h_r|^2, \quad r = 1, \ldots, NR, \tag{15}$$

where P_r is the rth diagonal element of $\mathbf{P}$ and h_r is the rth element of $\mathbf{h}$. If the rth column of $\mathbf{S}$ is written as $\mathbf{s}_r$, then the covariance matrix of the noise and interference with respect to the tap of current interest h_r can be expressed as

$$\mathbf{Q}(r) = \mathbf{R} - P_r \mathbf{s}_r \mathbf{s}_r^H, \tag{16}$$

where $\mathbf{R} \triangleq \mathbf{S}\mathbf{P}\mathbf{S}^H$. Then, the WLS cost function is given by[54,61–63]

$$(\mathbf{y} - h_r\mathbf{s}_r)^H \mathbf{Q}^{-1}(r)(\mathbf{y} - h_r\mathbf{s}_r). \tag{17}$$

TABLE I. IAA.

```
$P_r = \frac{|\mathbf{s}_r^H \mathbf{y}|^2}{(\mathbf{s}_r^H \mathbf{s}_r)^2}$, $r = 1, 2, \ldots, NR$
repeat
  $\mathbf{R} = \mathbf{S}\mathbf{P}\mathbf{S}^H$
  $\hat{h}_r = \frac{\mathbf{s}_r^H \mathbf{R}^{-1}\mathbf{y}}{\mathbf{s}_r^H \mathbf{R}^{-1}\mathbf{s}_r}$, $r = 1, 2, \ldots, NR$
  $P_r = |\hat{h}_r|^2$, $r = 1, 2, \ldots, NR$
until (convergence)
```

Minimizing Eq. (17) with respect to h_r yields

$$\hat{h}_r = \frac{\mathbf{s}_r^H \mathbf{Q}^{-1}(r)\mathbf{y}}{\mathbf{s}_r^H \mathbf{Q}^{-1}(r)\mathbf{s}_r}. \tag{18}$$

Using Eq. (16) and the matrix inversion lemma, Eq. (18) can be written as

$$\hat{h}_r = \frac{\mathbf{s}_r^H \mathbf{R}^{-1}\mathbf{y}}{\mathbf{s}_r^H \mathbf{R}^{-1}\mathbf{s}_r}. \tag{19}$$

This avoids the computation of $\mathbf{Q}^{-1}(r)$ for NR times and only one matrix inversion is needed per iteration. IAA for channel estimation is summarized in Table I. Since IAA requires $\mathbf{R}$, which itself depends on the unknown channel taps, it has to be implemented as an iterative approach. The initialization is done by a standard matched filter. Our empirical experience is that IAA does not provide significant improvements in performance after about 15 iterations. In IAA, $\mathbf{P}$ and hence $\mathbf{R}$ are obtained from the channel estimates of the previous iteration and not from the measurements $\mathbf{y}$ as done in conventional adaptive filtering algorithms.

If the computation of $\mathbf{R}$ becomes problematic due to numerical ill-conditioning during the iterations, a regularization approach can be used. IAA can be regularized by considering an additional noise term separately from the interference terms in the expression for $\mathbf{R}$:

$$\mathbf{R} = \mathbf{S}\mathbf{P}\mathbf{S}^H + \mathbf{\Sigma}, \tag{20}$$

where $\mathbf{\Sigma}$ is a diagonal matrix with unknown noise powers $\{\sigma_m^2\}_{m=1}^{d_y}$ on its diagonal. IAA is then implemented in the same way as before except that now there are $NR + d_y$ rather than NR unknowns. Consequently, $\{\sigma_m^2\}$ can be estimated by

$$\hat{\sigma}_m^2 = \frac{|\mathbf{i}_m^H \mathbf{R}^{-1}\mathbf{y}|^2}{(\mathbf{i}_m^H \mathbf{R}^{-1}\mathbf{i}_m)^2}, \quad m = 1, \ldots, d_y, \tag{21}$$

at each iteration, where $\mathbf{i}_m$ is the mth column of the $d_y \times d_y$ identity matrix. Since the diagonal loading levels are calculated automatically, the approach conserves the practicality of IAA. Setting $\{\hat{\sigma}_m\}_{m=1}^{d_y}$ to zero gives the original IAA algorithm. $\mathbf{\Sigma}$ can be initialized as all zeros.

2. IAA with BIC

In order to achieve sparsity with IAA, i.e., to retain only a few dominant channel taps, the BIC[48,49] can be used in conjunction with IAA. BIC is a model order selection tool that is widely used in the statistics and signal processing communities. The advantage of using BIC over a simple thresholding operation is that BIC does not require the manual specification of a threshold value. (Note that the se-

TABLE II. IAA with BIC.

```
$\mathcal{P}=\{1,\ldots,NR\}$
$\mathcal{I}=\{\emptyset\}$; $\eta=1$; quit=0; $\mathrm{BIC}^{\mathrm{old}}=\infty$
repeat
    $i'=\mathrm{argmin}_{i\in\mathcal{P}-\mathcal{I}}\,\mathrm{BIC}_i(\eta)$
    if $\mathrm{BIC}_{i'}(\eta)<\mathrm{BIC}^{\mathrm{old}}$
        $\mathcal{I}=\{\mathcal{I},i'\}$
        $\mathrm{BIC}^{\mathrm{old}}=\mathrm{BIC}_{i'}(\eta)$
        $\eta=\eta+1$
    else quit=1
until (quit=1)
```

lection of the threshold value has a significant effect on the overall performance and it is usually impractical to tune this value for best performance since the true CIR is unknown.) Let $\mathcal{P}$ denote a set containing the indices of all the channel taps. Also, let $\mathcal{I}$ denote the set of the indices of the taps selected by the BIC algorithm so far. The IAA with BIC algorithm works as follows: first, the tap from the set $\mathcal{P}$ giving the minimum BIC is selected and included in the set $\mathcal{I}$ (initially $\mathcal{I}=\emptyset$). Then the second tap, from the set $\mathcal{P}-\mathcal{I}$, which together with the first tap gives the minimum BIC is selected and so on until the BIC value does not decrease anymore. The IAA with BIC algorithm is summarized in Table II. $\mathrm{BIC}_i(\eta)$ is calculated as follows:[48]

$$\mathrm{BIC}_i(\eta)=2d_y\ln\left(\left\|\mathbf{y}-\sum_{j\in\{\mathcal{I}\cup i\}}\mathbf{s}_j\hat{h}_j\right\|_2^2\right)+1.5\,\eta\ln(2d_y), \tag{22}$$

where $\eta=|\mathcal{I}|+1$, with $|\mathcal{I}|$ denoting the size of $\mathcal{I}$, i is the index of the current tap under consideration, and $\hat{h}_j$ is the IAA estimate of the jth element of $\mathbf{h}$, $j\in\{\mathcal{I}\cup i\}$. After BIC is implemented, the indices of the surviving CIR taps can be found in $\mathcal{I}$. All other channel taps are then set to zero.

3. IAA with RELAX

The parametric and cyclic RELAX algorithm,[50,51] which was originally proposed for spectral estimation, can be used to improve the IAA with BIC results even further. Because RELAX is parametric, it requires the number of sources to be known. The IAA with BIC result can be used to estimate the number of sources and also to provide initial estimates for the last step of RELAX, as shown in Table III. Note that $\mathcal{I}(k)$ denotes the kth element in the set $\mathcal{I}$. The idea presented in Table III is to remove the contribution from all the components of $\hat{\mathbf{h}}$ other than the one of current interest $\hat{h}_{\mathcal{I}(k)}$ and then to update $\hat{h}_{\mathcal{I}(k)}$ in the minimum least squares sense. This procedure is repeated until the difference of the cost function $\|\mathbf{y}-\mathbf{S}\hat{\mathbf{h}}\|_2^2$ between two successive iterations becomes less than a certain threshold. (We used a threshold of 5×10^{-3} in our simulations herein.) For the best performance, it is recommended that before each RELAX iteration, $\{\hat{h}_k\}$ be sorted by their magnitude in descending order and the columns of $\mathbf{S}$ be permuted accordingly. This way, the tap with the largest magnitude will be updated first, the tap with the second largest magnitude will be updated next, and so on.

TABLE III. IAA with RELAX.

```
$\mathcal{I}$: Indices of the taps selected by IAA with BIC
$K=|\mathcal{I}|$, i.e., the number of selected taps
repeat
    for $k=1,2,\ldots,K$
        $\mathbf{y}_k=\mathbf{y}-\sum_{i=1,i\neq k}^{K}\mathbf{s}_{\mathcal{I}(i)}\hat{h}_{\mathcal{I}(i)}$
        $\hat{h}_{\mathcal{I}(k)}=\mathbf{s}_{\mathcal{I}(k)}^H\mathbf{y}_k/\|\mathbf{s}_{\mathcal{I}(k)}\|_2^2$
    end for
until (convergence)
```

4. Complexity analysis

The initialization step of IAA has complexity $\mathcal{O}(2d_y(NR)+3(NR))$ and each IAA iteration has complexity $\mathcal{O}(d_y^3+(2d_y^2+3d_y+2)(NR))$. These complexities are calculated by counting the multiplication and division operations in Table I. When $d_y>(NR)$, $\mathbf{R}^{-1}$ can be calculated only once at the initialization step of IAA and then it can be updated when every $\{P_r\}$ is estimated using the rank-1 matrix inverse update formula.[64] This way, the complexity of computing $\mathbf{R}^{-1}$ reduces to $\mathcal{O}((d_y^2+3)(NR))$ rather than $\mathcal{O}(d_y^3)$ at each IAA iteration. The resulting complexity of IAA is then given by $\mathcal{O}(d_y^3+(d_y^2+3d_y+3)(NR))$ for initialization and $\mathcal{O}((2d_y^2+2d_y+5)(NR))$ per IAA iteration. The complexity of IAA is smaller than those of MP and LSMP when $d_y\ll(NR)$ and larger when $d_y>(NR)$.[31] However, the computation time does not depend only on the number of computations but rather is a function of the memory access time, the implementation software and hardware, and the number of computations combined together.

Note that the regularization, BIC, and RELAX extensions will be applied in all of our numerical examples and henceforth this combined approach will simply be referred to as IAA.

IV. SYMBOL DETECTION

The symbol detection problem is to estimate the transmitted payload symbols using the CIR estimates. In this section we first describe how to obtain the MMSE filter coefficients for each symbol of interest and then describe three methods on how to apply these filter coefficients to the measurements.

A. Problem formulation

Once the CIR estimates are available, the transmitted symbols can be detected by expressing Eq. (6) as[15]

$$\mathbf{y}_m=\tilde{\mathbf{H}}_m\tilde{\mathbf{x}}+\mathbf{e}_m, \tag{23}$$

where

$$\mathbf{y}_m=[y_m(t_0),\ldots,y_m(t_0+R-1)]^T, \tag{24}$$

t_0 represents the index corresponding to the payload symbol that is to be detected, $\tilde{\mathbf{H}}_m=[\hat{\mathbf{H}}_{1,m}\cdots\hat{\mathbf{H}}_{N,m}]$,

$$\hat{\mathbf{H}}_{n,m} = \begin{bmatrix} \hat{h}_{n,m}(R) & \cdots & \hat{h}_{n,m}(1) & & \mathbf{0} \\ \vdots & \ddots & & \ddots & \vdots \\ \mathbf{0} & & \hat{h}_{n,m}(R) & \cdots & \hat{h}_{n,m}(1) \end{bmatrix}, \qquad (25)$$

$\tilde{\mathbf{x}} = [\mathbf{x}_1^T \cdots \mathbf{x}_N^T]^T$, and

$$\mathbf{x}_n = [x_n(t_0 - R + 1), \ldots, x_n(t_0), \ldots, x_n(t_0 + R - 1)]^T. \qquad (26)$$

Note that $\mathbf{y}_m$ is used to represent the measurements in Eqs. (5), (6), and (23) since $\mathbf{y}_m$ represents a portion of the received signal in any case. However, the use of $\mathbf{y}_m$ should be clear from the context. When detecting the symbols, the channel is assumed to be stationary, which allows keeping $\tilde{\mathbf{H}}_m$ in Eq. (23) constant.

If the measurements from all the receivers are stacked up together, we obtain

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \\ \vdots \\ \tilde{\mathbf{H}}_M \end{bmatrix} \tilde{\mathbf{x}} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{bmatrix} \qquad (27)$$

or

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{e}}. \qquad (28)$$

The transmitted symbols $\{x_n(t_0)\}$ are estimated using Eq. (28). When estimating $\{x_n(t_0+1)\}$, the measurement vector $\tilde{\mathbf{y}}$ is shifted by one symbol duration and so on. Yet, $\tilde{\mathbf{H}}$ remains constant since the channel is assumed to be stationary during the process.

B. The MMSE filter

In this section, we briefly review the Wiener filter,[65,66] which is optimal in the MMSE sense with respect to each transmitted symbol, for symbol detection. The Wiener filter is widely used in the communication literature[53,67,68] and the exposition provided in this section is for the sake of completeness. The steering vector corresponding to $\{x_n(t_0)\}$ in Eq. (28) is given by $\mathbf{d}_n \triangleq [\hat{\mathbf{h}}_{n,1}^T \cdots \hat{\mathbf{h}}_{n,M}^T]^T$, where $\hat{\mathbf{h}}_{n,m}$ are the estimates of $\mathbf{h}_{n,m}$ defined in Eq. (3). We let the symbol of current interest be $x_n(t_0)$. Then, the Wiener filter for this symbol, denoted as $\mathbf{g}_n$, can be derived by solving

$$\mathbf{g}_n = \underset{\mathbf{g}}{\operatorname{argmin}}\, E(\|\mathbf{g}^H\tilde{\mathbf{y}} - x_n(t_0)\|_2^2). \qquad (29)$$

The solution to Eq. (29) is[65,66]

$$\mathbf{g}_n = \mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1} E(x_n^H(t_0)\tilde{\mathbf{y}}), \qquad (30)$$

where $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ is the covariance matrix of $\tilde{\mathbf{y}}$, i.e., $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = E(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^H)$.

In the following, it is assumed that the payload sequences are pairwise uncorrelated, each payload sequence is uncorrelated with the noise $\tilde{\mathbf{e}}$, the noise has zero mean, each payload symbol is independent of the other payload symbols, and each payload symbol has zero mean. These assumptions can be stated mathematically as follows:

$$E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H) = \mathbf{I}, \quad E(\tilde{\mathbf{x}}\tilde{\mathbf{e}}^H) = \mathbf{0}. \qquad (31)$$

Using Eqs. (28) and (31), we obtain

$$\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = \tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \mathbf{R}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}} \qquad (32)$$

and $E(x_n^H(t_0)\tilde{\mathbf{y}}) = \mathbf{d}_n$. Equation (30) then becomes

$$\mathbf{g}_n = (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \mathbf{R}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}})^{-1}\mathbf{d}_n, \qquad (33)$$

and the soft estimate of the symbol $x_n(t_0)$ is given by $\mathbf{g}_n^H\tilde{\mathbf{y}}$. In our experiments we estimate $\mathbf{R}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}}$ from the residual error obtained during the channel estimation process, i.e., using $\mathbf{e}_m = \mathbf{y}_m - \mathbf{S}\hat{\mathbf{h}}_m$, $m = 1, \ldots, M$, in Eq. (14). Since digital communications require the receiver to make a hard decision, the nearest constellation point to $\mathbf{g}_n^H\tilde{\mathbf{y}}$ is selected as the symbol estimate.

C. Detection schemes

In the following, we will consider three approaches for applying the filters $\{\mathbf{g}_n\}$ to the measurements. We will note the relations between the approaches proposed in the communications literature with those in the spectral estimation area and propose a new scheme inspired by this relationship.

1. Linear combinatorial nulling (LCN)

In LCN,[52] $x_n(t_0)$ is detected using $\mathbf{g}_n^H\tilde{\mathbf{y}}$ for $n = 1, \ldots, N$ separately where for each n, other symbols are simply treated as interferences, i.e., the estimation of $x_n(t_0)$ has no effect on the estimation of $x_{n'}(t_0)$ $(n' \neq n)$. However, this approach shows poor performance when the channel coefficients for each transmitter differ significantly in magnitude. For instance, when the channel coefficients of the first transmitter dominate all the others, the symbol estimate for the first transmitter will be relatively accurate whereas the symbols sent from the other transmitters will be buried under the contribution from the first transmitter and hence they will be estimated inaccurately.

2. CLEAN-BLAST

The idea of sequential cancellation and nulling (SCN) can be used to alleviate the aforementioned drawback of LCN. As the name implies, SCN first detects the symbol with the strongest channel response. Then, the contribution of this symbol is removed from the measurements $\tilde{\mathbf{y}}$ (and the corresponding columns are removed from $\tilde{\mathbf{H}}$) before estimating the other symbols. This process continues until all the N symbols are estimated. The symbol with the strongest channel coefficients is detected first because it can be estimated more accurately than the other symbols with weaker channel coefficients. After the dominant symbols are subtracted from the measurements, the weaker symbols can be estimated more accurately. Sequential cancellation, from the viewpoint of the remaining symbols, can be recognized as interference removal. Eventually, when detecting the symbol with the weakest channel coefficients, no more interferences are present. The detection algorithm featuring SCN is called V-BLAST.[53] Herein, we name the algorithm as CLEAN-BLAST to emphasize its analogy to the CLEAN algorithm used in spectral estimation.[69]

3. RELAX-BLAST

As we have already pointed out, the relationship between LCN and CLEAN-BLAST is analogous to that of the periodogram and CLEAN.[54] In spectral estimation, RELAX is also called SUPER-CLEAN[50,51] since it is a recursive version of CLEAN but with much better performance. In the same spirit as RELAX, RELAX-BLAST first detects the symbol with the dominant channel taps and subtracts it out from $\tilde{\mathbf{y}}$. Then, it estimates the next dominant symbol from the residue signal. Unlike CLEAN-BLAST, however, which at this time estimates the third strongest symbol, RELAX-BLAST instead updates the two already detected symbols in an iterative manner until the difference of the RELAX-BLAST estimates between two successive iterations becomes less than a certain threshold. Once these two symbols are subtracted from the measurements and the third strongest symbol is estimated, the three symbols are again updated in an iterative manner until all the three estimates do not improve anymore. This process is repeated until all the N symbols are detected and updated.

Finally, note that when $N=1$, i.e., for a SIMO or SISO system, LCN, CLEAN-BLAST, and RELAX-BLAST become identical approaches.

V. NUMERICAL AND EXPERIMENTAL RESULTS

In this section we evaluate the performance of the CA training sequences, compare IAA with MP, OMP, and LSMP for channel estimation, and compare CLEAN-BLAST with RELAX-BLAST for symbol detection using simulations and/or the RACE'08 experimental results. Throughout this section, all the CIR estimation algorithms are followed by BIC to achieve sparsity.

A. Simulations

1. CIR estimation performance

To begin with, we consider the problem of CIR estimation for a 4×1 multi-input single-output (MISO) system with a time-invariant channel. The simulated CIR coefficients resemble real UWA conditions encountered in the RACE'08 experimental measurements. Figure 3 shows the modulus of the CIRs corresponding to the four transmitters where $R=30$ delay taps are considered. Given the training symbols, the received data samples are constructed using Eq. (5), where $\mathbf{e}_1$ is assumed to be a circularly symmetric independent and identically distributed (i.i.d.) complex-valued Gaussian random process with mean zero and variance σ^2.

Figure 4 shows the mean-squared error (MSE) of the channel estimates obtained by MP, OMP, LSMP, and IAA with two different training sequences: QPSK training and CA training. In QPSK training, each training symbol is randomly selected to be one of the four QPSK constellation points whereas in CA training each symbol is selected by using the CA algorithm described in Sec. III B. The training sequence length is set at $P=128$ symbols. Each point in Fig. 4 is obtained by averaging 100 Monte-Carlo trials. We observe that when the QPSK training is used, IAA significantly outperforms the other channel estimation methods. OMP and LSMP show similar performance whereas MP shows the worst performance. On the other hand, when the CA training sequences are used, the performance gap between IAA and the MP based channel estimation algorithms diminishes and all algorithms yield very similar performance although IAA still gives the lowest MSE. Moreover, the performance of IAA is not affected very much from the characteristics of the training sequences used. This is an advantage over the other methods since in the decision-directed mode, the channel has to be updated using the previously decoded symbols, which do not have as good auto- and cross-correlation properties as the specifically designed training sequences.

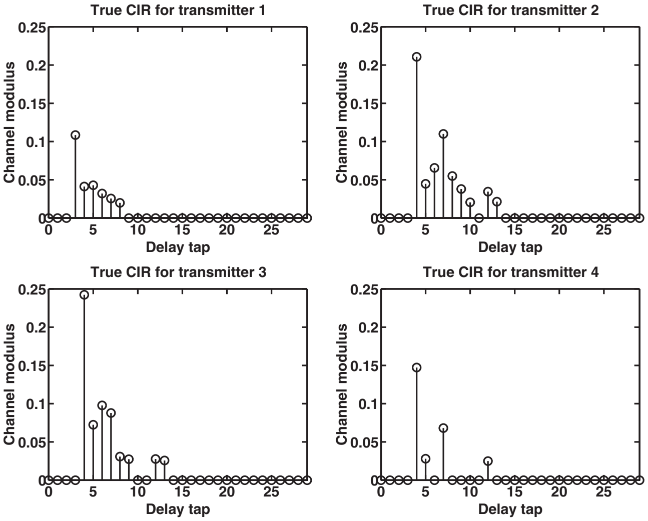


FIG. 3. The modulus of the simulated CIRs between the four transmitters and the receiver in a 4×1 MISO system.

2. Symbol detection performance

We now evaluate the bit error rates (BERs) of CLEAN-BLAST and RELAX-BLAST for a 4×12 MIMO system. The package structure shown in Fig. 1 is used in the simulations with CA training sequences consisting of $P=512$ symbols, a payload sequence consisting of 6000 QPSK modulated symbols, and two gaps consisting of 80 mute symbols each. IAA is used for channel estimation. The detection order for the algorithms is 3, 2, 4, and 1, i.e., the third channel is assumed to have the strongest channel response at all the receivers and the first channel the weakest. The average BERs over 100 Monte-Carlo trials are shown for the data transmitted from all four transducers in Fig. 5. We observe that RELAX-BLAST shows much better performance than

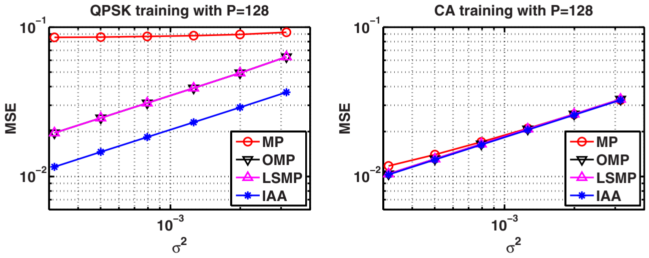


FIG. 4. (Color online) MSE of the CIR estimates for a 4×1 MISO system using the QPSK and CA training sequences with $P=128$ symbols. Each point is averaged over 100 Monte-Carlo trials.

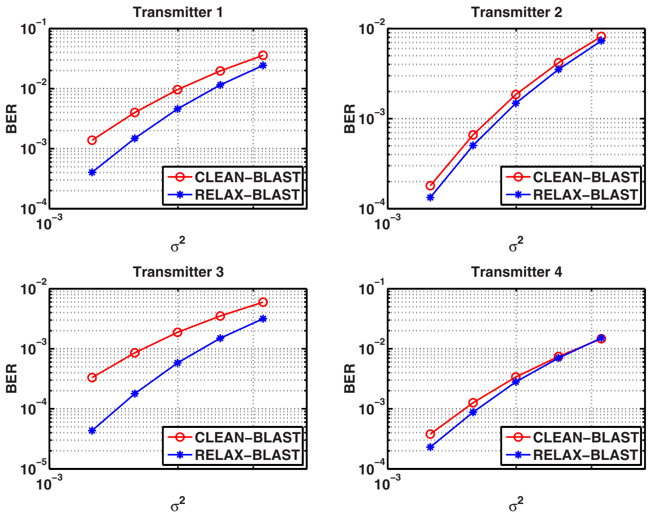


FIG. 5. (Color online) The BERs of each of the four transmitted payload sequences for a 4×12 MIMO system. The training sequences consist of $P=512$ symbols and are designed by the CA algorithm. The detection performance of CLEAN-BLAST and RELAX-BLAST is compared in terms of BER averaged over 100 independent Monte-Carlo trials for varying levels of the noise variance σ^2.

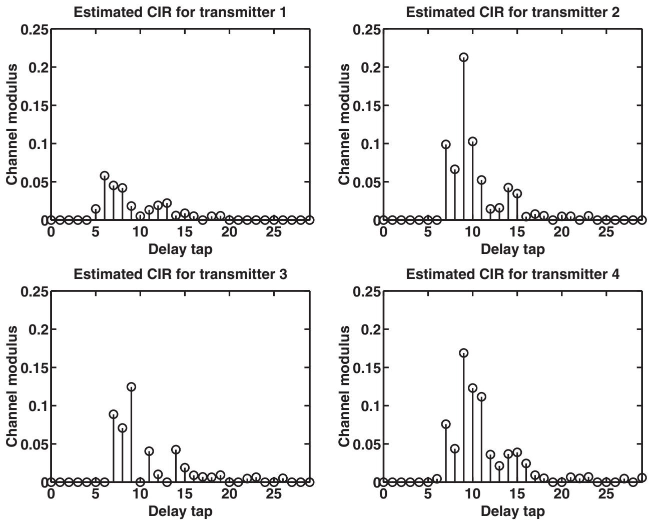


FIG. 6. The modulus of the four RACE'08 CIRs estimated by IAA for the first receiver from epoch "0830156".

CLEAN-BLAST as long as severe error propagation does not exist. This result is supported by the fact that similar performance improvements in spectral estimation are obtained when RELAX is used instead of CLEAN.[50,51] Due to this reason, we will use RELAX-BLAST when analyzing the RACE'08 data in the following.

B. RACE'08 experimental results

In this part, we evaluate our proposed MIMO underwater communications scheme using the RACE'08 experimental data set. RACE'08 was conducted by WHOI in Narragansett Bay. The water depths ranged from 9 to 14 m during the experiments. Surface conditions were primarily wind blown chop. A 4×24 MIMO system was used in the experiments. The primary transmitter was located approximately 4 m above the bottom of the ocean using a stationary tripod. Below the primary transmitter, a source array consisting of three transducers was deployed vertically with a spacing of 0.6 m between the elements. The top element of the source array was 1 m below the primary source. 24 receiving transducers were mounted at a range of approximately 400 m. Receivers were deployed vertically with a spacing of 0.05 m between the individual elements. The carrier frequency and the bandwidth employed in the RACE'08 experiments were 12 and 3.9 kHz, respectively.

The data packet that we will consider herein is from epoch "0830156". Some epochs could not be evaluated due to the severe conditions of sea. Among the many epochs that result in reasonable performance, epoch "0830156" was chosen arbitrarily. The package structure shown in Fig. 1 was used in the experiments with CA training sequences consisting of $P=512$ symbols, a payload sequence consisting of 2000 QPSK modulated symbols, and two gaps consisting of 80 mute symbols each. The symbol rate was 3906.25 symbols per second. By applying QPSK modulation and using four transmitters simultaneously, a 31.25 kbps uncoded payload data rate was achieved. The coding scheme we used for the experiments was a 1/2 convolutional encoder with constraint length of 5, and generator polynomials (1 0 0 1 1) and (1 1 0 1 1).[3] This coding scheme reduces the net payload data rate to 15.63 kbps.

The selection of the number of delay taps, R, to consider is very important. A value too small will loose important channel features whereas a value too large will complicate the receiver and may result in overfitting as well as increased noise. We found out empirically that $R=30$ yields reasonable results. Figure 6 shows the modulus of the training-directed IAA estimate of the CIR at receiver 1. The CIRs for the other receivers, i.e., $\{\hat{\mathbf{h}}_m\}_{m=2}^{24}$, share similar structure with $\hat{\mathbf{h}}_1$. As shown in Fig. 6, the detection order should be 2 (strongest coefficients), 4, 3, and 1 (weakest coefficients).

The channel tracking approach we follow is summarized in Fig. 7. In the first step, the CIR is estimated using the training sequences. Based on the initial CIR estimate, the first $L+50$ payload symbols are obtained using RELAX-BLAST, where L was defined after Eq. (8). Next, a decision-directed CIR estimation is done using the first L estimated symbols. The reason for not using all the $L+50$ estimated symbols will be explained shortly. With the updated CIR, starting from the $(L-49)$th symbol, the subsequent $L+100$ symbols are detected again using RELAX-BLAST. This process is repeated until all the 2000 payload symbols are detected. Figure 7 shows that 100 more symbols (50 more sym-

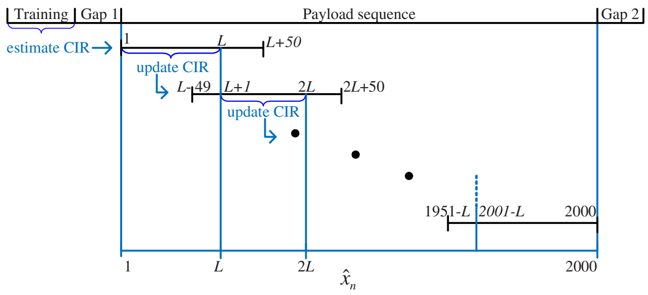


FIG. 7. (Color online) The channel tracking procedure.

TABLE IV. BER for L=200. Tx 1–4 stand for transmitters 1–4.

	Uncoded BER (%)				Coded BER (%)			
	Tx 1	Tx 2	Tx 3	Tx 4	Tx 1	Tx 2	Tx 3	Tx 4
MP	30.45	6.80	14.38	3.83	46.70	0	8.85	0
OMP	12.15	0.60	2.00	0.35	2.40	0	0	0
LSMP	12.15	0.60	2.00	0.35	2.40	0	0	0
IAA	4.63	0.10	0.35	0	0	0	0	0

bols at the first and last steps) are detected other than the L symbols used to update the CIR at each step. These 50 margin symbols on either end serve as guard intervals because the errors tend to happen at the beginning and end of each block. This is partly due to no mute symbols being available within the payload sequence.

In Table IV we show the uncoded and coded BERs obtained via MP, OMP, LSMP, or IAA as the channel estimation algorithm. For the results presented in this table, the number of payload symbols used for updating the channel coefficients is 200, i.e., L=200. We observe that IAA provides the best performance among all four algorithms. The average uncoded BER for IAA is 1.27%, MP is 13.86%, and OMP and LSMP is 3.78% and the coded average BER for IAA is 0%, MP is 13.89%, and OMP and LSMP is 0.6%. As expected, the sequence with the strongest (weakest) channel coefficients is estimated with the highest (lowest) accuracy, see Fig. 6.

In Table V the uncoded and coded BERs are shown for L=400. This means that the channel will be updated less frequently than in the case where L=200. We observe that now IAA, OMP, and LSMP show almost identical performance. The average uncoded BER for IAA is 0.38%, MP is 2.09%, and OMP and LSMP is 0.37% and the coded average BER for IAA is 0%, MP is 0.01%, and OMP and LSMP is 0%. As we mentioned previously, when L is large or the sequence used for updating the channel is well-structured, the performance of MP type of algorithms approaches that of IAA. However, it might not be always possible to select L large in practice.

The choice of L determines the rate at which the CIR will be updated in the decision-directed mode. It also determines the accuracy of the CIRs. As can be seen from Eq. (6), the larger the L, the more accurate the channel estimates will be assuming that the previously detected symbols are correct and the channel is stationary. On the other hand, for larger L, the channel will be updated less frequently and hence the results will be inaccurate for a rapidly varying channel. Therefore, the choice of L has a direct effect on the performances of MP, OMP, LSMP, and IAA. Moreover, L also determines the computational complexities of these algorithms. For the current set of data, we observed that the channel is rather benign and using a large L value results in better estimates than using a lower one, as seen in Tables IV and V. However, for a rapidly varying channel where L has to be selected small, IAA appears to be the best candidate for channel estimation as its performance is still good with small L whereas MP type of algorithms show relatively worse performance.

Note that in our experiments, neither the training sequence length P nor the gap lengths have been optimized for the best performance as no prior information of the experimental conditions was available. Moreover, for the current experimental conditions, a 1/2 rate convolutional code appears to be on the conservative side to achieve 0% coded BER.

VI. CONCLUSIONS

In this paper, we have focused on the various aspects of using a MIMO acoustic communications system in an underwater environment where delay spread is present. The problem was divided into two main parts: (i) channel estimation, which involves the design of the training sequences and the design of the algorithm to estimate the channel coefficients using the training sequences or previously detected symbols, and (ii) symbol detection. We have presented the CA for designing training sequences with good auto- and cross-correlation properties. IAA coupled with BIC and RELAX was presented as an approach for estimating the CIR. It was shown via simulations that IAA outperforms MP type of algorithms with arbitrary training sequences and it was shown

TABLE V. BER for L=400. Tx 1–4 stand for transmitters 1–4.

	Uncoded BER (%)				Coded BER (%)			
	Tx 1	Tx 2	Tx 3	Tx 4	Tx 1	Tx 2	Tx 3	Tx 4
MP	6.98	0.23	1.05	0.13	0.05	0	0	0
OMP	1.48	0	0	0	0	0	0	0
LSMP	1.48	0	0	0	0	0	0	0
IAA	1.50	0	0	0	0	0	0	0

via experimental data that IAA gives better results than MP type of algorithms when the number of symbols used for updating the channel is relatively small (a situation encountered in rapidly varying sea conditions). An extension to the widely used V-BLAST algorithm, namely, RELAX-BLAST, has been presented to improve detection performance. The validity of the proposed scheme was shown via both simulations and field data from the RACE'08 experiment.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Research (ONR) under Grant No. N00014-07-1-0193, the National Aeronautics and Space Administration (NASA) under Grant No. NNX07AO15A, and the National Science Foundation (NSF) under Grant No. ECS-0621879.

[1] J. Catipovic, "Performance limitations in underwater acoustic telemetry," IEEE J. Ocean. Eng. **15**, 205–216 (1990).

[2] J. Preisig, "Acoustic propagation considerations for underwater acoustic communications network development," ACM SIGMOBILE Mobile Computing Communications Review **11**, 2–10 (2007).

[3] J. G. Proakis, *Digital Communications*, 3rd ed. (McGraw-Hill, New York, NY, 1995).

[4] M. Stojanovic, J. Catipovic, and J. Proakis, "Phase-coherent digital communications for underwater acoustic channels," IEEE J. Ocean. Eng. **19**, 100–111 (1994).

[5] M. Chitre, S. Shahabudeen, and M. Stojanovic, "Underwater acoustic communications and networking: Recent advances and future challenges," Mar. Technol. Soc. J. **42**, 103–116 (2008).

[6] D. Kilfoyle and A. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).

[7] J. C. Preisig and G. B. Deane, "Surface wave focusing and acoustic communications in the surf zone," J. Acoust. Soc. Am. **116**, 2067–2080 (2004).

[8] D. J. Garrood, "Applications of the MFSK acoustic communications system," in Proceedings of the Oceans, Boston, MA (1981), pp. 67–71.

[9] A. Baggeroer, D. Koelsch, K. von der Heydt, and J. Catipovic, "DATS—A digital acoustic telemetry system for underwater communications," in Proceedings of the Oceans, Boston, MA (1981), pp. 55–60.

[10] E. Appleton, "Automatic synchronization of triode oscillators," Proc. Cambridge Philos. Soc. **21**, 231–248 (1922).

[11] G. Hsieh and J. Hung, "Phase-locked loop techniques—A survey," IEEE Trans. Ind. Electron. **43**, 609–615 (1996).

[12] T. H. Eggen, A. B. Baggeroer, and J. C. Preisig, "Communication over Doppler spread channels—Part I: Channel and receiver presentation," IEEE J. Ocean. Eng. **25**, 62–71 (2000).

[13] T. H. Eggen, J. C. Preisig, and A. B. Baggeroer, "Communication over Doppler spread channels—Part II: Receiver characterization and practical results," IEEE J. Ocean. Eng. **26**, 612–621 (2001).

[14] M. Stojanovic, "Recent advances in high-speed underwater acoustic communications," IEEE J. Ocean. Eng. **21**, 125–136 (1996).

[15] J. C. Preisig, "Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment," J. Acoust. Soc. Am. **118**, 263–278 (2005).

[16] D. R. Dowling, "Acoustic pulse compression using passive phase-conjugate processing," J. Acoust. Soc. Am. **95**, 1450–1458 (1994).

[17] T. C. Yang, "Temporal resolutions of time-reversal and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng **28**, 229–245 (2003).

[18] T. C. Yang, "Differences between passive-phase conjugation and decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **29**, 472–487 (2004).

[19] M. Johnson, L. Freitag, and M. Stojanovic, "Improved Doppler tracking and correction for underwater acoustic communications," in IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany (1997), Vol. **1**, pp. 575–578.

[20] L. Freitag, M. Johnson, and M. Stojanovic, "Efficient equalizer update algorithms for acoustic communication channels of varying complexity," in IEEE/MTS Oceans Conference (1997), Vol. **1**, pp. 580–585.

[21] T. C. Yang, "Channel Q function and capacity," in IEEE/MTS Oceans Conference (2005), Vol. **1**, pp. 273–277.

[22] T. C. Yang, "Correlation-based decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **30**, 865–880 (2005).

[23] T. C. Yang and A. Al-Kurd, "Performance limitations of joint adaptive channel equalizer and phase locking loop in random oceans: Initial test with data," in IEEE/MTS Oceans Conference (2000), Vol. **2**, pp. 803–808.

[24] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," IEEE Trans. Aerosp. Electron. Syst. (to be published).

[25] B. Song and J. Ritcey, "Spatial diversity equalization for MIMO ocean acoustic communication channels," IEEE J. Ocean. Eng. **21**, 505–512 (1996).

[26] S. Gray, J. Preisig, and D. Brady, "Multiuser detection in a horizontal underwater acoustic channel using array observations," IEEE Trans. Signal Process. **45**, 148–160 (1997).

[27] M. Nordenvaad and T. Oberg, "Iterative reception for acoustic underwater MIMO communications," in Proceedings of the Oceans, Boston, MA (2006), pp. 1–6.

[28] S. Roy, T. Duman, V. McDonald, and J. Proakis, "High-rate communication for underwater acoustic channels using multiple transmitters and space-time coding: Receiver structures and experimental results," IEEE J. Ocean. Eng. **32**, 663–688 (2007).

[29] D. Kilfoyle, J. C. Preisig, and A. B. Baggeroer, "Spatial modulation over partially coherent multiple-input/multiple-output channels," IEEE Trans. Signal Process. **51**, 794–804 (2003).

[30] D. Kilfoyle, J. C. Preisig, and A. B. Baggeroer, "Spatial modulation experiments in the underwater acoustic channel," IEEE J. Ocean. Eng. **30**, 406–415 (2005).

[31] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," IEEE J. Ocean. Eng. **32**, 927–939 (2007).

[32] H. Ryser, *Combinatorial Mathematics* (Wiley, New York, NY, 1963).

[33] W. Pratt, *Digital Signal Processing* (Wiley, New York, NY, 1978).

[34] S. Yang and J. Wu, "Optimal binary training sequence design for multiple-antenna systems over dispersive fading channels," IEEE Trans. Veh. Technol. **51**, 1271–1276 (2002).

[35] S. Wang and A. Abdi, "MIMO ISI channel estimation using uncorrelated Golay complementary sets of polyphase sequences," IEEE Trans. Veh. Technol. **56**, 3024–3039 (2007).

[36] P. Fan and W. H. Mow, "On optimal training sequence design for multiple-antenna systems over dispersive fading channels and its extensions," IEEE Trans. Veh. Technol. **53**, 1623–1626 (2004).

[37] J. Li, X. Zheng, and P. Stoica, "MIMO SAR imaging: Signal synthesis and receiver design," in The Second International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Saint Thomas, VI (2007), pp. 89–92.

[38] J. Li, P. Stoica, and X. Zheng, "Signal synthesis and receiver design for MIMO radar imaging," IEEE Trans. Signal Process. **56**, 3959–3968 (2008).

[39] B. K. Natarajan, "Sparse approximate solutions to linear systems," SIAM J. Comput. **24**, 227–234 (1995).

[40] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process. **41**, 3397–3415 (1993).

[41] S. F. Cotter, R. Adler, R. D. Rao, and K. Kreutz-Delgado, "Forward sequential algorithms for best basis selection," IEE Proc. Vision Image Signal Process. **146**, 235–244 (1999).

[42] W. Li, "Estimation and tracking of rapidly time-varying broadband acoustic communication channels," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2005).

[43] S. F. Cotter and B. D. Rao, "The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels," in 34th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA (2000), Vol. **2**, pp. 1772–1776.

[44] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," IEEE Trans. Commun. **50**, 374–377 (2002).

[45] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in IEEE/MTS Oceans Conference (1995), Vol. **3**, pp. 1417–1422.

[46] M. Stojanovic, L. Freitag, and M. Johnson, "Channel-estimation-based adaptive equalization of underwater acoustic signals," in IEEE/MTS Oceans Conference (1999), Vol. **2**, pp. 590–595.

[47] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," in IEEE International Conference on Communications, Paris, France (2004), Vol. **6**, pp. 3173–3177.

[48] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," IEEE Signal Process. Mag. **21**, 36–47 (2004).

[49] G. Schwarz, "Estimating the dimension of a model," Ann. Stat. **6**, 461–464 (1978).

[50] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," IEEE Trans. Signal Process. **44**, 281–295 (1996).

[51] J. Li, D. Zheng, and P. Stoica, "Angle and waveform estimation via RELAX," IEEE Trans. Aerosp. Electron. Syst. **33**, 1077–1087 (1997).

[52] R. L. Cupo, G. D. Golden, C. C. Martin, K. L. Sherman, N. R. Sollenberger, J. H. Winters, and P. W. Wolniansky, "A four-element adaptive antenna array for IS-136 PCS base stations," in 47th IEEE Vehicular Technology Conference (1997), Vol. **3**, pp. 1577–1581.

[53] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in Proceedings of the ISSSE, Pisa, Italy (1998), pp. 295–300.

[54] P. Stoica and R. L. Moses, *Spectral Analysis of Signals* (Prentice-Hall, Upper Saddle River, NJ, 2005).

[55] H. L. Van Trees, *Optimum Array Processing*, Detection, Estimation, and Modulation Theory (Wiley, New York, NY, 2002), Pt. IV.

[56] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," Astron. Astrophys. Suppl. Ser. **15**, 417–426 (1974).

[57] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," IEEE Trans. Inf. Theory **51**, 188–209 (2005).

[58] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1985).

[59] P. Tsai, H. Kang, and T. Chiueh, "Joint weighted least-squares estimation of carrier-frequency offset and timing offset for OFDM systems over multipath fading channels," IEEE Trans. Veh. Technol. **54**, 211–223 (2005).

[60] T. Koike, "Optimum-weighted RLS channel estimation for time-varying fast fading MIMO channels," in IEEE International Conference on Communications, Glasgow, Scotland (2007), pp. 5015–5020.

[61] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," IEEE Trans. Signal Process. **44**, 1469–1484 (1996).

[62] P. Stoica, H. Li, and J. Li, "A new derivation of the APES filter," IEEE Signal Process. Lett. **6**, 205–206 (1999).

[63] P. Stoica, A. Jakobsson, and J. Li, "Matched-filter bank interpretation of some spectral estimators," Signal Process. **66**, 45–59 (1998).

[64] G. H. Golub and C. F. V. Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, MD, 1989).

[65] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley, New York, NY, 1949).

[66] A. H. Sayed, *Fundamentals of Adaptive Filtering* (Wiley, New York, NY, 2003).

[67] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," IEEE Trans. Commun. **42**, 3178–3188 (1994).

[68] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," IEEE Trans. Inf. Theory **43**, 858–871 (1997).

[69] U. J. Schwarz, "Mathematical-statistical description of the iterative beam removing technique (Method CLEAN)," Astron. Astrophys. **65**, 345–356 (1978).

Noise reduction utilizing cross time-frequency ε-filter

Tomomi Abe[a)]
Pure and Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Mitsuharu Matsumoto
The Education and Research Center for Frontier Science, the University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan

Shuji Hashimoto
Department of Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

(Received 3 October 2008; revised 24 February 2009; accepted 24 February 2009)

A time-frequency ε-filter (TF ε-filter) is an advanced ε-filter applied to complex spectra along the time axis. It can reduce most kinds of noise while preserving a signal that varies frequently such as a speech signal. The filter design is simple and it can effectively reduce noise. It is applicable not only to small amplitude stationary noise but also to large amplitude nonstationary noise. However, when the filter is applied to the noise that varies much frequently along the time axis, TF ε-filter cannot reduce noise without the signal distortion. This paper introduces an advanced method for noise reduction that applies ε-filter to complex spectra not only along the time axis but also along the frequency axis labeled cross TF ε-filter. It can reduce the noise where the neighboring frequency bins have similar powers. To show the effectiveness of the proposed method, some comparative experiments are also given, such as the performance of noise reduction and the robustness concerning input signal-to-noise ratio and parameter changes.

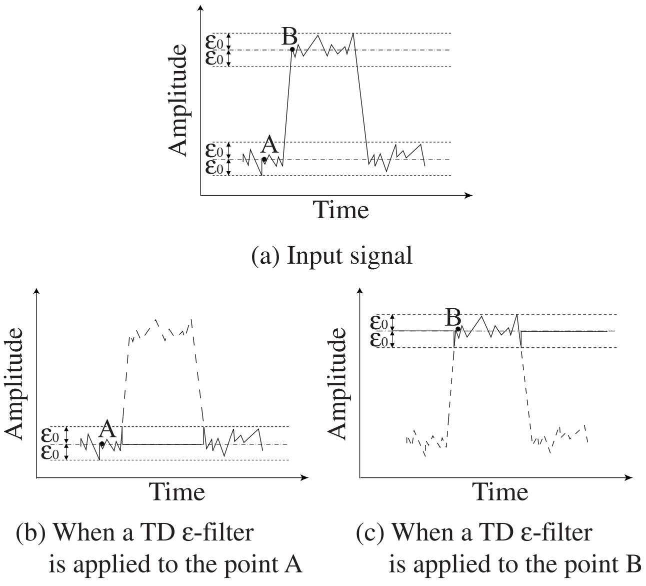
© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106126]

PACS number(s): 43.60.Hj, 43.60.Wy, 43.60.Ac [EJS] Pages: 3079–3087

I. INTRODUCTION

Although there are many studies about multichannel signal processing for noise reduction in acoustical signal processing such as microphone array,[1–3] independent component analysis,[4,5] and sparseness approaches,[6–8] single channel approaches have several advantages compared to multichannel approaches, e.g., system downsizing, system applicability, and system simplification. The spectral subtraction (SS) is a well-known approach for reducing the noise signal of single channel signal.[9–11] It can reduce the noise effectively despite the simple procedure. However, it can handle only the stationary noise. It also needs to estimate the noise in advance. Although noise reduction utilizing Kalman filter has also been reported,[12,13] the calculation cost is large. Some authors have reported a model based approach for noise reduction.[14] In this approach, we can extract the objective sound by learning the sound model in advance. However, it is not applicable to the signals with the unknown noise as well as SS. There are some approaches utilizing comb filter.[15] In this approach, we first estimate the pitch of the speech signal and reduce the noise signal utilizing comb filter. However, the estimation error results in the degradation of the speech quality. Some authors have reported the method using ε-filter.[16,17] The early ε-filter labeled a time-domain ε-filter (TD ε-filter) is a nonlinear filter, which can reduce the noise signal with preserving the signal. TD ε-filter is simple and has some desirable features for noise reduction. It does not need to have the model not only of the signal but also of the noise in advance. It is easy to be designed and the calculation cost is small. It can reduce not only the stationary noise but also the nonstationary noise. However, it can reduce only the small amplitude noise in principle. To solve the problems, the method labeled time-frequency ε-filter (TF ε-filter) was proposed.[18] TF ε-filter is an improved ε-filter applied to the complex spectra along the time axis in time-frequency domain. By utilizing TF ε-filter, we can reduce not only small amplitude stationary noise but also large amplitude nonstationary noise. However TF ε-filter cannot reduce the noise without distortion when the noise changes frequently along the time axis such as impulse noise. To solve the problem, we apply ε-filter to complex spectra not only along the time axis but also along the frequency axis labeled cross TF ε-filter. By applying ε-filter to the complex spectra along the two axes, we can reduce the noise even if it changes frequently along the time axis. It does not require the noise information as well as TF ε-filter in advance. We also show the experimental results of the proposed method compared to the other methods such as SS and TF ε-filter. In Sec. II, we explain an ε-filter and TF ε-filter to clarify the problems. In Sec. III, we describe the algorithm of cross TF ε-filter. In Sec. IV, we show the experimental results. To compare the performance of the proposed method, we also show the experimental results using SS and the method combining TF ε-filter and TD ε-filter. Conclusions are given in Sec. V.

[a)] Author to whom correspondence should be addressed. Electronic mail: tomomi@shalab.phys.waseda.ac.jp



(a) Input signal

(b) When a TD ε-filter is applied to the point A

(c) When a TD ε-filter is applied to the point B

FIG. 1. Basic concept of a TD ε-filter.

II. NOISE REDUCTION UTILIZING TD ε-FILTER AND TF ε-FILTER

To clarify the problems of a TD ε-filter,[16,17] 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 the output signal of the ε-filter at time k as follows:

$$y(k) = x(k) + \sum_{i=-P}^{P} a(i)F(x(k+i) - x(k)), \tag{1}$$

where $a(i)$ represents the filter coefficient. $a(i)$ is usually constrained as follows:

$$\sum_{i=-P}^{P} a(i) = 1. \tag{2}$$

The window size of the ε-filter is $2P+1$. $F(x)$ is the nonlinear function described as follows:

$$|F(x)| \leqslant \varepsilon_0: \ -\infty \leqslant x \leqslant \infty, \tag{3}$$

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:

$$F(x) = \begin{cases} x & (-\varepsilon_0 \leqslant x \leqslant \varepsilon_0) \\ 0 & (\text{otherwise}). \end{cases} \tag{4}$$

Figure 1 shows the basic concept of a TD ε-filter when Eq. (4) is utilized as $F(x)$. Figure 1(a) shows the waveform 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 larger than ε_0 by the value of point A. We then summate the signals in the same window. Figure 1(b) shows the basic concept of this procedure. The dotted line represents the points where the distance from A is larger 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. Similar operations are executed in all the points [see point B in Figs. 1(a) and 1(c)]. Consequently, we can reduce small amplitude noise near the processed point while preserving the speech signal.

A TD 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, TF ε-filter was proposed.[18] TF ε-filter utilizes the distribution difference of the speech signal and the noise in the frequency domain. The following assumptions regarding the sound sources are used.

- *Assumption 1*. Speech signal has greater variation in power than noise signal in the time-frequency domain.
- *Assumption 2*. Noise signal is distributed more uniformly and has less variation in the time-frequency domain compared to in the time domain.

Figure 2 depicts the speech signal and the white noise signal in the time and the time-frequency domains.

As shown in Fig. 2, assumptions 1 and 2 are fulfilled in the case of various noises 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 corresponding to the presence of active speech signal, the power of the noise with respect to the power of the signal is smaller than the power of the noise with respect to the power of the signal in the time domain. In TF ε-filter, we utilize this feature to apply an ε-filter to high-level noise.

Figure 3 illustrates the method combining TF ε-filter and TD ε-filter 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:

$$X(\kappa,\omega) = \sum_{l=-\infty}^{\infty} x(\kappa+l)W(l)e^{-j\omega l}, \tag{5}$$

where $W(l)$, κ, and ω represent the window function, the time frame in the time-frequency domain, and the angular frequency, respectively. κ and ω are discrete numbers. j represents the imaginary unit. Next we execute a TF ε-filter, which is an ε-filter applying to complex spectra along the time axis in the time-frequency domain, as shown in Fig. 3(2). In this procedure, $V(\kappa,\omega)$ is obtained as follows:

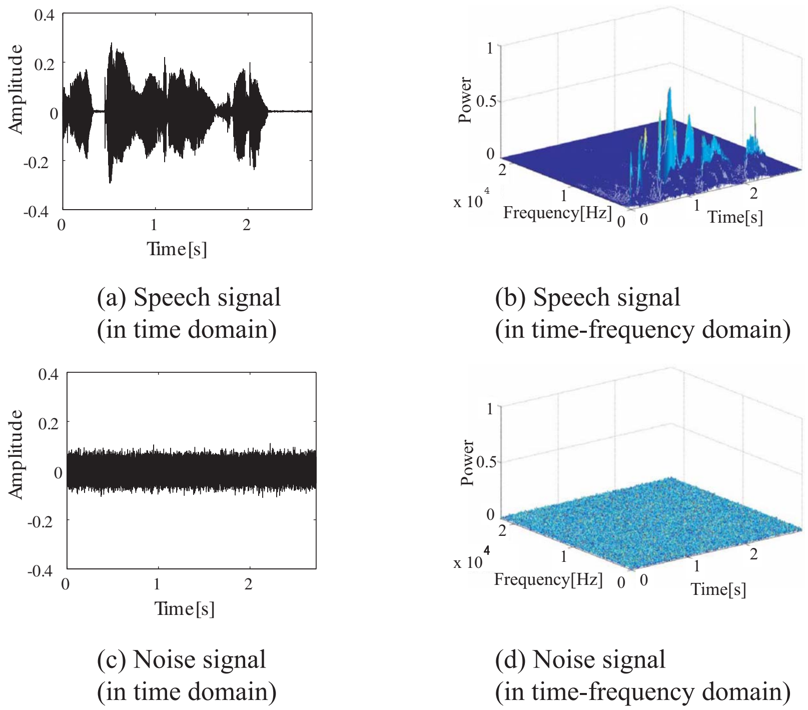
$$V(\kappa,\omega) = \sum_{i=-Q}^{Q} a(i)X'(\kappa+i,\omega), \tag{6}$$

where the window size of ε-filter is $2Q+1$,

$$X'(\kappa+i,\omega) = \begin{cases} X(\kappa,\omega)(\big||X(\kappa,\omega)| - |X(\kappa+i,\omega)|\big| > \varepsilon_T) \\ X(\kappa+i,\omega)(\big||X(\kappa,\omega)| - |X(\kappa+i,\omega)|\big| \leqslant \varepsilon_T) \end{cases} \tag{7}$$

and ε_T 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



(a) Speech signal (in time domain)

(b) Speech signal (in time-frequency domain)

(c) Noise signal (in time domain)

(d) Noise signal (in time-frequency domain)

FIG. 2. (Color online) A speech signal or noise signal in the time and time-frequency domains.

and the imaginary axis, respectively. In the following explanations, we basically use the word "signal" when we handle them as the symbols while we use the word "complex spectra" when we handle them as the values. We used the word "signal" as the mean of "all the signal points." We also used the word "complex spectra of the points" as the "all the complex amplitudes of the points." In Fig. 4, $*$ and $\times$ represent the processed point and the other signal points in the same window, 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 the points A and B, 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 summate the complex spectra of all the points in the same window. Due to handling complex spectra, when we have many signals that have similar powers but different phases, they are filtered out by the TF ε-filter and the complex amplitudes of the filter outputs become small. In other words, even if the absolute value of the noise is large, the noise is reduced because the phases of signal points are generally different in noise. Note that the noise is reduced not only when the power of the noise is small but also when the power of the noise is large because of this procedure. Figure 4(a) represents the basic concept in the case that the power varies frequently like in a speech signal. When we consider a signal whose power varies frequently, the differ-

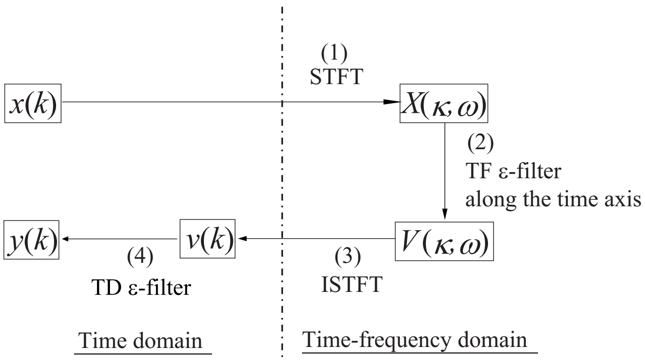
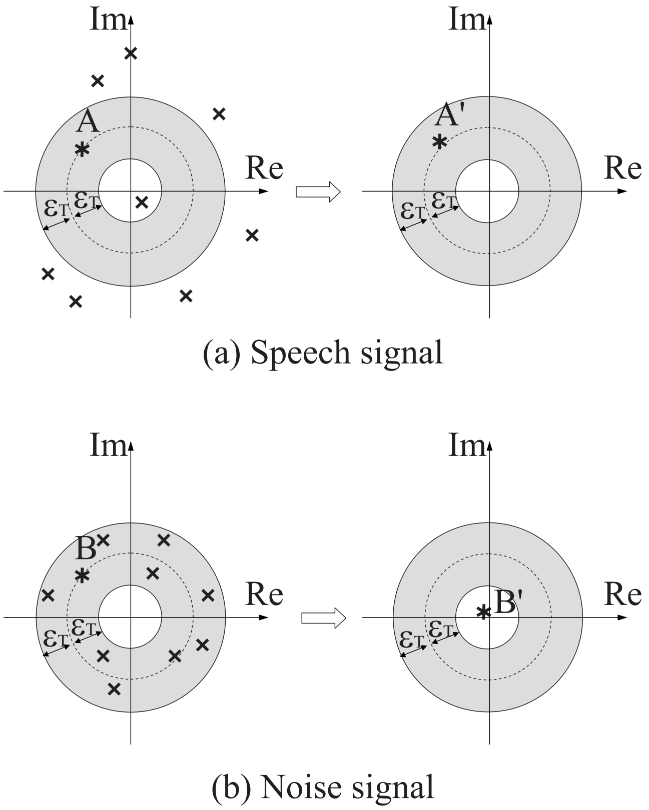


FIG. 3. Block diagram of combining TF ε-filter and TD ε-filter.



(a) Speech signal

(b) Noise signal

FIG. 4. Differences in performance when a TF ε-filter is applied to the speech signal and noise.

ence between the absolute value of A and that of the other signals is large, as shown in Fig. 4(a). For this reason, many signals in the same window as the point A are replaced by A. As a result, when we handle the speech signal, the complex amplitude of the processed point is almost preserved. Figure 4(b) represents the basic concept in case that the power does not vary so much like in 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_T appropriately. Hence, the TF ε-filter is effective even when the power of the noise with respect to the power of the signal is large. Additionally, under assumption 2, the TF ε-filter becomes more effective. When assumption 2 is satisfied, the variation of the noise with respect to the variation of the signal in the frequency domain becomes smaller than it is the case in the time domain. As a consequence, even if the noise varies frequently in the time domain, the ε-filter can be applied in the time-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:

$$y(k) = \sum_{i=-P}^{P} a(i)v'(k+i), \tag{8}$$

where the window size of ε-filter is $2P+1$,

$$v'(k+i) = \begin{cases} v(k) & (|v(k+i)-v(k)| > \varepsilon_t) \\ v(k+i) & (|v(k+i)-v(k)| \leqslant \varepsilon_t) \end{cases} \tag{9}$$

and ε_t is a constant.

III. NOISE REDUCTION UTILIZING CROSS TF ε-FILTER

TF ε-filter can reduce various types of noise effectively. However, when we use the noise that varies much frequently along the time axis, TF ε-filter cannot reduce noise without the signal distortion. When we consider the noise where the neighboring frequency bins have similar powers such as impulse noise, we can reduce the noise by using ε-filter applying to complex spectra not along the time axis but along the frequency axis.

Figure 5 shows the basic concept of the proposed method. At first, as shown in Fig. 5 "Step 1," we apply the ε-filter to the complex spectra along the frequency axis. This is to reduce the noise where the neighboring frequency bins have similar powers. By executing this process, we can reduce the noise whose amplitude varies frequently such as the impulse noise and white noise with large variation. Next we apply ε-filter to complex spectra along the time axis, as shown in Fig. 5 "Step 2." In Step 1, the noise where the neighboring frequency bins have similar powers is roughly reduced. Hence, ε-filter to complex spectra along the time axis is effective after Step 1. Figure 6 illustrates the proposed method with a block diagram. Let us consider $x(k)$ as the input signal. $x(k)$ is first transformed to the complex amplitude $X(\kappa,\omega)$ by STFT, as shown in Fig. 6(1). Next we apply ε-filter to complex spectra along the frequency axis, as shown in Fig. 6(2). In this procedure, $U(\kappa,\omega)$ is obtained as follows:

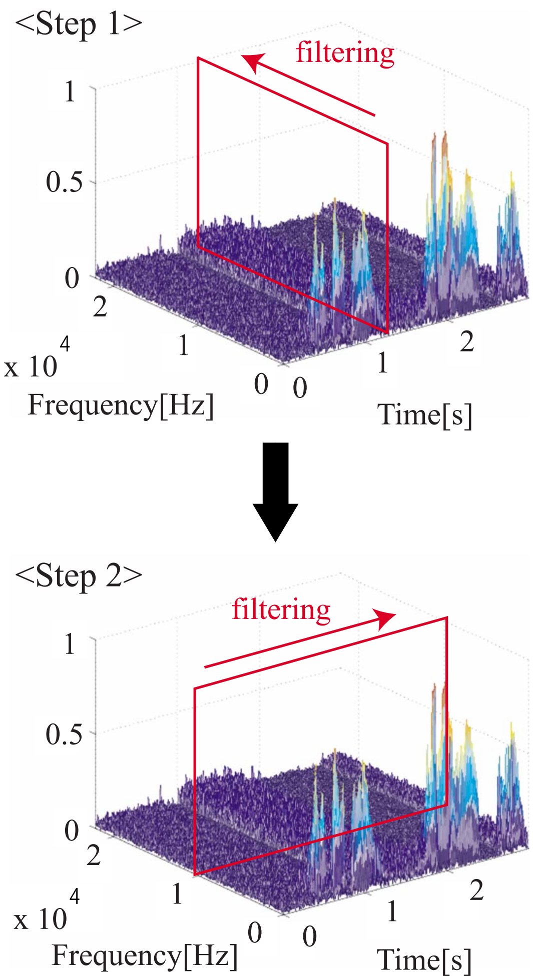


FIG. 5. (Color online) Basic concept of cross TF ε-filter.

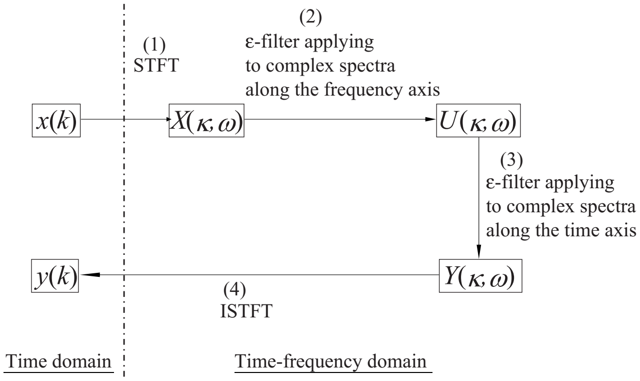


FIG. 6. Block diagram of the proposed method.

TABLE I. Common parameters.

Parameter	Value
Sampling frequency (Hz)	44 100
STFT Block size	512
Hop size	256
Window function	Hanning window

$$U(\kappa,\omega)=\sum_{i=-N}^{N} a(i)X'(\kappa,\omega+i), \tag{10}$$

where

$$X'(\kappa,\omega+i)=\begin{cases} X(\kappa,\omega) \\ (||X(\kappa,\omega)|-|X(\kappa,\omega+i)||>\varepsilon_F) \\ X(\kappa,\omega+i) \\ (||X(\kappa,\omega)|-|X(\kappa,\omega+i)||\leqslant\varepsilon_F). \end{cases} \tag{11}$$

ε_F is a constant and the window size is $2N+1$. Then we employ ε-filter applying to complex spectra along the time axis, as shown in Fig. 6(3). In this procedure, $Y(\kappa,\omega)$ is obtained as follows:

$$Y(\kappa,\omega)=\sum_{i=-M}^{M} a(i)U'(\kappa+i,\omega), \tag{12}$$

where

$$U'(\kappa+i,\omega)=\begin{cases} U(\kappa,\omega) \\ (||X(\kappa,\omega)|-|X(\kappa+i,\omega)||>\varepsilon_T) \\ U(\kappa+i,\omega) \\ (||X(\kappa,\omega)|-|X(\kappa+i,\omega)||\leqslant\varepsilon_T). \end{cases} \tag{13}$$

ε_T is a constant and $2M+1$ is the window size. Next we transform $Y(\kappa,\omega)$ to $y(k)$ by inverse STFT as shown in Figure 6(4). We label this process "cross TF ε-filter."

IV. EXPERIMENT

A. Experimental condition

We conducted the experiments utilizing a speech signal with a noise signal. As the speech signal, 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 were mixed in the computer. To compare the effectiveness of the proposed method to other methods, we conducted the experiments utilizing three methods; SS, the method combining TF ε-filter and TD ε-filter, and cross TF ε-filter. Table I shows the value of common parameters for all the experiments.

To evaluate the performance of noise reduction, we used signal-to-noise ratio (SNR), noise-reduction-ratio (NRR), and signal-to-distortion ratio (SDR). SNR is defined as follows:

TABLE II. SNR and SDR when a signal with stationary noise was utilized.

	SNR (dB)	SDR (dB)
Input signal	10.0	⋯
SS	18.3	21.7
Combination of TF ε-filter and TD ε-filter	40.6	18.9
Cross TF ε-filter	44.4	19.3

$$\mathrm{SNR}=10\log_{10}\left(\frac{\sum_{k=1}^{L}s(k)^2}{\sum_{k=1}^{L}n(k)^2}\right), \tag{14}$$

where $s(k)$, $n(k)$, and L represent the speech signal at time k, the noise signal at time k, and the length of the signal, respectively. To calculate SNR of the output signal, we separately applied each method to the signal and noise, and calculated the output SNR by using the obtained signal and noise. 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. To calculate $\mathrm{SNR}_{\mathrm{out}}$, we separately applied each method to the signal and noise, and calculated $\mathrm{SNR}_{\mathrm{out}}$ by using the obtained signal and noise. 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 and the output signal at time k respectively, when we used only the speech signal. SDR represents how much the signal is distorted by reducing the noise. Throughout all the experiments, the optimal parameters of SS and the method combining TF ε-filter and TD ε-filter were set empirically. To optimize the parameter, we conducted the experiment with changing the parameters. We employed the parameter that showed the best SNR. On the other hand, in cross TF ε-filter, ε_T was set at 0.1. We only changed ε_F depending on the noise to show the robustness concerning the parameter setting although we could reduce the noise more effectively. ε_F was set empirically to show the effectiveness compared to the other methods. SNR of the input signal was set at 10 dB throughout all of the experiments.

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 as a desired signal. We also prepared white noise that distributes uniformly as the stationary noise. We set ε_F in the proposed method at 0.7. We also set the window size of ε-filter applied to complex spectra in the proposed method along the frequency axis and the time axis at 101 and 11, respectively.

Table II shows the results of the experiments for stationary noise. As shown in Table II, the proposed method could reduce the noise compared to the other methods with pre-

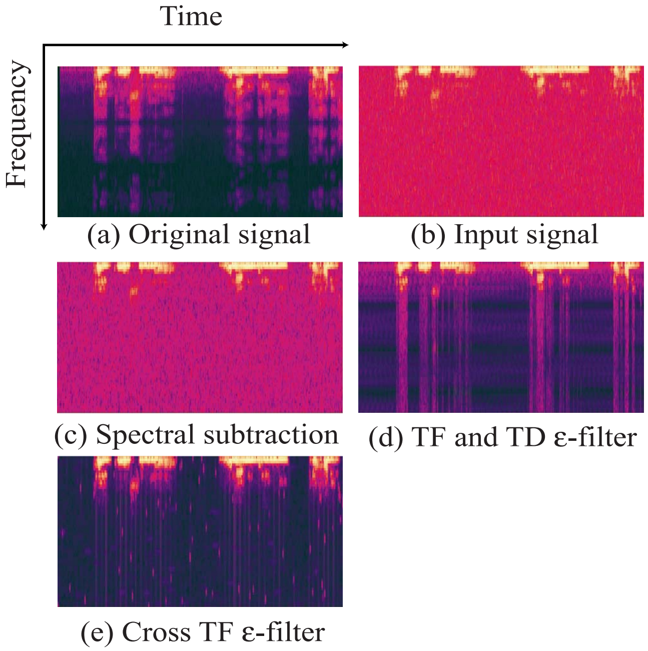


FIG. 7. (Color online) Experimental results when a signal with stationary noise is utilized.

serving the signal. Figure 7 shows the sound spectrograms. In Fig. 7, bright color represents high signal power while dark color represents low signal power. Figure 7(a) shows the spectrogram of the original signal. Figure 7(b) shows the spectrogram of the signal with stationary noise. Figures 7(c)–7(e) show the spectrograms of the output of SS, the output of the method combining TF ε-filter and TD ε-filter, and the output of the proposed method, respectively. As shown in Fig. 7, when SS and the method combining TF ε-filter and TD ε-filter are employed, the noises remained in the high-frequency bins, while the proposed cross TF ε-filter could reduce them. By using the proposed method, the noise could be reduced more effectively than using the other methods.

C. Experimental results in the case of nonstationary noise

The experiment was conducted using a signal with nonstationary noise. We used the same speech signal as in Sec. IV B. We prepared white noise with the amplitude that sometimes varied, as shown in Fig. 8. We set ε_F in the proposed method at 1.1. We also set the window size of ε-filter applied to complex spectra in the proposed method along the frequency axis and the time axis to 81 and 11 samples, respectively. Table III shows the results of the experiments on nonstationary noise. As shown in Table III, the SNR of the proposed method is superior to those of the other methods. Figure 9(a) shows the spectrogram of the original signal. Figure 9(b) shows the spectrogram of the signal with nonstationary noise. Figures 9(c)–9(e) show the spectrograms of the outputs of SS, the output of the method combining TF ε-filter and TD ε-filter, and the output of the proposed method, respectively. The relation between the color and signal power is the same as in Sec. IV B. As shown in Fig. 9, when we used the proposed method, the noise could be reduced more effectively than using the other methods even if we use the nonstationary noise.

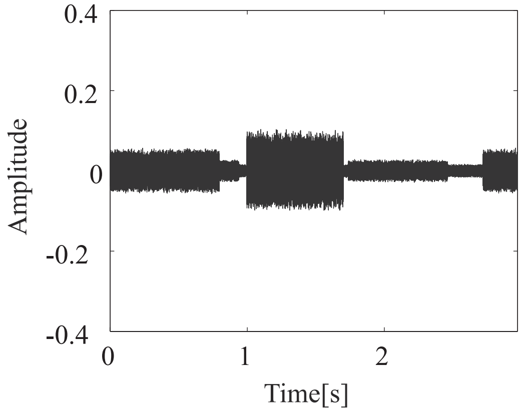


FIG. 8. Waveform of the nonstationary noise.

TABLE III. SNR and SDR when a signal with nonstationary noise was utilized.

	SNR (dB)	SDR (dB)
Input signal	10.0	⋯
SS	15.5	21.9
Combination of TF ε-filter and TD ε-filter	40.8	16.3
Cross TF ε-filter	44.2	17.3

D. Experimental results in the case of natural noise

To evaluate the performance of the proposed method for natural noise, we conducted the experiment utilizing a speech signal and a noise generated from the cooling fan of a personal computer. Most powers of noise used in this experi-

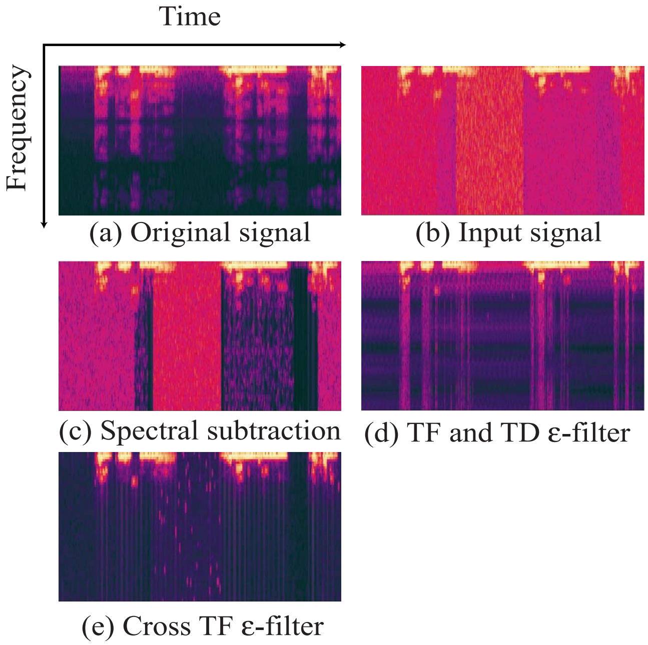


FIG. 9. (Color online) Experimental results when a signal with nonstationary noise is utilized.

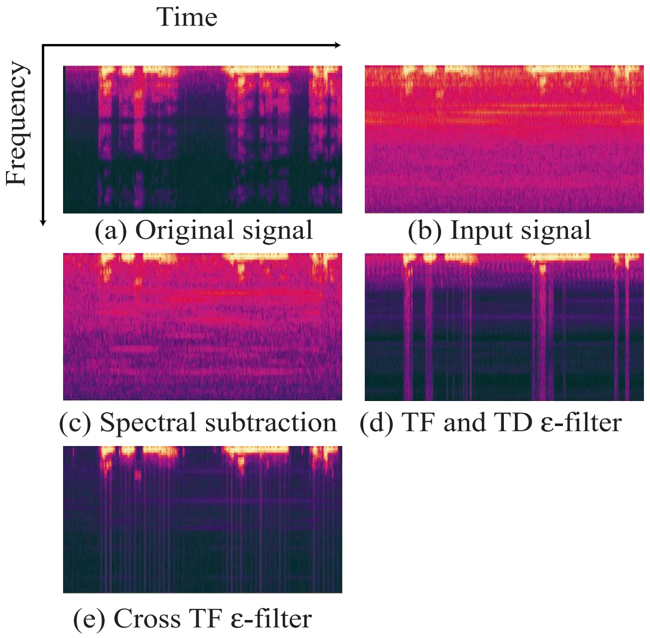
TABLE IV. SNR and SDR when a signal with natural noise was utilized.

	SNR (dB)	SDR (dB)
Input signal	10.0	···
SS	17.4	20.1
Combination of TF ε-filter and TD ε-filter	38.2	13.4
Cross TF ε-filter	40.2	15.0

ment are distributed in the low-frequency range. We set ε_F in the proposed method at 1.8. We also set the window size of ε-filter applied to complex spectra in the proposed method along the frequency axis and the time axis to 51 and 11 samples, respectively. Table IV shows the results of the experiments for natural noise. As shown in Table IV, the SNR of the proposed method is superior to those of the other methods as well as in the case of stationary noise and nonstationary noise. Figure 10(a) shows the spectrogram of the original signal. Figure 10(b) shows the spectrogram of the signal with natural noise. Figures 10(c)–10(e) show the spectrograms of the outputs of SS, the output of the method combining TF ε-filter and TD ε-filter, and the output of the proposed method, respectively. The relation between the color and signal power is the same as in Sec. IV B. As shown in Fig. 10, when we use the proposed method, the noise could be reduced more effectively than the other methods even if the natural noise was used as noise.

E. Robustness for various SNR

We also conducted the experiments utilizing the signal with various noise levels to confirm that the proposed method can be applied not only to the small amplitude noise but also to the large amplitude noise. We used five signals with nonstationary noise whose SNRs are −5, 0, 5, 10, and 15 dB, respectively. Figure 11 shows the experimental results about NRR. As shown in Fig. 11, NRR of SS is only about 5 dB; however, the proposed method can reduce much more noise than the conventional methods. Figure 12 depicts the experimental results about SDR. NRR and SDR of the proposed method are better than those of the method combining TF ε-filter and TD ε-filter.



(a) Original signal (b) Input signal

(c) Spectral subtraction (d) TF and TD ε-filter

(e) Cross TF ε-filter

FIG. 10. (Color online) Experimental results when a signal with natural noise is utilized.

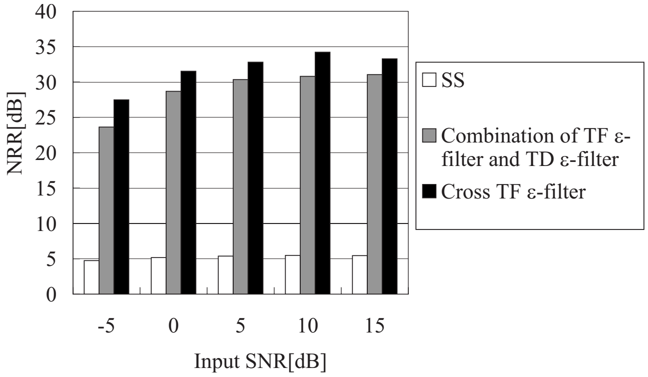


FIG. 11. Experimental results about NRR.

F. Evaluation experiment concerning the change of ε_F and the window size of the first ε-filter

We also conducted evaluation experiments concerning the change of ε_F and window size of the first ε-filter of the proposed method. This is because it seems to us that it is more useful to know the relation between NRR and these parameters for practical use. We can expect the performance change by the shift of each parameter with knowledge about the relation among the values of the parameters and NRR. We utilized the signal with the nonstationary noise utilized in Sec. IV C. The ε_F is changed from 0.1 to 1.0 by 0.1 interval. Window size of the first ε-filter is also changed from 21 to 181 samples by 40-sample interval. Figure 13 shows the experimental results regarding NRR. As shown in Fig. 13, the NRR does not depend on the window size of the first ε-filter very much. NRR increases depending on ε_F.

G. Evaluation using MSE

Although NRR and SDR are evaluated in the previous sections, they are measured not by the signal with noise but by the independent signal and noise. Due to this reason, it may not be adequate as the evaluation functions to show the

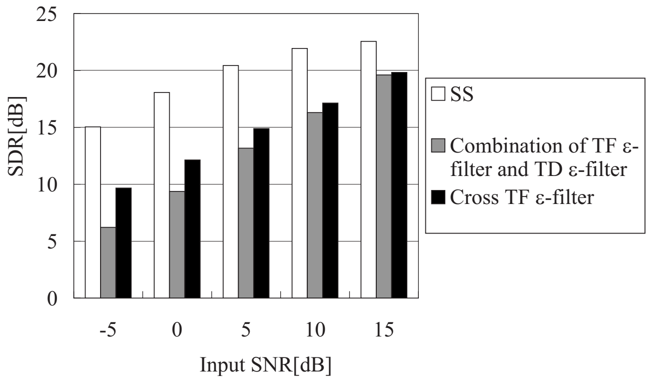


FIG. 12. Experimental result about SDR of the output signals.

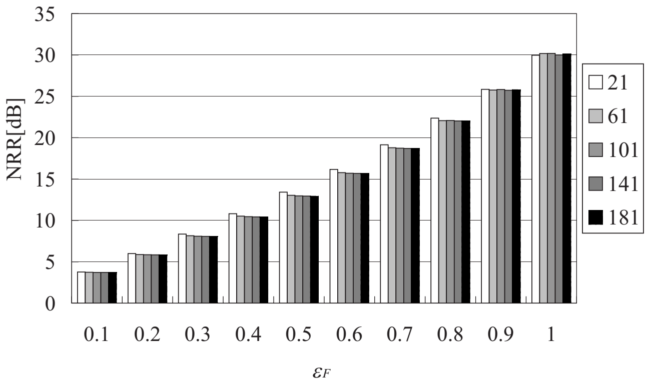


FIG. 13. NRR results by change of the ε_F and the window size.

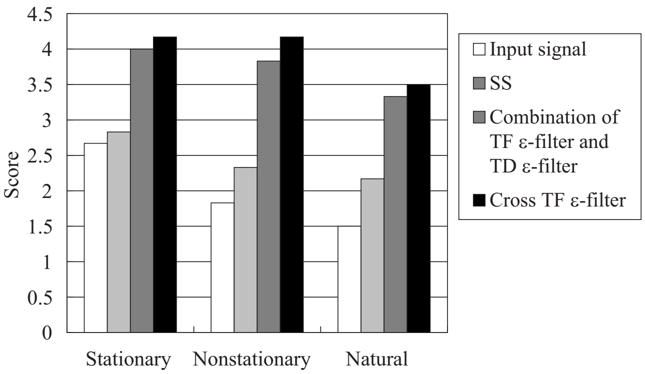


FIG. 14. Experimental results on subjective evaluation.

real performance of the proposed method on the signal with noise. Hence, we conducted the additional experiment on quantitative evaluation using the signal with noise. The performance of the proposed method is confirmed by using mean square error (MSE).

We used the signal that is identical to Sec. IV C as input signal. We calculated MSE concerning the output of SS, that of the method combining TF ε-filter and TD ε-filter, and that of the proposed method. MSE is defined as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y(i) - s(i))^2. \qquad (17)$$

Table V shows the results of the experiment on MSE. As shown in Table V, the signal processed by the proposed method is closer to the original signal than the signal processed by the other methods. The results show that the proposed method can reduce even when we employ not the signal and noise independently but the signal with noise.

H. Results of the experiment on subjective evaluation

We conducted an experiment on subjective evaluation. We used three signals that are identical to Secs. IV B–IV D as input signals. The examinees listened to the four signals: input signal, the output of SS, the output of the method combining TF ε-filter and TD ε-filter, and the output of the proposed method for three types of input signals. The examinees rated each signal on a scale of 1 to 5. The score 1 is the worst rating while score 5 is the best as auditory impression. The six examinees participated in this experiment.

Figure 14 shows the results of the experiment on subjective evaluation. As shown in Fig. 14, the signal processed by the proposed method shows better results than any other signals.

TABLE V. MSE when a signal with nonstationary noise was utilized.

	MSE ($\times 10^{-4}$)
Input signal	4.62
SS	1.74
Combination of TF ε-filter and TD ε-filter	1.23
Cross TF ε-filter	0.97

V. DISCUSSION AND CONCLUSION

In this paper, we introduced an algorithm for noise reduction applying ε-filter to complex spectra not only along the time axis but also along the frequency axis in time-frequency domain. The proposed method can reduce not only stationary noise but also nonstationary and natural noise effectively with preserving signal clarity. The experimental results showed that the proposed method could be applied to various kinds of noise. The proposed method could reduce the louder noise effectively compared with the conventional methods such as the method combining TF and TD ε-filter and SS. The proposed method is also robust for the change of the window size and the input SNR. Figure 15 shows the application range of the proposed method compared to the ε-filter and TF ε-filter. As shown in Fig. 15, the application range of the proposed method is expanded. It is applicable to the signal with noise whose amplitude varies widely along the time axis in addition to the signal that TF ε-filter handles. For future works, we would like to improve the performance of noise reduction to fill the shortage in Fig. 15. We would like to evaluate the performance of the proposed method as preprocessing of speech recognition system. We also aim to determine each parameter adaptively. We are considering to the parameter optimization of ε by using statistical constraints.

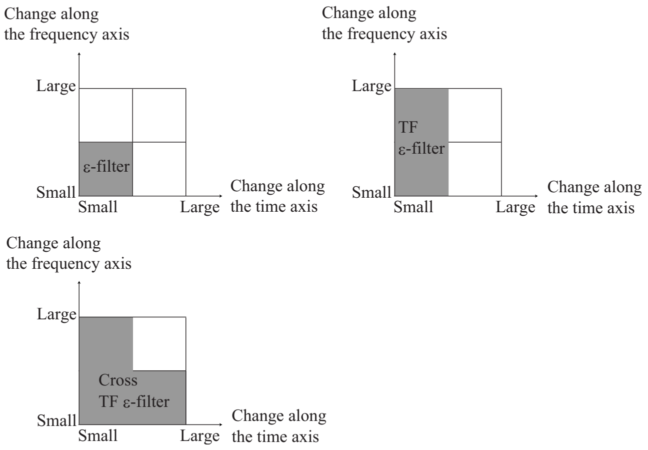

Change along the frequency axis
Large
Small
ε-filter
Small
Large
Change along the time axis
Change along the frequency axis
Large
Small
TF ε-filter
Small
Large
Change along the time axis
Change along the frequency axis
Large
Small
Cross TF ε-filter
Small
Large
Change along the time axis

FIG. 15. Application range of ε-filter, TF ε-filter, and cross TF ε-filter.

ACKNOWLEDGMENTS

This research was supported by the research grant of Support Center for Advanced Telecommunications Technology Research (SCAT), by the research grant of Foundation for the Fusion of Science and Technology, by the research grant of Tateisi Science and Technology Foundation, and by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 20700168, 2008. This research was also supported by Waseda University Grant for Special Research Projects (B), Nos., 2007B-142, 2007B-143, and 2007B-168, 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, Ministry of Education, Culture, Sports, Science and Technology, Japan, by the CREST project "Foundation of technology supporting the creation of digital media contents" of JST, by the Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture, the 21st Century Center of Excellence Program, "The innovative research on symbiosis technologies for human and robots in the elderly dominated society," Waseda University, by Project for Strategic Development of Advanced Robotics Elemental Technologies (NEDO: 06002090), and the Global-COE Program, "Global Robot Academia," Waseda University.

[1] K. Sasaki and K. Hirata, "3D-localization of a stationary random acoustic source in near-field by using 3 point-detectors," Trans. CSICE **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. Matsumoto and S. Hashimoto, "A miniaturized adaptive microphone array under directional constraint utilizing aggregated microphones," J. Acoust. Soc. Am. **119**, 352–359 (2006).

[4] A. J. Bell and T. J. Sejowski, "An information maximization approach to blind separation and blind deconvolution," Neural Comput. **7**, 1129–1159 (1995).

[5] H. Saruwatari, S. Kurita, and K. Takeda, "Blind source separation combining frequency-domain ICA and beamforming," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Process 2001*, Utah, Canada (2001), pp. 146–157.

[6] T. Ihara, M. Handa, T. Nagai, and A. Kurematsu, "Multi-channel speech separation and localization by frequency assignment," IEICE Trans. Fundamentals **J86-A**, 998–1009 (2003).

[7] S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Process 2002*, Orlando, FL (2002), pp. 529–532.

[8] M. Aoki, Y. Yamaguchi, K. Furuya, and A. Kataoka, "Modifying SAFIA: Separation of the target source close to the microphones and noise sources far from the microphones," IEICE Trans. Fundamentals **J88-A**, 468–479 (2005).

[9] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**, 113–120 (1979).

[10] J. S. Lim, *Speech Enhancement* (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[11] 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**, 471–472 (1978).

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," J. Basic Eng. **82**, 35–45 (1960).

[13] 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).

[14] P. Daniel, W. Ellis, and R. Weiss, "Model-based monaural source separation using a vector-quantized phase-vocoder representation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Process* (2006), pp. V-957–V-960.

[15] 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).

[16] 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).

[17] 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).

[18] T. Abe, M. Matsumoto, and S. Hashimoto, "Noise reduction combining time-domain ε-filter and time-frequency ε-filter," J. Acoust. Soc. Am. **122**, 2697–2705 (2007).

Ray-based acoustic localization of cavitation in a highly reverberant environment

Natasha A. Chang and David R. Dowling
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109

(Received 3 October 2008; revised 9 February 2009; accepted 11 February 2009)

Acoustic detection and localization of cavitation have inherent advantages over optical techniques because cavitation bubbles are natural sound sources, and acoustic transduction of cavitation sounds does not require optical access to the region of cavitating flow. In particular, near cavitation inception, cavitation bubbles may be visually small and occur infrequently, but may still emit audible sound pulses. In this investigation, direct-path acoustic recordings of cavitation events are made with 16 hydrophones mounted on the periphery of a water tunnel test section containing a low-cavitation-event-rate vortical flow. These recordings are used to localize the events in three dimensions via cross correlations to obtain arrival time differences. Here, bubble localization is hindered by reverberation, background noise, and the fact that both the pulse emission time and waveform are unknown. These hindrances are partially mitigated by a signal-processing scheme that incorporates straight-ray acoustic propagation and Monte-Carlo techniques for compensating ray-path, sound-speed, and hydrophone-location uncertainties. The acoustic localization results are compared to simultaneous optical localization results from dual-camera high-speed digital-video recordings. For 53 bubbles and a peak-signal to noise ratio frequency of 6.7 kHz, the root-mean-square spatial difference between optical and acoustic bubble location results was 1.94 cm. Parametric dependences in acoustic localization performance are also presented.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097465]

PACS number(s): 43.60.Jn, 43.60.Uv, 43.30.Nb [AIT] Pages: 3088–3100

I. INTRODUCTION

Fluid flows are commonly studied via scaled experiments in wind and water-tunnels. Here aero- and hydro-acoustic noise sources may be the focus of the experiments or they may provide information about invisible or otherwise obscure flow phenomena. Hence, acoustic detection, localization, and characterization of flow-induced sound sources are common experimental objectives in wind and water tunnel testing. Unfortunately, the geometry of wind and water tunnel test sections typically leads to echoes and reverberation while the tunnel's prime mover, flow-control, and manipulation elements (turning vanes, flow straighteners, screens, etc.) may unintentionally produce background noise. Thus, acoustic techniques for investigating flow-induced noise sources in wind and water-tunnels must remain effective when reverberation and noise are present. This paper describes a ray-based acoustic localization technique that should have generic applicability in wind and water tunnel testing, and in other reverberant and noisy environments. The technique, a mild extension of prior work, relies on direct-path sound to mitigate the effects of reverberation, and it uses multiple recording transducers and statistical signal-processing methods to address the deleterious effects of noise and environmental uncertainties.

Multiple recording transducers with known locations are commonly used in reverberant and noisy environments to detect and locate sound sources, and these transducers may be mounted on a carriage that moves across the domain of interest to search for the loudest source. In wind tunnels, it is possible to use different acoustic apertures and configurations. For example, an acoustic mirror may be used to obtain aero-acoustic sound-source locations by scanning scaled test models of trains (Nagakura, 2006; Grosche and Meier, 2001) or airframes (Dobrzynski *et al.*, 2001). Alternatively, multiple microphones (tens or hundreds) can be used to mitigate reverberation and locate point- and distributed-sound sources (Wang *et al.*, 2004; Gerard *et al.*, 2005; Arnold, *et al.*, 2003). However, there is a relative dearth of published articles on the topic of acoustic localization of sound sources in water-tunnels. The U.S. Navy's Large Cavitation Channel has an array of hydrophones for locating sound sources along the centerline of its test section via near-field beamforming (Etter *et al.*, 2005). Instead of localization, the acoustic studies conducted in water-tunnels typically characterize the hydro-acoustic sound source, and excellent summaries and reviews of cavitation and other hydro-acoustic sound sources are available (Arndt, 2002; Blake, 1986a, 1986b; Brennen, 1995).

For cavitating flows, optical localization methods have been preferred, and are successful when the domain of interest is relatively well defined, near a hydrofoil tip for example, and the cavitation bubbles are large enough to be visualized. Unfortunately optical methods become tedious, imprecise, or impractical when the domain of interest is less well defined, the cavitation bubbles are too small to be readily visualized, the cavitation event rate is low, or the test facility or test geometry does not allow optical access to the domain of interest. The acoustic localization technique described here overcomes these limitations, and it provides comparable accuracy to optical methods when applied to an unsteady cavitating vortex flow near inception. Furthermore, the time-domain technique described here is based on the simplest and most generic sound propagation, straight rays, so that it can be deployed in the widest possible range of test facilities. More sophisticated frequency-domain sound localization schemes, like matched field processing (Jensen *et al.*, 1994), require extensive experiment-specific geometrical and environmental information, and detailed sound field calculations. Collection of such information and/or completion of such calculations may be too burdensome for many sound-source localization applications.

For this study, an unsteady cavitating flow composed of two counter-rotating vortices of unequal strength was investigated. Here, the stronger vortex distorted and stretched the weaker one. When the stretching was severe enough and the static pressure was low enough for inception, the weaker vortex cavitated at a low rate (~1 Hz) in a spatial volume (10 l) too large for routine optical interrogation. This type of vortex-interaction cavitation can take place in shear flows (Iyer and Ceccio, 2002), jets (Katz and O'Hern, 1986; O'Hern, 1990), and ducted propulsor systems where the interaction of leakage and tip vortices (weak and strong vortices) may induce cavitation before the primary tip vortex (Chesnakas and Jessup, 2003; Oweis *et al.*, 2006a, 2006b). For military propulsor systems that require stealth, a high priority is placed on knowing the flow characteristics that lead to cavitation inception. Thus, the ability to acoustically detect and locate incipient cavitation while developing and testing prototype propulsion systems is desirable, especially since optical access to the region of interest may not be available.

Ray-based acoustic localization techniques have been developed for numerous tracking and localization applications involving: autonomous underwater vehicles (Caiti *et al.*, 2005), a human voice in a room (Handzel and Krishnaprasad, 2002), animals calls (Spiesberger and Fristrup, 1990; Spiesberger, 1998, 1999a), aircraft transit (Ferguson, 1999), and underwater sound sources (Westwood and Knobles, 1997; Voltz and Lu, 1994; Skarsoulis, 2005). Inverse active-sonar problems are also of interest here. For example, if a source at a known location broadcasts a sound signal, then the relative location of a receiver array of fixed geometry may be determined (Aarabi, 2002). Similarly, the location of obstacles can be estimated for robotics (Jiménez *et al.*, 2005). The majority of such applications are in multipath environments where the characteristic signal wavelength is smaller than the characteristic dimensions of the environment so that direct-path and reflected-path signals are commonly distinct at a receiver. In this case the reflected-path signal and knowledge of the environment can potentially aid source localization, and just few receivers can be used to obtain an accurate source location (Perkins and Kuperman, 1990).

For the most general localization problem, the signal waveform and emission time are not known. In the present effort, the dimensions of the environment are such that the first ~0.1 ms of the signal may not contain reflected-path signal energy. Consequently, the most useful signal bandwidth lies at and above 10 kHz. However, in a water-tunnel-test section, the many-possible once-reflected paths are commonly about the same length so multiple reflected-path signals may arrive at any receiver at approximately the same time. Without a detailed model of the environment, these once-reflected signals and later arrivals are of little use for localization purposes since the individual path contributions cannot be separated. Thus, in this investigation, the various hydrophone recordings are time-gated to suppress reflections and the reverberant signal coda.

In ocean and geoacoustic applications, small- or sparse-array acoustic localization techniques may exploit the multipath environment to estimate the source location and/or properties of the environment. Collins and Kuperman (1991) developed an extension of conventional match-field processing, focalization, that simultaneously obtains the source location and environmental parameters. Similarly, Westwood and Knobles (1997) used time delays of signal arrivals from three hydrophones (direct and reflected) to obtain a correlogram that was then matched to the best fit of a series of simulated correlograms for source tracks where the environmental parameters and locations were varied.

An alternative localization method, which was developed to track marine and terrestrial animals, relies on direct-path signals and more receivers. Here each ray path is assumed to be straight but the sound speed may be different for each path so that refraction is accounted for approximately but not resolved (Spiesberger, 2004). In an environment with constant sound speed, this method reduces to triangulation (Schmidt, 1972). In general, three-dimensional localization requires a minimum of five receivers (Spiesberger, 2001). Uncertainties in the acoustic environment and the location of the receivers can then be handled by assigning appropriate distributions for the uncertain geometrical parameters, and solving for potential source locations from array subgroups having the minimum required number of receivers. Here, the final location and correct environmental parameters would be a common solution from all of the subgroups (Spiesberger, 1999b, 2005).

The present localization effort is based on the adaptation of ray-methods developed for the ocean environment to water tunnel experiments. As many as 16 receiving hydrophones are used and the possible signal bandwidth extends from 1 to 200 kHz. The water tunnel test-section is a three-dimensional highly reverberant environment where even direct-path acoustic rays are typically refracted at one or more material interfaces, and direct-path signal waveforms can be distorted by passage through different materials. Although such complexities could be overcome with a detailed environmental model, the present effort seeks to determine the performance of a facility-nonspecific acoustic localization system based on generic straight-ray acoustic propagation. Here, statistical signal processing is used to at least partially mitigate the errors that arise from this oversimplification of the actual acoustic propagation.

The remainder of this paper is divided into four sections. Section II describes the experimental setup and data acquisition techniques. Section III covers the signal-processing algorithms used. The comparisons of optical and acoustical localization results are provided in Sec. IV. Section V summarizes this effort and states the conclusions drawn from it.

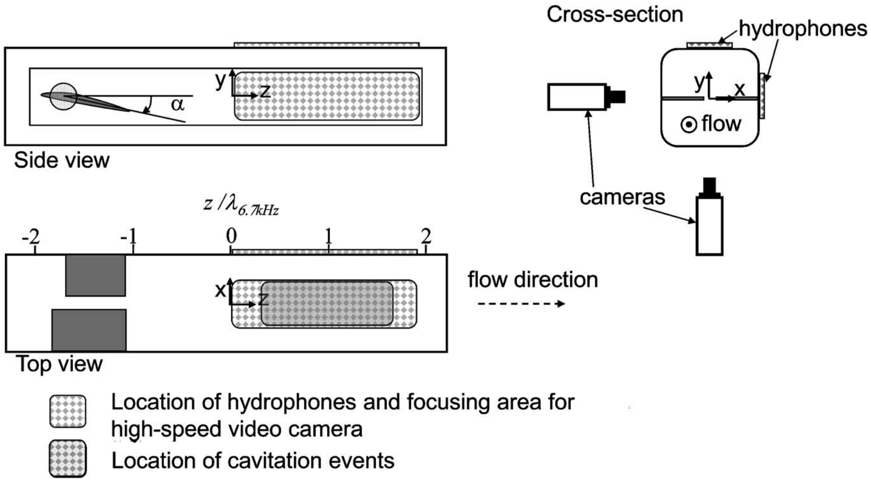


FIG. 1. Schematic of the water tunnel test section with the acoustical and high-speed video camera setup. The experimental coordinate system was centered in the middle of the water tunnel test section with z-axis pointing in the downstream direction.

II. EXPERIMENTAL SETUP

This experimental investigation involved three primary elements: a water-tunnel, an optical localization system, and an acoustic localization system. These elements are described in turn below.

The water-tunnel had a nominally square test section, 23×23 cm^2, with an overall length of 1.0 m. The test-section flow speed and static pressure could be controlled between 0–18 m s^{-1} and ~30–200 kPa absolute, respectively. In addition, the cavitation nuclei content and distribution in the tunnel water could be altered with a de-aeration system, and monitored with a dissolved oxygen sensor (Orion model 810) and a cavitation susceptibility meter (GEC Alsthom ACB). Two hydrofoils were placed at the upstream boundary of the test section to generate a pair of parallel counter-rotating vortices of different strengths that would interact in the remainder of the test section. At certain flow speeds and foil attack angles, the resulting vortex interactions would cause the weaker vortex to cavitate first as the static pressure in the test section was lowered. When weaker vortex inception occurred first, individual cavitation events were erratic and their locations were spread out over the final 40% of the test section. The coupling of the vortex hydrodynamics and the resulting cavitation in this flow is described in Chang (2007). The cavitation nuclei content and distribution in the tunnel water set the rate of intended cavitation events in the test section (signal sources), and the amount of unintended sheet cavitation elsewhere in the water-tunnel (noise sources). After some exploration of the available parameter space, an acceptable experimental configuration was found that allowed the acoustic localization system to be developed in parallel with an independent visual localization method for validation. Here, the test-section-inlet flow speed U, static pressure P_s, dissolved oxygen content, and test-section inlet cavitation number $\sigma_\infty=(P_s-P_v)/(\frac{1}{2}\rho U^2)$ were 10.0 m/s, 157 kPa, 25% of saturation, and 3.1, respectively, where P_v is water vapor pressure and ρ is water density, both at the nominal test-section temperature of 25 °C. All results presented here are derived from this experimental configuration which provided the best acoustical signal-to-noise ratio (SNR) with visually discernible bubbles. Unfortunately, the unintentional cavitation noise produced in this configuration curtailed the usable signal bandwidth to 1 to 30 kHz. Other test conditions more favorable for acoustic localization did not produce visually discernible bubbles and therefore were not suitable for validation of the acoustic localization technique.

At the chosen test condition, the cavitation bubbles were elongated along the weaker vortex's axis and were 1–10 or so millimeters in length, just large enough to be reliably imaged. For the comparisons below, the optically-determined locations of extended bubbles were bubble-image centroids. The bubbles were illuminated through a light diffuser from behind by four 300 W incandescent lights and imaged by two 8-bit Phantom V9.0 high-speed movie cameras with 50 or 85 mm Nikon lenses in conjunction to 12 mm extension rings. Bubble images were recorded under a variety of experimental settings. Image spatial resolution was either 526×1200 pixels (60 mm across by 150 mm along the flow) or 504×528 pixels (approximately 45 mm square) with frame rates of 2500 and 5400 fps, respectively. Exposure times varied between 31 and 180 μs. To determine bubble locations in three dimensions, the cameras were focused in the same region of the tunnel test section but their viewing directions were at right angles, i.e., from the side and from below (see Fig. 1). Bubble locations and sizes were calibrated by imaging a ruler placed in the cameras' overlapping field of view. The error in optically locating a bubble in the 10 m/s flow was estimated to be ±4 mm based on the camera's frame rate and calibration accuracy.

The acoustic measurements were made with an array of 16 hydrophones, Reson TC-4013, with a receiving sensitivity of -211 ± 3 dB re 1 V/μPa and 3-dB bandwidth from 1 Hz to 170 kHz. Each phone was driven by a Reson VP-2000 voltage preamplifier powered by 12 V from a Hewlett-Packard E3610 A dc power supply. The preamplifiers' pass-band was 100 Hz–1 MHz with a gain of 10 dB. The signal from each preamplifier was further bandpass filtered between 1 and 200 kHz and amplified with a gain of 40 dB with a Khron-Hite 3364 four-pole tunable active filter. Here the filter type was Butterworth and the attenuation was 24 dB/octave. The 16 amplified and bandpass-filtered signals where digitized with four four-channel 12-bit National Instruments A/D boards, NI-PCI-6110 S, acquiring data synchronously

TABLE I. Location of the hydrophones in the water-tunnel.

Hydrophone	x (cm)	y (cm)	z (cm)
1	12.8	2.4	37.4
2	12.8	−2.4	35.5
3	12.8	0.0	31.7
4	12.8	2.4	27.3
5	12.8	−2.4	25.4
6	12.8	0.0	21.6
7	12.8	2.4	17.1
8	12.8	−2.4	15.2
9	−2.5	12.8	38.5
10	−2.4	12.8	30.3
11	0.0	12.8	31.5
12	2.4	12.8	30.3
13	−2.5	12.8	24.5
15	1.5	12.8	17.5
15	−1.0	12.8	4.5
16	0.0	12.8	2.0

with a sample-timing error of less than 50 ns at a sampling rate 1 MHz per channel. Timing errors between the different hydrophones (phase error) were measured with all the phones placed equidistant from a broadcast transducer, an ITC 1042, and simultaneously recording sound pulses broadcast to every receiving phone. Here geometric positioning error of the receiving phones was estimated to be less than ± 0.2 mm [a time delay uncertainty of O(10^{-7}) s]. Cross correlations of the recorded pulses from different receiving phones produced a time delay uncertainty of $\sim 10^{-6}$ s which corresponds to a distance of ~ 1.5 mm in water.

The hydrophones were mounted on two adjoining sides of the water tunnel test section in cavities within two of the tunnel's acrylic windows. The cavities were filled with quiescent de-aerated water, and were separated from the test-section flow by a 1 cm thick acrylic plate. These separating plates had slightly more than twice the characteristic impedance of water, 3.2 MPa s/m vs 1.5 MPa s/m. The tip of each hydrophone was spaced 1 cm from the separating plate to avoid the evanescent pressure fluctuations from the turbulent boundary layer on the inner test-section wall. Figure 1 is a schematic drawing showing hydrophone geometry, the experimental coordinate system, and the region where cavitation events occurred. The placement of the hydrophones was not optimized. Nonetheless, it followed some of the basic principles of array design for improving the localization results by placing receivers with the greatest span possible in each coordinate direction. Plus, the hydrophone placement was not symmetrical to help the localization scheme in avoid dominant side lobes (see Steinberg, 1976 or Ziomek, 1995). For the tests reported here, vortex cavitation events occurred in the downstream half of the water tunnel test-section, so the phones were placed in that half, eight on top and eight on one side, as shown in Fig. 1 and listed in Table I.

A sound projector, ITC-1042, was used in the process of developing and validating the acoustic localization system. The description of the cavitation signature from work by Brennen (1995), Chesnakas and Jessup (2003), and Oweis *et al.* (2004) was used as a reference to generate a series of representative synthetic signals. An incepting cavitation bubble produces sound via volume acceleration that occurs during bubble growth, collapse, or oscillation. The sound signature generally is broadband spanning from a few hertz to several hundreds of kHz, Brennen (1995) and Oweis *et al.* (2004). In a recent study of vortex cavitation, Chesnakas and Jessup (2003) recorded cavitation bubbles that produced popping noises typically associated bubble growth and collapse, as well as tonal bursts with a center frequency of a few kilohertz that were labeled chirps. Hydrodynamic interactions between the bubble and the vortex are believed to set the oscillation frequency (Choi *et al.*, 2009). A function generator, Tektronix AFG 320, was used to generate a series of single-period sine wave pulses from 10 to 180 kHz. This pulse was then amplified with a Khron-Hite 7602 M wideband amplifier and broadcast in the tunnel at a known location $(x,y,z)=(-11\text{ mm},-21\text{ mm},152\text{ mm})$. The synthetic cavitation signal was recorded with the acoustic localization system described above. Sample recorded signals are shown in Fig. 2.

During an experiment where cavitation events were being located, both the optical and acoustical systems acquired data continuously into data buffers. To synchronize the two data streams, a trigger from one of the hydrophones prompted each system to tag and save its respective buffer. However, the point that each system regarded as $t=0$ was the first datum after the trigger. This means that the video system, which has a sampling rate two orders of magnitude slower than the acoustic system, would have a time delay with respect to the acoustic data that could range from zero to the frame period (0.400 or 0.185 ms). In addition, the acoustic signal received was time delayed with respect to the image viewed. This time delay is estimated to be 0.080 ms based on the general location of the cavitation events with respect to the trigger hydrophone. Thus, overall error in timing of the acoustic signal with respect to the video data was -0.080 to $+0.400$ or -0.080 to $+0.185$ ms.

It was found that the acoustic signal from a cavitation event was approximately 1–15 ms long. Specifically there were two types of cavitation signals, a *pop* that was generally broadband and lasted at most 2 ms (see Fig. 3) and a *chirp* that persisted for several milliseconds which would embody at least one strong tone, typically 3–6 kHz (see Fig. 4). The maximum amplitude of the acoustic signal typically occurred during bubble growth; thus, the bubbles were localized both acoustically and visually at this point, as illustrated in Fig. 4.

III. SIGNAL PROCESSING

A. Determining time delays

For the chosen experimental condition, 10.0 m/s, 25% dissolved oxygen content, and 157 kPa static pressure, acoustic data were collected from all 16 receiving hydrophones at 1.0 MHz per channel synchronously with the dual-camera high-speed video data. Visually detectable cavitation events created acoustic signatures that had relatively high signal to noise ratio, so such events could be detected by

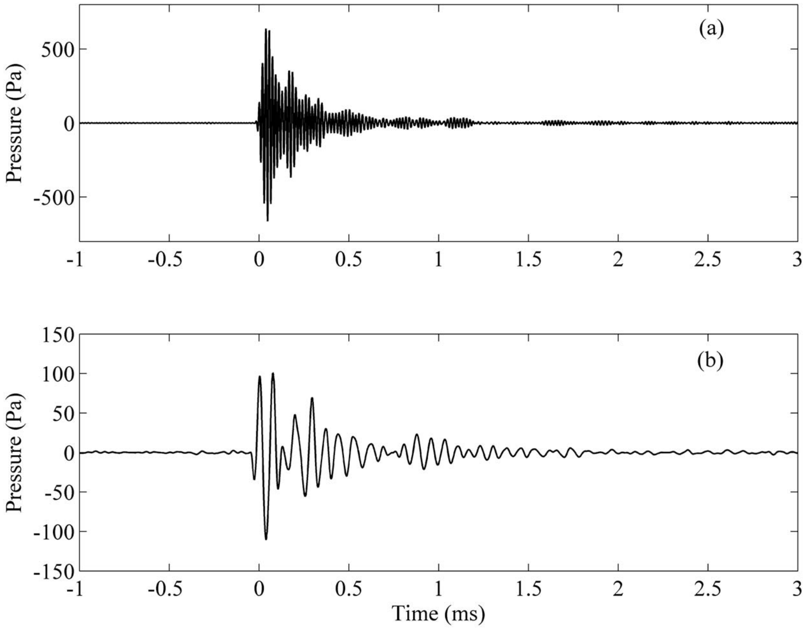


FIG. 2. Typical recorded signals (hydrophone 14) when the sound projector (ITC-1042) broadcasted a synthetic cavitation pulse, 1 cycle of a sine wave, at different frequencies: 50 kHz (a) and 10 kHz (b).

thresholding the output from one centrally located receiver. The signal from this centrally located receiver was used to trigger the tagging and saving of the recorded acoustic and video data.

The technique for isolating the direct-path sound in each hydrophone's recording does require some knowledge of the experimental geometry and signal characteristics. During the experiments, constructive interference of multiply-reflected signals, leading to a signal amplitude peak, was not observed with any regularity. Thus, the detection scheme was based on the assumption that reliable signal peaks from cavitation events were composed of first reflections only, see Fig. 5. Furthermore, estimates of the amplitude and timing differences between direct and once-reflected paths are needed to set the final time window.

The experimental SNR was determined by averaging 153 uncorrelated spectral samples of signal plus noise, $\tilde{M}(f)=\tilde{S}(f)+\tilde{N}(f)$, and noise-alone, $\tilde{N}(f)$, that were collected from each hydrophone. For the samples of $\tilde{M}(f)$, the strongest part of the cavitation signal was windowed from the acoustic time series using a Tukey window of 0.3 ms in length and a taper-to-constant-duration ratio of 0.5. Here, 2^{12}-point fast Fourier transforms were used, and the noise-alone time series were windowed identically to the signal-plus-noise time series. Zero padding was used as necessary. The SNR as function frequency f was then estimated via

$$\mathrm{SNR}(f) = 10\log_{10}\left(\frac{|\tilde{M}(f)|^2 - |\tilde{N}(f)|^2}{|\tilde{N}(f)|^2}\right), \tag{1}$$

and SNR results are plotted in Fig. 6.

The time delay information necessary for acoustic localization was determined from the acoustic signal from the cavitation bubble that traveled to each receiver along the most direct path. Unfortunately, once-reflected paths in many cases arrived sometime during the direct-path time period so determining and placing a temporal window to isolate direct-path sound in each hydrophone recording involved a trade-off; enough of the signal must be kept for effective cross-correlation processing, but the signal coda must be suppressed to prevent reflected-path corruption of the signal timing results. The Tukey window described above was found to be adequate for the effort described here. Unfortunately, the symmetry of the tunnel's test-section geometry generally leads to multiple once-reflected signals arriving together to reinforce each other at any receiver. In addition, propagation through the acrylic plate that separated the hydrophones from the flow distorted the direct-path signal

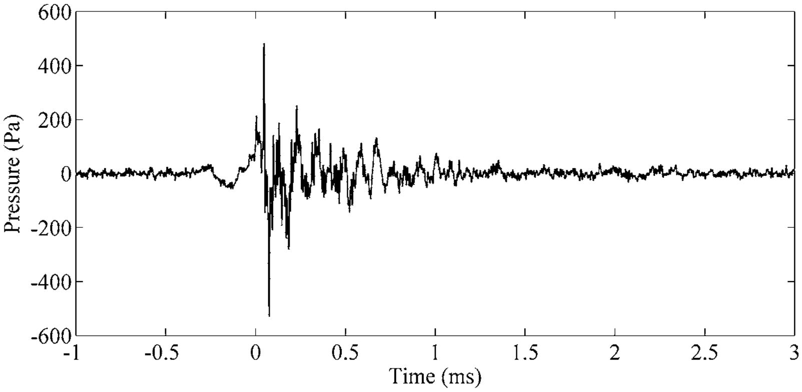


FIG. 3. Typical recorded signal (hydrophone 14) from a cavitation sound source generating a *pop*, a broadband signal lasting less than 2 ms.

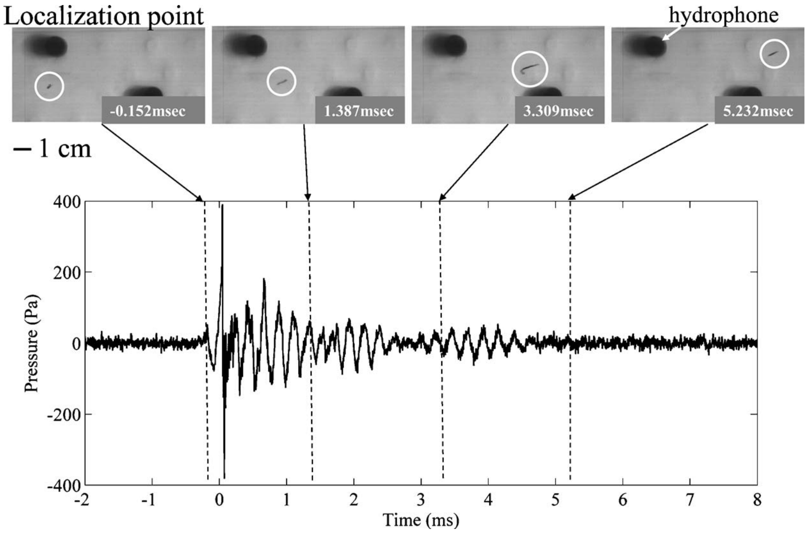


FIG. 4. Typical recorded signal (at hydrophone 14) and camera images from a cavitation bubble producing a *chirp* signal. This signal includes a tone at approximately 6 kHz that persists with some modulation for several milliseconds. In this investigation, cavitation-bubble localization occurred just before the highest amplitude portion of the signal.

waveform, especially at angles of incidence greater than the critical angle. Sound may pass through the acrylic plate as compression, shear, or evanescent waves (see Fig. 7) and the combination of these leads to geometry-dependent signal distortion at the recording hydrophones. In spite of this, the loudest part of the signal typically could be detected via a simple threshold, but this loudest part did not necessarily correspond to direct-path sound. This is illustrated in Fig. 8 which shows synthetic [(a)–(c)] and actual [(d)–(f)] cavitation pulse shapes for near normal [(a) and (d)], near critical [(b) and (e)], and above critical [(c) and (f)] angles of incidence. Here the critical angle is $\sin^{-1}(c/c_a)$, where c and c_a are the compression wave sound speed in the water and in the acrylic plate, respectively. Thus, careful placement of the direct-path-timing window within any of the 16 measured time series was required for robust localization. The following paragraphs describe the window-placement method developed in this effort.

First, a cavitation event was detected by thresholding the output from one centrally located receiver. The 16 time series were then coarsely time windowed from 10 ms before to 10 ms after the time of detection. These 16 coarsely-windowed time series could potentially contain signals from multiple cavitation events; however, the nuclei content in the water-tunnel and the static pressure were set so that the event rate in the tunnel test section was at most 3/s. The event rate was examined both acoustically and visually throughout the entire water tunnel test section. Thus, examining only 20 ms of data at a time decreased the likelihood that the coarsely-windowed time series included multiple events. These 16 coarsely-windowed time series contained the sought-after direct-path signal, reflected-path echoes, and background noise, so a finer time windowing was needed.

This final time windowing was accomplished by considering the average SNR between 1 and 30 kHz in a sliding 0.3 ms Tukey window. This frequency range contained half of the signal energy and was found to provide a reliable indication of the time of peak signal strength. Thus, the peak of the resulting fine-window SNR for each receiver was located and assumed to correspond to the arrival of multiple first reflections that reinforced each other. For the current test-section geometry, three first-reflected paths could be anticipated, so

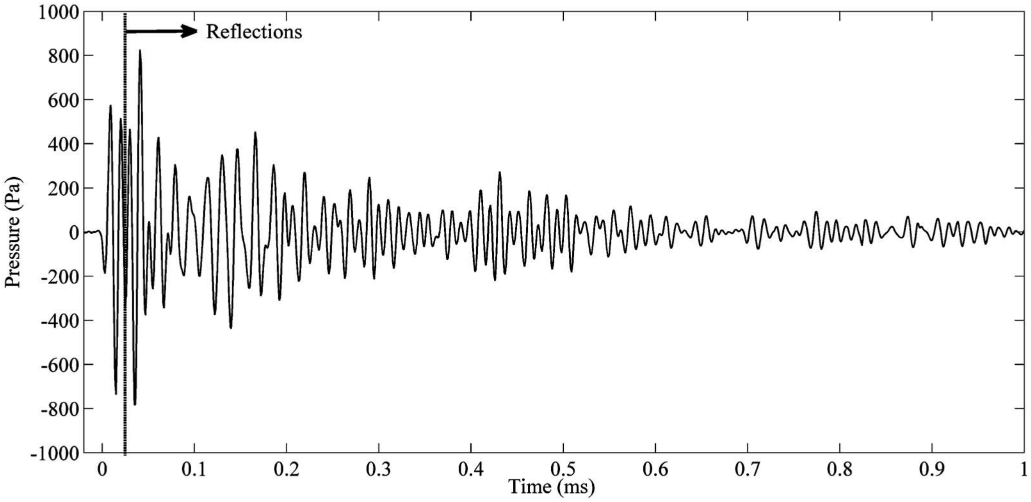


FIG. 5. Recorded signal (at hydrophone 6) from a sound source, ITC-1042, emitting a at a synthetic cavitation pulse with center frequency of 100 kHz. The approximate angle of incidence is 35° of the sound path to the acrylic plate. After the first 50 μs, once-reflected acoustic waves begin to arrive, and their aggregate can be stronger than the direct path.

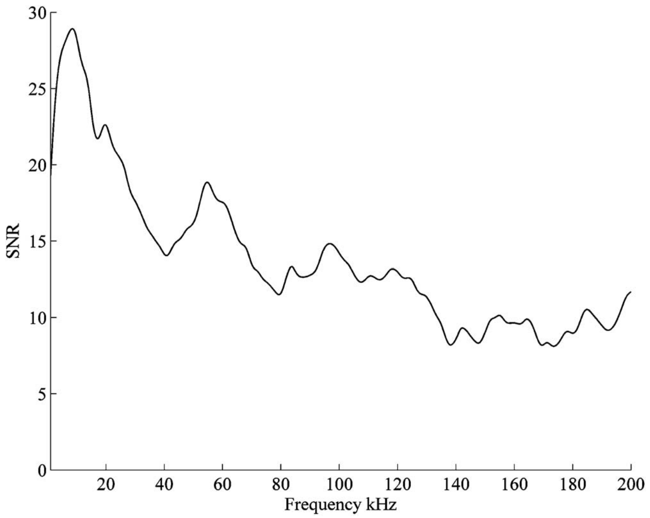


FIG. 6. SNR as a function of frequency of the cavitation events located.

the lone direct-path signal was estimated to be on average 6–12 dB lower than the peak fine-window SNR, depending on the coherence of the first three reflections. Here it was assumed that reality was closer to the incoherent limit, so the final fine-window location in each time series was set by determining the point prior to the maximum fine-window SNR having a SNR that was 6 dB lower. Similarly, for the synthetic cavitation pulses, the peak-SNR frequency was found, and the fine-window SNR_a was calculated within a frequency band that contained half the signal energy. Sample signals that illustrate fine-window placement are shown in Fig. 9.

The windowed direct-path signal may contain distortions due to the bubble motion in the 10 m/s flow. However, such distortions were ignored since the maximum possible time delay of 0.26 ms or the window size of 0.3 ms leads to only 3 mm of bubble motion, and the average bubble length in the downstream direction is 8 mm. If sound-source location uncertainty, from bubble motion and finite size effects, were included, then statistical inversion for the source location might be possible (see Tarantola, 2005; Dosso and Sotirin, 1999). However, for simplicity, this approach was not attempted in this study.

Once the time series from each hydrophone was fine

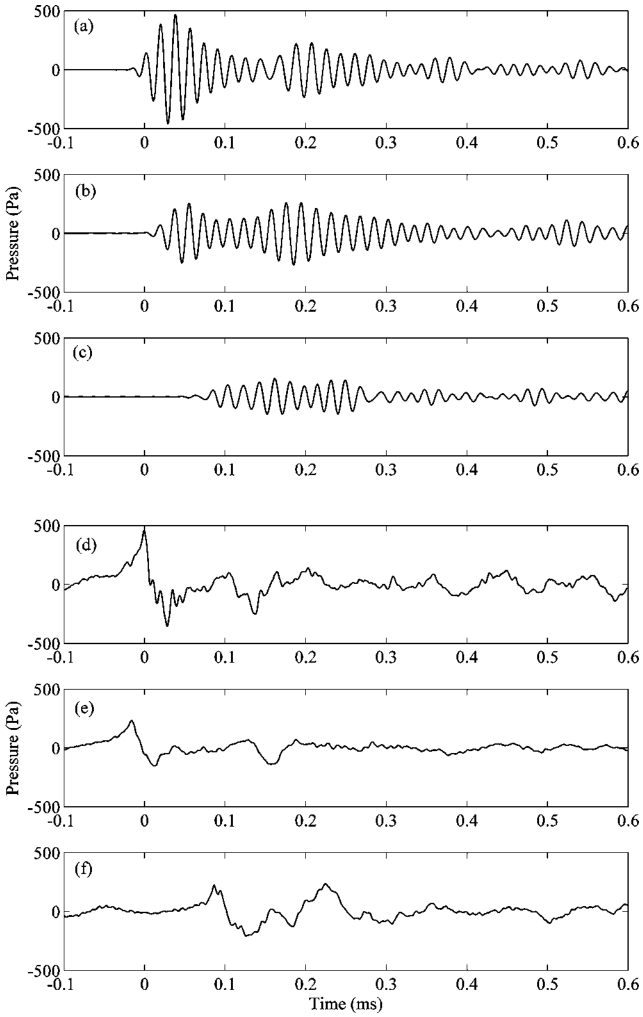


FIG. 8. Recorded signal waveforms for a synthetic cavitation pulse with center frequency of 50 kHz (a)–(c), and an actual cavitation pulse with center frequency of 6.7 kHz (d)–(f) for incidence angle near normal (a),(d), near the water-acrylic critical angle of ~35° (b),(e), and at 50°–60°, well above the water-acrylic critical angle (c),(f). For both pulse types, signal distortion in the first few tenths of a millisecond increases with increasing incidence angle. In particular, the loudest sound in (b), (c), and (f) is not near the leading edge of signal pulse as it is in (a), (d), and (e).

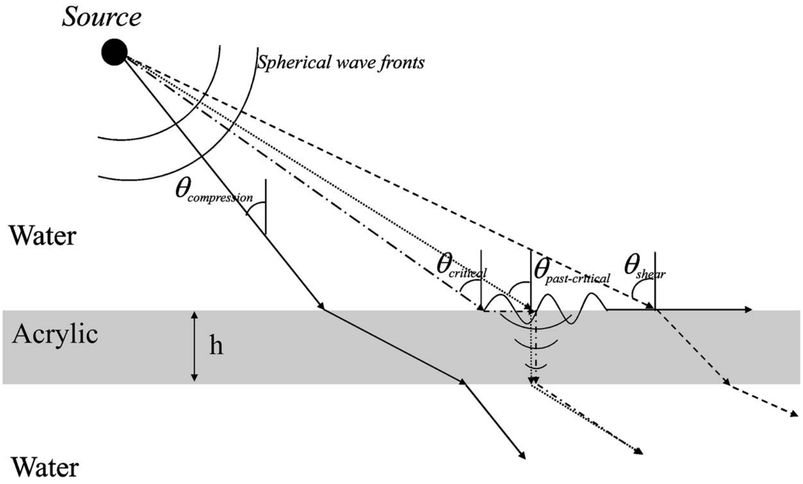


FIG. 7. Spherical sound waves that originate at a point in the flow may travel through a solid acrylic plate via a compression wave in the acrylic plate (solid lines), a shear wave in the acrylic plate (dashed lines), or as an evanescent wave in the acrylic plate (dash-dot lines) when the incidence angle exceeds the water-acrylic critical angle. Because of this complexity, the path that carries the strongest signal is geometry dependent, and significant signal distortion may occur, especially as skimming incidence is approached.

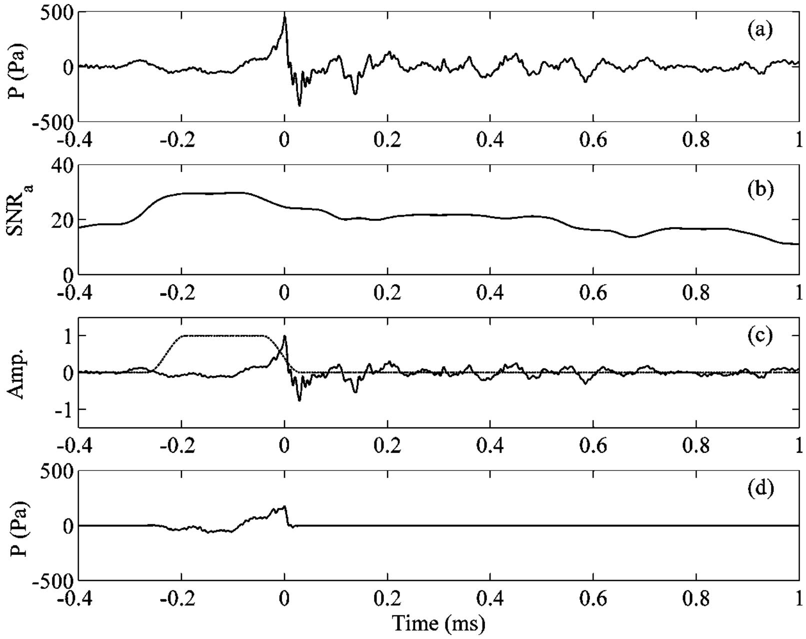


FIG. 9. Sample wave forms that illustrate the selection of the final time window for a single hydrophone recording: (a) coarsely-windowed raw signal, (b) time-dependent fine-window SNR of the signal shown in (a), (c) placement of the final time window superimposed on the signal shown in (a), and (d) the final fine-windowed signal used in cross-correlation processing. Here the window location is specified by the time at its left edge.

windowed to create m_i, the cross-correlation A_{ij} of the these fine-windowed time series was computed without normalization,

$$A_{ij}(l) = \sum_{n=0}^{N-l-1} m_i(t_{n+l})m_j(t_l) \quad \text{for} \quad l = 1, \ldots, 2N-1, \qquad (2)$$

and the measured time delays, τ_{ij}, were found from maxima of A_{ij}. Unfortunately, corrupt or spurious time delays were occasionally generated unintentionally by this signal-processing effort. To partially mitigate such problems, the various experimental time delays were compared to maximum possible values based on the geometry of the array and physically impossible time delays were eliminated from further consideration.

During the localization process of these cavitation events, the time of maximum fine-window SNR was compared to the final sound-source location and it was found that the time lag of the maximum SNR with respect to the direct-path fine window was of the expected duration based on the geometry of the test section and the hydrophone locations.

The system and method described here has limitations. If the direct and once-reflected arrival time differences are less than two periods of the signal's peak-SNR frequency, the localization technique may fail. The availability of sufficient direct-path signal will depend on the ratio of the peak-SNR wavelength, λ, of the source to the smallest characteristic dimension of the water tunnel test section, d (or the enclosure). If λ/d is approximately 1.5 then the direct path should be distinct enough. A smaller λ/d would imply that the direct path is corrupted by the first-reflected paths, while at larger λ/d the direct-path signal would be easily distinguishable from the reflected signal, and a coarse windowing technique or methods such as Spiesberger (1998) would suffice. In addition, the current approach is based on a combination of first reflections being loudest. Its performance is likely to decrease if later-arriving multiply-reflected paths are loudest.

B. Straight-ray Monte-Carlo processing

To be successful, the straight-ray propagation model used in the localization algorithm needs to mitigate some of complications inherent in the water tunnel environment: the reverberation, the acrylic plate between the water tunnel interior and the receivers, finite SNR, and errors in the locations of the receivers. In addition, there may be some Doppler differences between phones with the estimated largest error in the observed frequency being ±0.67% depending on whether the hydrophone was upstream or downstream from the source; however, no Doppler corrections were made to the recorded signals. The acrylic plate causes geometry-dependent refraction of acoustic waves but its effects on the ray paths and the effective speed of sound, c, were assumed to be the same for all phones during the signal processing. In addition, all sound waves were assumed to travel on straight paths. Thus, the Euclidean distance between the known receiver locations $\boldsymbol{R}_j = (R_{x,j}, R_{y,j}, R_{z,j})$ and the unknown source location $\boldsymbol{S} = (S_x, S_y, S_z)$ was presumed to depend on the measured time delays, $\tau_{ij} = t_i - t_j$, an unknown reference time of flight, t_j, from the source to reference receiver, j, and the unknown speed of sound, c. The overall source localization scheme is largely a synthesis of prior techniques with two mild extensions. For straight rays:

$$\|\boldsymbol{R}_i - \boldsymbol{S}\|^2 = (c(\tau_{ij} + t_j))^2. \qquad (3)$$

This equation was cast in an invertible via least-squares matrix form.

$$\mathbf{W}\mathbf{A}\boldsymbol{x} = \mathbf{W}\boldsymbol{B}, \qquad (4)$$

where matrix $\mathbf{A}$ contains all the known and experimentally measured information,

$$\mathbf{A}\boldsymbol{x} = \begin{pmatrix} (R_{x,i} - R_{x,j}) & (R_{y,i} - R_{y,j}) & (R_{z,i} - R_{z,j}) & \tau_{ij}^2 & \tau_{ij} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (R_{x,k} - R_{x,j}) & (R_{y,k} - R_{y,j}) & (R_{z,k} - R_{z,j}) & \tau_{kj}^2 & \tau_{kj} \end{pmatrix} \times \begin{pmatrix} S_x \\ S_y \\ S_z \\ c^2/2 \\ c^2 t_j \end{pmatrix}, \quad (5)$$

the column vector $\boldsymbol{x}$ includes all the unknowns, the column vector $\boldsymbol{B}$ is formed from the receiver locations

$$\boldsymbol{B} = \frac{1}{2}\begin{pmatrix} \|\boldsymbol{R}_i\|^2 - \|\boldsymbol{R}_j\|^2 \\ \vdots \\ \|\boldsymbol{R}_k\|^2 - \|\boldsymbol{R}_j\|^2 \end{pmatrix}, \quad (6)$$

and $\mathbf{W}$ is a diagonal weighting matrix. The derived least-squares matrix form is similar to that in Spiesberger and Fristrup (1990), which is extended here to include sound-speed estimation and a data weighting scheme. The weighting scheme (Strang, 1986) was introduced to at least partially address the distortions in the acoustic signal that are not modeled. Here, the ith diagonal element of $\mathbf{W}$ is set to zero if the measured time delays generated from hydrophone i are larger than the theoretical maximum based on the known geometry. Otherwise the ith diagonal element of $\mathbf{W}$ was set equal to $(m_i)_{\max}/(m_1, m_2, \ldots, m_{16})_{\max}$. The geometry of the hydrophone array in relation to the sound-source locations implies that the loudest recordings predominantly occur when the sound signal travels on paths having small angles of incidence with the acrylic plate. Therefore, the weighting scheme favors the loud recordings.

The current experimental setup has 16 hydrophones; consequently, 16 different sets of 15 model equations can be created by cycling through each receiver as the reference. The inputs for the acoustic propagation model (4) are the time delays and the locations of the receivers. The size of a typical time delay was found to be on the order of 30 μs based on the relative location of the average cavitation event and the hydrophones with an uncertainty due to the signal distortion that varies depending on the angle of incidence of the acoustic ray and the frequency content of the signal. The largest errors occur at large angles of incidence and are of the order of 300 μs. Based on the geometry of the tunnel and array, the maximum physically possible value is of 270 μs. Time delays larger than 270 μs were eliminated with the weighting scheme described above. Timing errors that are of the same (or smaller) size as the real time delays could not be eliminated.

Monte-Carlo calculations were used to account for receiver location uncertainty, similar to Spiesberger (1999b), though the data from all 16 receivers were used here. The nominal three-dimensional coordinate locations of the receivers were presumed to have Gaussian-distributed errors of 4 mm. The sets of 15 model equations were then solved 2000 times each with different perturbed receiver locations. The resulting sound-source locations and sound speeds were then paired down to those which were physically possible, i.e., locations within the volume of the test section with sound speed between 1480 and 1600 m/s. The precise value of the effective sound speed could not be estimated because the acrylic separating the source from the hydrophones has Young's modulus that can take on values from 2.2 to 5.5 GPa (Ngoepe *et al.*, 1990; Johnson and Jones, 1994), and varies with temperature and frequency (Mulliken and Boyce, 2006). Furthermore, the ratio of the acoustic path in the acrylic and water cannot be predetermined. The location solutions having sound speeds within ±20 m/s of the most common sound speed were averaged to determine the final sound-source location estimate.

IV. LOCALIZATION RESULTS

The localization system was initially tested with the synthetic sound source and the localization method performance was gauged for a variety of parametric changes involving: the Monte-Carlo calculations, deteriorating SNR, signal bandwidth, and the relative location of the geometric center of the array to the sound source.

When the number of Monte-Carlo trials was varied from 100 to 10 000, the localization error fell to within 5% of its asymptotic minima at 2000 calculations. Therefore, this number of trials was used in the remainder of this investigation. Furthermore, the hydrophone placement error was varied from 1 to 5 mm, and 4 mm minimized the average localization error.

To gauge the effect of signal frequency and noise, synthetic cavitation pulses were generated with an ITC-1042 in the water tunnel test section without flow. Here, the location of this sound projector was known. The peak-SNR frequency was changed from 10 to 180 kHz while a bandwidth of approximately of 20 kHz was maintained. The SNR was incrementally decreased from 30 to 5 dB by adding synthetic random noise to each time series. For each frequency and SNR, 70 trials were conducted with different noise realizations. The average localization error, L_{error}, divided by the peak-SNR wavelength, λ, is plotted in Fig. 10(b). As expected, L_{error}/λ generally increases with increasing noise. In addition, L_{error} increases with increasing frequency [Fig. 10(a)]. This frequency effect is presumed to be caused by signal-pulse distortion arising from propagation through the acrylic plate that separated the hydrophones from the test-section flow. This distortion increased with increasing angle of incidence, especially past the critical angle [see Figs. 8(c)–8(f)]. Consequently, when the direct-path sound reaches the array at large incidence angles, signal timing accuracy is degraded and this adversely impacts localization. In addition, at the lower SNR levels (10 dB or less) and the higher peak-SNR frequencies (40 kHz or more), the signal-processing algorithm fails to localize the sound projector 10%–100% of the time.

The acoustic localization system was tested with 53 cavitation events that were also located optically to ±4 mm. The results for L_{error} vs downstream distance z and L_{error}/λ vs dimensionless downstream distance, z/λ, in the test section are provided in Fig. 11. Here, the average length of the

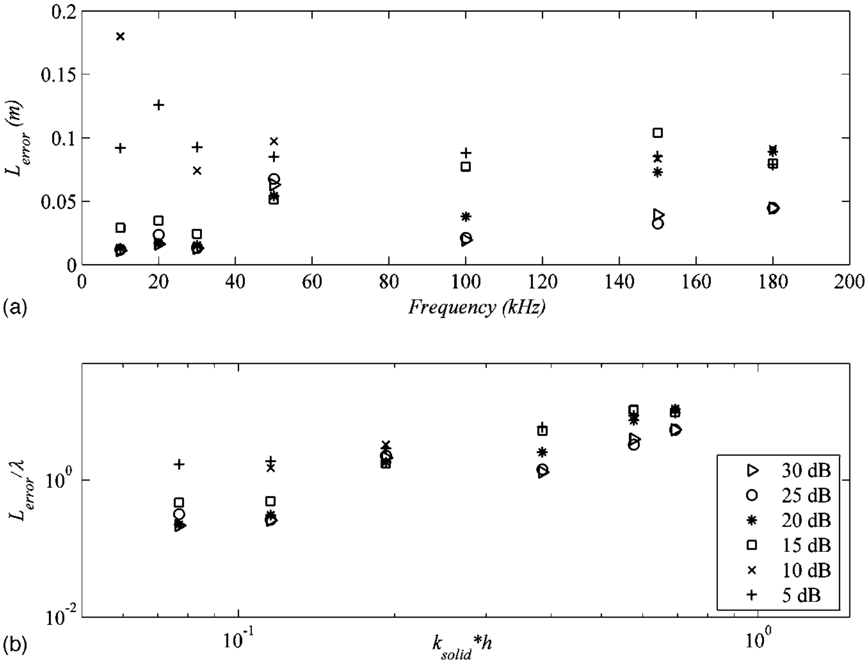


FIG. 10. (a) Localization error, $L_{\rm error}$, vs peak-SNR frequency. (b) Localization error, $L_{\rm error}$, divided by the peak-SNR wavelength in water, λ vs the product of the nominal compression-speed wavenumber, $k_{\rm solid}=2\pi f/(2600$ m/s), and the thickness, h, of the acrylic plate. Each symbol is the average of 70 different realizations of the added noise. The average 95% confidence interval (unplotted) for $L_{\rm error}/\lambda$ is 10%–20% of the plotted averages. As expected, localization error is generally lower at higher SNR, but it increases with increasing frequency.

bubbles was 8 mm, and the peak-SNR wavelength was λ =22.1 cm (6.7 kHz). Overall, the average distance between the optically- and acoustically-measured bubble locations, $L_{\rm error}$, was 19.4 mm.

The effect of the relative location of the geometric center of the array with respect to the sound-source location was gauged by processing the data collected from the 53 cavitation events with eight hydrophones selected to represent a different array configurations. The localization error was then determined in each of the different coordinate directions. If the hydrophones are distributed similarly to the full 16-element array, the error was found to be the same in the two cross stream directions, but 30% less in the z-direction compared to x and y. It was found that if all the hydrophones were placed in the y-z or x-z plane, the respective error in the x- or y-coordinate direction was approximately twice that of the other two directions. In the z-direction, the hydrophones were separated into upstream and downstream groups for processing purposes. The upstream group was centered at approximately the same z-location as the average of the cavitation events. The localization error in this case was similar in all three directions and to that from an 8-hydrophone array with a distribution similar to the 16-hydrophone case. The downstream array was centered approximately 15 cm downstream from the average event location. For this group, the error in the z-direction was an order of magnitude larger than in the other two directions.

The localization error was also studied with respect to the number, N, of hydrophones used. For each $N<16$, 16 different combinations of hydrophones where chosen to locate all 53 cavitation events. The minimum group size was 6, and the maximum was 16, for which there is only one group-

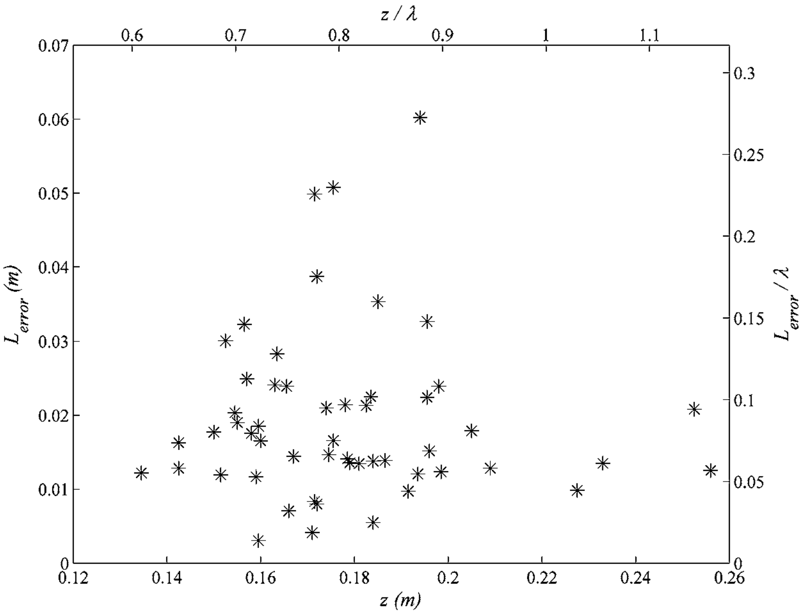


FIG. 11. The distance between acoustically and optically-determined bubble locations, $L_{\rm error}$, in meters vs downstream distance, z, in meters, for 53 cavitation bubbles. The top and right sides of the plot list values normalized by the peak-SNR wavelength in water, λ. Here, λ=22.1 cm (6.7 kHz), and average $L_{\rm error}$ was 19.4 mm or 0.087λ. The average length of the bubbles was 8 mm.

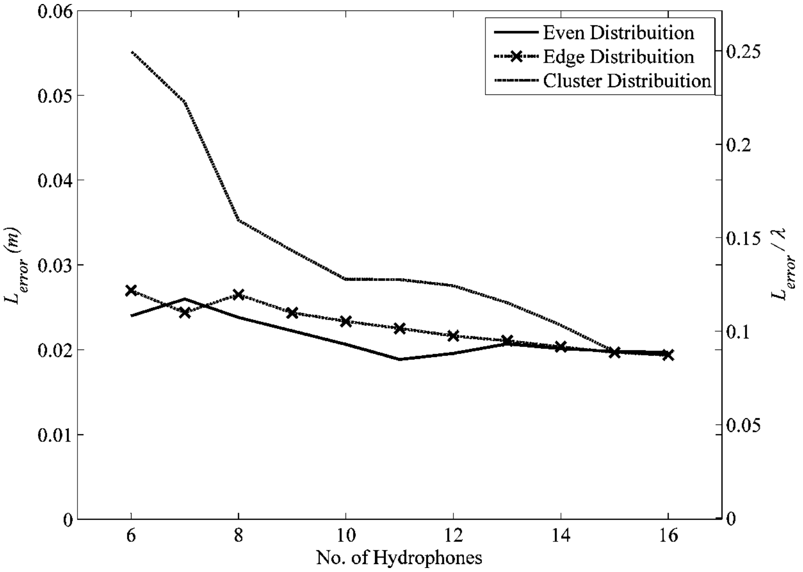


FIG. 12. The distance between acoustically and optically-determined bubble locations, L_{error}, in meters vs the number of hydrophones in grouped: evenly along (solid line), at the edges of (wide spaced dashes), and clustered in the middle (closely spaced dashes) of the full 16-element array. The right side shows L_{error} divided by the peak-SNR wavelength in water, λ.

ing. Here again, the hydrophones used were geometrically arranged into groups that were (i) evenly distributed over the full extent of the 16-element array, (ii) clustered near the center of the full 16-element array, or (iii) selected to emphasize the upstream and downstream edges of the full 16-element array. The criterion used for creating the evenly-distributed, clustered, and edge-distributed groups was based on the average three-dimensional location of the hydrophones, and their spatial-location standard deviation. In all three cases, the average three-dimensional location of the hydrophones of each subgroup was as close as possible as to that of the full 16-element array. The evenly-distributed, clustered, and edge-distributed groupings had standard deviations in the hydrophone location that were 1.0, 0.6, and 3.0 times that of the full 16-element array.

Localization and failure probability for the three types of hydrophone groupings are shown in Figs. 12 and 13 as a function of hydrophone number. As expected, the error decreases as N increases, but in the edge and evenly-distributed arrays the error does not change by more than 20% while for the clustered groupings the error drops by 150% from $N=6$ to $N=15$. This result is due in part to the low failure probability of the clustered groupings, as shown in Fig. 13. The minimum required number of hydrophones for all groupings was $N=6$, and even for this low number, the clustered groupings successfully produced a sound-source location estimate more than 90% of the time, while for the even- and edge-distributed arrays 8 and 9 phones were required for a similar success rate. The clustered groupings are more likely to robustly produce a less accurate answer than the other groupings. In all cases, the signal-processing algorithm is more likely to converge with larger arrays, and the current results suggest that at least nine hydrophones are necessary to achieve reasonable results. The low failure probability of the clustered groupings compared to the other two distributions again points out the waveform distortion caused by acoustic propagation through the acrylic as being a primary limitation on the acoustic localization performance of the current technique.

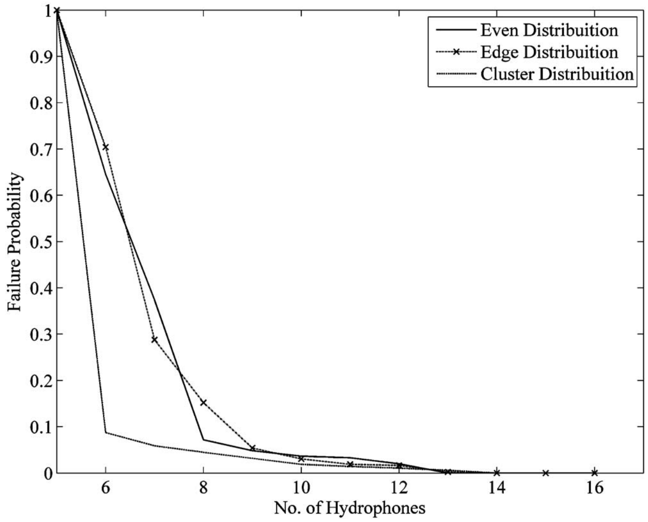


FIG. 13. Failure probability for the localization algorithm vs the number of hydrophones. The clustered groupings are robust in producing a sound-source location estimate than the other groupings. And, for the current localization algorithm, at least eight or nine hydrophones are needed.

V. SUMMARY AND CONCLUSIONS

A straight-ray-based sound-source localization routine for highly reverberant environments has been developed that at least partially compensates for parametric and geometric uncertainties, and even some propagation complexities via signal processing, data weighting, and Monte-Carlo trials. This current technique is a combination and mild extension of those described in Spiesberger and Fristrup (1990) and Spiesberger (1999b, 2005) with refinements for low-event-rate cavitation-bubble sounds in a water tunnel test section. The technique described here has only been validated for a low-event-rate cavitating flow where bubbles are typically present individually. Application of this technique to flows involving bubble clusters, clouds, or sheet cavitation has not (yet) been attempted. The technique's performance was compared to simultaneous optical localization measurements of 53 discreet (cavitation) sound sources having unknown timing and waveform in a domain extending over several peak-SNR acoustic wavelengths. In the present experiments, the sound sources were small cavitation bubbles that produced broadband pulses of less than 2 ms duration and longer tone bursts of 2–15 ms duration. When all 16 receiving hydrophones were used, all 53 bubbles—representing both signal types—were acoustically localized within an average of 2 cm of their optically-determined locations. However, subsampling the present array measurements suggests that as few as 8 or 9 receivers may be sufficient. Moreover, the acoustic technique described here should be applicable in many circumstances where optical techniques are not.

This study leads to the following conclusions. Three-dimensional localization of sound sources in water tunnel test section is possible when direct-path sound can be isolated from hydrophone array recordings. As expected, localization robustness and accuracy improve with increasing SNR and increasing numbers of receivers, but the accuracy does not increase with increasing signal frequency between 10 and 180 kHz. This performance limitation is likely the result of the straight-ray propagation model used in the localization routine. However, straight-ray simplicity should still be an overall advantage for this localization technique since it should allow it to be readily and successfully implemented in other test facilities. The extensive facility characterization necessary for other acoustic localization techniques can be bypassed with the current approach. Yet, improvements to the localization accuracy of the current technique are likely possible if features of the actual acoustic environment can be incorporated into the propagation model. Specifically in the present experimental study, the acoustic waveform distortion caused by the acrylic plate that separates the receivers from the test-section flow appears to be the primary limitation on localization accuracy. This limitation can be overcome in part by using hydrophone recordings that correspond to the lowest-possible direct-path water-acrylic incidence angles.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-07-1-0471, Dr. Ki-Han Kim program monitor. The authors would also like to acknowledge Dr. Steven L. Ceccio, Dr. Jaehyug Choi, and Mr. Ryo Yakushiji for their assistance.

Aarabi, P. (**2002**). "Self-localizing dynamic microphone arrays," IEEE Trans. Syst. Man Cybern. **32**, 474.

Arndt, R. E. A. (**2002**). "Cavitation in vortical flows," Annu. Rev. Fluid Mech. **34**, 143–175.

Arnold, D. P., Nishida, T., Cattafesta, L. N., and Sheplak, M. (**2003**). "A directional acoustic array using silicon micromachined piezoresistive microphones," J. Acoust. Soc. Am. **113**, 289–298.

Blake, W. K. (**1986a**) *Mechanics of Flow-Induced Sound and Vibration I* (Academic, New York), pp. 403–407.

Blake, W. K. (**1986b**) *Mechanics of Flow-Induced Sound and Vibration II* (Academic, New York), pp. 428–491.

Brennen, C. E. (**1995**) *Cavitation and Bubble Dynamics* (Oxford University Press, New York), pp. 65–91.

Caiti, A., Garulli, A., Livide, F., and Prattichizzo, D. (**2005**). "Localization of autonomous underwater vehicles by floating acoustic buoys: A set-membership approach," IEEE J. Ocean. Eng. **30**, 140–152.

Chang, N. A. (**2007**). "Acoustic characterization of cavitation in reverberant environments," Ph.D. thesis, University of Michigan, Ann Arbor, MI.

Chesnakas, C. and Jessup, S. (**2003**). "Tip vortex induced cavitation on a ducted propulsor," in Proceedings of the Fourth ASME-JSME Joint Fluids Engineering Conference, Honolulu, HI, Paper No. FEDSM2003-45320.

Choi, J., Hsiao, C.-T., Chahine, G., and Ceccio, S. L. (**2009**). "Growth, oscillation, and collapse of vortex cavitation bubbles," J. Fluid Mech. **624**, 255–279.

Collins, M. D. and Kuperman, W. A. (**1991**). "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **90**, 1410–1422.

Dobrzynski, W., Gehlhar, B., and Buchholz, H. (**2001**). "Model and full scale high-lift wing wind tunnel experiments dedicated to airframe noise reduction," Aerosp. Sci. Technol. **5**, 27–33.

Dosso, S. E., and Sotirin, B. J. (**1999**). "Optimal array element localization," J. Acoust. Soc. Am. **106**, 3445–3459.

Etter, R. J., Cutbirth, J. M., Ceccio, S. L., Dowling, D. R., and Perlin, M. (**2005**). "High Reynolds number experimentation in the US Navy's William B Morgan large cavitation channel," Meas. Sci. Technol. **16**, 1701–1709.

Ferguson, B. G. (**1999**). "Time-delay estimation techniques applied to the acoustic detection of jet aircraft transits," J. Acoust. Soc. Am. **106**, 255.

Gerard, A., Berry, A., and Masson, P. (**2005**). "Control of tonal noise from subsonic axial fan. Part 1: Reconstruction of aeroacoustic sources from far-field sound pressure," J. Sound Vib. **288**, 1049–1075.

Grosche, F.-R., and Meier, G. E. A. (**2001**). "Research at DLR Gottingen on bluff body aerodynamics, drag reduction by wake ventilation and active flow control," J. Wind. Eng. Ind. Aerodyn. **89**, 1201–1218.

Handzel, A. A., and Krishnaprasad, P. S. (**2002**). "Biomimetic sound-source localization," IEEE Sens. J. **2**, 607.

Iyer, C. O., and Ceccio, S. L. (**2002**). "The influence of developed cavitation on the flow of a turbulent shear layer," Phys. Fluids **14**, 3414–3431.

Jensen, F. B., Kuperman, W. A., Portor, M. B., and Schmidt, H. (**1994**) *Computational Ocean Acoustics* (American Institute of Physics, New York).

Jiménez, J. A., Mazo, M., Ureña, J., Hernández, A., Álvarez, F., García, J. J., and Santiso, E. (**2005**). "Using PCA in time-of-flight vectors for reflector recognition and 3-D localization," IEEE Trans. Robot. Autom. **21**, 909.

Johnson, J. A., and Jones, D. W. (**1994**). "The mechanical properties of PMMA and copolymers with ethyl methacrylate and butyl methacrylate," J. Mater. Sci. **29**, 870–876.

Katz, J., and O'Hern, T. J. (**1986**). "Cavitation in large scale shear flow," ASME Trans. J. Fluids Eng. **108**, 373–376.

Mulliken, A. D., and Boyce, M. C. (**2006**). "Mechanics of the rate-dependent elastic–plastic deformation of glassy polymers from low to high strain rates," Int. J. Solids Struct. **43**, 1331–1356.

Nagakura, K. (**2006**). "Localization of aerodynamic noise sources of Shinkansen trains," J. Sound Vib. **293**, 547–556.

Ngoepe, P. E., Lambson, E. F., and Saunders, G. A. (**1990**). "The elastic behaviour under hydrostatic pressure of poly(methyl methacrylate) and its fully deuterated form," J. Mater. Sci. **25**, 4654–4657.

O'Hern, T. J. (**1990**). "An experimental investigation of turbulent shear flow cavitation," J. Fluid Mech. **215**, 365–391.

Oweis, G. F., Choi, J., and Ceccio, S. L. (**2004**). "Dynamics and noise emission of laser induced cavitation bubbles in a vortical flow field," J. Acoust. Soc. Am. **115**, 1049–1058.

Oweis, G. F., Fry, D., Chesnakas, C. J., Jessup, S. D., and Ceccio, S. L. (**2006a**). "Development of a tip-leakage flow—Part 1: The flow over a range of Reynolds numbers," ASME J. Fluids Eng. **128**, 751–765.

Oweis, G. F., Fry, D., Chesnakas, C. J., Jessup, S. D., and Ceccio, S. L. (**2006b**). "Development of a tip-leakage flow—Part 2: Comparison between the ducted and un-ducted rotor," ASME J. Fluids Eng. **128**, 765–773.

Perkins, J. S., and Kuperman, W. A. (**1990**). "Environmental signal processing: Three-dimensional matched-field processing with a vertical array," J. Acoust. Soc. Am. **87**, 1553–1556.

Schmidt, R. O. (**1972**). "A new approach to geometry of range difference location," IEEE Trans. Aerosp. Electron. Syst. **AES-8**, 821–835.

Skarsoulis, E. K. (**2005**). "Ray-theoretic localization of an impulsive source in a stratified ocean using two hydrophones," J. Acoust. Soc. Am. **118**, 2934–2943.

Spiesberger, J. L. (**1998**). "Linking auto- and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipaths," J. Acoust. Soc. Am. **104**, 300–312.

Spiesberger, J. L. (**1999a**). "Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration," J. Acoust. Soc. Am. **106**, 837–847.

Spiesberger, J. L. (**1999b**). "Probability density functions for hyperbolic and isodiachronic locations," J. Acoust. Soc. Am. **106**, 837–847.

Spiesberger, J. L. (**2001**). "Hyperbolic location errors due to insufficient numbers of receivers," J. Acoust. Soc. Am. **109**, 3076–3079.

Spiesberger, J. L. (**2004**). "Geometry of locating sounds from differences in travel time: Isodiachrons," J. Acoust. Soc. Am. **116**, 3168–3177.

Spiesberger, J. L. (**2005**). "Probability distributions for locations of calling animals, receivers, sound speeds, winds, and data from travel time differences," J. Acoust. Soc. Am. **118**, 1790–1801.

Spiesberger, J. L., and Fristrup, K. M. (**1990**). "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," Am. Nat. **135**, 107–153.

Steinberg, B. D. (**1976**). *Principles of Aperture and Array System Design* (Wiley, New York), Chap. 1.

Strang, G. (**1986**). *Introduction to Applied Mathematics* (Wellesley-Cambridge, Wellesley, MA).

Tarantola, A. (**2005**). *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, Philadelphia).

Voltz, P., and Lu, I. T. (**1994**). "A time-domain backpropagating ray technique for source localization," J. Acoust. Soc. Am. **95**, 805–812.

Wang, Y., Lia, J., Stoica, P., Sheplak, M., and Nishida, T. (**2004**). "Wideband RELAX and wideband CLEAN for aeroacoustic imaging," J. Acoust. Soc. Am. **115**, 757–767.

Westwood, E. K., and Knobles, D. P. (**1997**). "Source track localization via multipath correlation matching," J. Acoust. Soc. Am. **102**, 2645–2654.

Ziomek, L. J. (**1995**). *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing* (CRC, Ann Arbor, MI), Chaps. 6 and 7.

Scattering calculation and image reconstruction using elevation-focused beams

David P. Duncan and Jeffrey P. Astheimer
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

Robert C. Waag
Department of Electrical and Computer Engineering and Department of Imaging Sciences, University of Rochester, Rochester, New York 14627

(Received 13 June 2008; revised 17 February 2009; accepted 17 February 2009)

Pressure scattered by cylindrical and spherical objects with elevation-focused illumination and reception has been analytically calculated, and corresponding cross sections have been reconstructed with a two-dimensional algorithm. Elevation focusing was used to elucidate constraints on quantitative imaging of three-dimensional objects with two-dimensional algorithms. Focused illumination and reception are represented by angular spectra of plane waves that were efficiently computed using a Fourier interpolation method to maintain the same angles for all temporal frequencies. Reconstructions were formed using an eigenfunction method with multiple frequencies, phase compensation, and iteration. The results show that the scattered pressure reduces to a two-dimensional expression, and two-dimensional algorithms are applicable when the region of a three-dimensional object within an elevation-focused beam is approximately constant in elevation. The results also show that energy scattered out of the reception aperture by objects contained within the focused beam can result in the reconstructed values of attenuation slope being greater than true values at the boundary of the object. Reconstructed sound speed images, however, appear to be relatively unaffected by the loss in scattered energy. The broad conclusion that can be drawn from these results is that two-dimensional reconstructions require compensation to account for uncaptured three-dimensional scattering.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097497]

PACS number(s): 43.60.Pt, 43.20.Fn, 43.35.Wa, 43.80.Qf [TDM] Pages: 3101–3119

I. INTRODUCTION

Ultrasonic *b*-scan imaging is a valuable tool for the detection and diagnosis of disease. Applications include imaging the heart, abdominal organs, and the developing fetus.[1,2] Ultrasound has the advantages of being non-ionizing at intensities used for diagnostic imaging, showing motion in real time, and being inexpensive in comparison to other imaging modalities.

While *b*-scan imaging based on transmitting pulses and receiving echoes with focused beams is widespread, imaging based on inverse scattering is relatively undeveloped in medical applications. This lack of development can be attributed to a number of reasons. The most important of these is that a constructive solution of the inverse problem in two dimensions does not yet exist and current solutions now require *a priori* knowledge or iteration or both. Other important reasons are the complex apparatus to acquire data and the substantial computation needed to reconstruct images.

In x-ray tomography, an uncomplicated straight-line model facilitates data collection and fast reconstruction of images. In magnetic resonance imaging, a Fourier transform relation between the measured data and the image facilitates fast image reconstruction. In ultrasound tomography, however, acoustic scattering is a three-dimensional process that complicates data collection and image reconstruction. Nevertheless, if two-dimensional data collection and algorithms are applicable, then imaging complexity is reduced from that required for the three-dimensional case.

Cross sections reconstructed using a two-dimensional algorithm can be influenced by scattering into the third dimension. This study examines the effects that three-dimensional scattering has on these two-dimensional reconstructions. The scattering objects used in this study are cylinders and spheres. However, the methods developed to examine these objects are also applicable to more complicated objects.

Analytic expressions for plane-wave scattering by a cylinder and by a sphere are extended to include elevation-focused illumination and reception using an angular spectrum representation of the focused beams. The extended expressions are employed to calculate scattering at multiple frequencies. The scattering data are used to reconstruct elevation-offset cross sections of a sphere. These reconstructions show variations in image accuracy that result from elevation focusing and scattering object variability in the width of the focus.

The choice of parameters in this study is motivated by the ring transducer system at the University of Rochester.[3] The transducer in this system is a 150-mm diameter ring of 2048 equally spaced transducer elements each with an elevation of 25 mm. The elements are focused in elevation with an acoustic lens that creates a beam with an elevation *f*-number of 5. The focus of the beam is in the imaging plane at the center of the ring. At the focus, the beam has a half-amplitude width of 4 mm in the elevation dimension and a 100-mm imaging-plane depth of field essentially centered around the focus. The transmit subsystem has 128 channels each with an individually programmable wave form. The receive subsystem has 16 channels each sampled at 20 MHz with 12 bits of resolution. Transmit and receive multiplexers are used to connect the transmit and receive channels to the transducer elements.

Ultrasound image reconstruction at a given temporal frequency is equivalent to finding a solution for the acoustic scattering potential in the Lippman–Schwinger equation given measurements of the incident pressure and the total pressure. Available reconstruction algorithms can be grouped in three classes. One class utilizes so-called time-of-flight approximations, another class finds a linearized solution (i.e., uses a Born or Rytov approximation), and the other class is based on iterative, nonlinear, full-wave methods to obtain a solution. Algorithms representative of the first type are those in Refs. 4 and 5. Algorithms representative of the second type are filtered backpropagation[6] and Fourier-domain interpolation.[7] Algorithms representative of the third type are nonlinear algebraic reconstruction techniques such as the Born iterative methods,[8,9] the sinc-basis moment method,[10,11] and a method using eigenfunctions of a scattering operator.[12]

The eigenfunction method is used in this paper to reconstruct images. The eigenfunction method was originally developed for single-frequency illumination,[12] later extended to use all the frequencies in an incident pulse and include adaptive demodulation that extends the validity of the Born approximation,[13] and recently further extended to use iteration for imaging large-scale high-contrast objects.[14] Although the eigenfunction method is also applicable in three dimensions, the ring transducer system available for measurements acquires data in two dimensions and, as noted, motivates the choice of parameters in this study.

The remainder of this paper is comprised of the following sections. Section II develops the theoretical basis for the scattering calculations. Section III describes the computational methods. Then, results of calculations are presented in Sec. IV for various elevation-focused scattering geometries and scattering objects. Section V discusses the results and implications of using elevation-focused beams and also discusses key assumptions and other steps used in the analysis and computations. Conclusions are presented in Sec. VI.

II. THEORY

The application of a two-dimensional reconstruction algorithm to scattering in three dimensions requires that the scattering in three dimensions be related to the scattering in two dimensions. For a specific temporal frequency f with time dependence $e^{-j2\pi ft}$, this is accomplished by expressing the three-dimensional scattered pressure p_s in a Born series as

$$p_s(\mathbf{x},z,f) = p_0(\mathbf{x},z,f) + p_1(\mathbf{x},z,f) + p_2(\mathbf{x},z,f) + \cdots, \qquad (1)$$

where $\mathbf{x}$ has Cartesian components (x,y) in the x-y plane and z is the coordinate in the orthogonal (elevation) direction. Each term in this series is determined recursively using the formula

$$p_{n+1}(\mathbf{x},z,f) = k^2 \int_{z'} \int_{\mathbf{x}'} G_0^{(3\mathrm{D})}(\mathbf{x}-\mathbf{x}', z-z', f)\, \eta(\mathbf{x}',z') \times p_n(\mathbf{x}',z',f)\, d\mathbf{x}'\, dz', \qquad (2)$$

where $\eta(\mathbf{x},z)$ is the three-dimensional variation in the medium and $G_0^{(3\mathrm{D})}(\mathbf{x}-\mathbf{x}', z-z', f)$ is the three-dimensional free-space out-going Green's function given by

$$G_0^{(3\mathrm{D})}(\mathbf{x}-\mathbf{x}', z-z', f) = \frac{e^{j2\pi f\sqrt{\|\mathbf{x}-\mathbf{x}'\|^2+(z-z')^2}/c}}{4\pi\sqrt{\|\mathbf{x}-\mathbf{x}'\|^2+(z-z')^2}}, \qquad (3)$$

in which c is the speed of sound in the background medium.

If the medium variations are essentially independent of elevation, i.e., $\eta(\mathbf{x},z) \approx \eta(\mathbf{x})$, then Eq. (2) becomes

$$p_{n+1}(\mathbf{x},z,f) = k^2 \int_{\mathbf{x}'} \eta(\mathbf{x}') \int_{z'} \frac{e^{j2\pi f\sqrt{\|\mathbf{x}-\mathbf{x}'\|^2+(z-z')^2}/c}}{4\pi\sqrt{\|\mathbf{x}-\mathbf{x}'\|^2+(z-z')^2}} \times p_n(\mathbf{x}',z',f)\, dz'\, d\mathbf{x}'. \qquad (4)$$

Taking the Fourier transform with respect to the elevation coordinate z converts the convolution in the elevation coordinate into a product in the corresponding spatial-frequency coordinate ν. The result, as shown in the Appendix, is that Eq. (4) can be written[a] as

$$p_{n+1}(\mathbf{x},\nu,f) = k^2 \int_{\mathbf{x}'} G_0^{(2\mathrm{D})}(\mathbf{x}-\mathbf{x}', f_e)\, \eta(\mathbf{x}')\, p_n(\mathbf{x}',\nu,f)\, d\mathbf{x}', \qquad (5)$$

where $G_0^{(2\mathrm{D})}(\mathbf{x}-\mathbf{x}', f_e)$ is the two-dimensional free-space out-going Green's function as a function of the effective temporal frequency $f_e = \sqrt{f^2 - c^2\nu^2}$ and is given by

$$G_0^{(2\mathrm{D})}(\mathbf{x}-\mathbf{x}', f_e) = (j/4) H_0^{(1)}(\|\mathbf{x}-\mathbf{x}'\| 2\pi\sqrt{(f/c)^2 - \nu^2}). \qquad (6)$$

$2\pi\nu$ is the wavenumber in the z coordinate, $2\pi\sqrt{(f/c)^2-\nu^2}$ is the wavenumber in the x-y plane, and $p_n(\mathbf{x}',\nu,f)$ is a representation of the temporal-harmonic pressure at the nth iteration in terms of the spatial variable $\mathbf{x}'$ in the x-y plane and the spatial-frequency variable ν in the elevation dimension.

The foregoing analysis shows that scattering in three dimensions can be expressed as an uncoupled set of relations in two dimensions when the medium variations are independent of elevation. The scattering in three dimensions is then the sum of the two-dimensional scattering that results from each plane wave or spatial-frequency in the elevation dimension. The explicit relation is

$$p_n(\mathbf{x},z,f) = \int_{\nu} p_n(\mathbf{x},\nu,f)\, e^{j2\pi\nu z}\, d\nu. \qquad (7)$$

Since the spatial-frequency ν in the elevation dimension is given as a function of the elevation angle ϑ by

$$\nu = \frac{f}{c}\cos\vartheta, \qquad (8)$$

Eq. (7) can be expressed as

$$p_n(\mathbf{x},z,f) = \int_{\vartheta} p_n(\mathbf{x},(f\cos\vartheta/c),f)e^{j(2\pi f/c)z\cos\vartheta} \times d(f\cos\vartheta/c). \tag{9}$$

When $n=0$, the left side of this equation is the beam of illumination that is denoted as ψ_t in the next section and the integrand on the right side is the angular spectrum of plane waves that comprise the beam and is denoted as ϕ_t in that section. The angular spectrum form for the illumination pattern and also a corresponding form for reception sensitivity lead to convenient expressions for measured values of scattered pressure when the illumination pattern and reception sensitivity are focused in elevation.

Although the foregoing expressions and expressions in subsequent sections are for a single temporal frequency, each expression can be extended to the time domain by considering the temporal frequency to be a variable, weighting the frequency-domain expression with the spectrum of a temporal pulse such as the Gaussian-shaped bandpass pulse given by

$$g(l) = A_t e^{-t^2/2\sigma_t^2}\sin(2\pi f_0 t), \tag{10}$$

in which A_t is a constant amplitude scale factor, σ_t is a pulse-width parameter, and f_0 is the pulse center frequency, and then calculating the corresponding time-domain expression via an inverse Fourier transform of the weighted frequency-domain expression.

A. Focusing in elevation

The strength of acoustic illumination at a spatial location $(\mathbf{x},z)$ is described by a beam that is focused in elevation and has the form of a plane wave in the x-y coordinates and a direction determined by the azimuthal angle ϕ_i. For monochromatic illumination with time dependence $e^{-j2\pi ft}$, this beam is defined by an angular spectrum of plane waves in the elevation coordinate by using the relation

$$\psi_t(\mathbf{x},z) = \int_{\vartheta_i} \varphi_t(\vartheta_i)e^{j(2\pi f/c)\mathbf{u}(\phi_i,\vartheta_i)\cdot(\mathbf{x},z)}d\vartheta_i, \tag{11}$$

where $\varphi_t(\vartheta_i)$ is the illumination angular spectrum as a function of the varying polar angle ϑ_i and $\mathbf{u}(\phi_t,\vartheta_i)$ is a unit vector in the plane-wave direction specified by a fixed azimuthal angle ϕ_i and the varying polar angle ϑ_i. The vector $\mathbf{u}(\phi_t,\vartheta_i)$ has an elevation-dimension projection of $\cos\vartheta_i$ and an x-y plane projection of $\sin\vartheta_i\mathbf{u}_{xy}(\phi_i)$, where $\mathbf{u}_{xy}(\phi_i)$ is a unit vector with angle ϕ_i in the x-y plane.

Although both the beam pattern ψ_t and the angular spectrum φ_t depend on frequency, the frequency in this section is assumed fixed so the frequency dependence of ψ_t and φ_t is left implicit to simplify notation.

The angular spectrum $\varphi_t(\vartheta_i)$ of the illumination is determined in the x-y plane by the excitation of the ring transducer elements and in the elevation dimension by an acoustic lens that is the same for all elements in the ring. The excitation of the elements that are spaced evenly around the circumference of the ring $\{\mathbf{x}:\|\mathbf{x}\|=r_0\}$ is represented by a function $E(\mathbf{x},\phi_i)$ for $\|\mathbf{x}\|=r_0$, and the variation in the elevation dimension is represented by a function $\lambda(z)$ so that the field on the cylinder $\|\mathbf{x}\|=r_0$ at the ring radius r_0 is given by the product $\lambda(z)E(\mathbf{x},\phi_i)$. Equating this field with Eq. (11) for points $(\mathbf{x},z)$ on a portion of the cylinder $\|\mathbf{x}\|=r_0$ gives

$$\lambda(z)E(\mathbf{x},\phi_i) = \int_{\vartheta_i} \varphi_t(\vartheta_i)e^{j(2\pi f/c)\mathbf{u}(\phi_i,\vartheta_i)\cdot(\mathbf{x},z)}d\vartheta_i. \tag{12}$$

To solve for φ_t the first step is to take the Fourier transform of Eq. (12) with respect to the z coordinate to obtain

$$\lambda(\nu)E(\mathbf{x},\phi_i) = \int_{\vartheta_i} \varphi_t(\vartheta_i)\left[\int_z e^{j(2\pi f/c)\mathbf{u}(\phi_i,\vartheta_i)\cdot(\mathbf{x},z)}e^{-j2\pi\nu z}dz\right]d\vartheta_i. \tag{13}$$

Next, writing the spatial-frequency variable ν of the Fourier transform using Eq. (8) and expressing $\mathbf{u}(\phi_i,\vartheta_i)$ in $(\mathbf{x},z)$ components as described after Eq. (11) gives

$$\lambda(f\cos\vartheta_\nu/c)E(\mathbf{x},\phi_i) = \int_{\vartheta_i} \varphi_t(\vartheta_i)e^{j(2\pi f/c)\sin\vartheta_i\mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}} \times\left[\int_z e^{j(2\pi f/c)(\cos\vartheta_i-\cos\vartheta_\nu)z}dz\right]d\vartheta_i. \tag{14}$$

Evaluation of the inner integral and the use of a small angle approximation in the exponential yields

$$\lambda(f\cos\vartheta_\nu/c)E(\mathbf{x},\phi_i) = \frac{\varphi_t(\vartheta_\nu)}{f\sin\vartheta_\nu/c}e^{j(2\pi f/c)\mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}}. \tag{15}$$

This implies that

$$\varphi_t(\vartheta_\nu) = (f\sin\vartheta_\nu/c)\lambda(f\cos\vartheta_\nu/c) \tag{16}$$

and that

$$E(\mathbf{x},\phi_i) = e^{j(2\pi f/c)\mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}}. \tag{17}$$

For calculations of multiple scattering in this paper, the lens factor

$$\lambda(z) = a(z)e^{-j(2\pi f/c)\ell(z)} \tag{18}$$

is used. In this expression, $a(z)$ is a Gaussian-weighted amplitude term,

$$a(z) = A_G e^{-z^2/2\sigma_a^2}, \tag{19}$$

and $\ell(z)$ is a length factor given by

$$\ell(z) = \sqrt{(r_0+r_a)^2+z^2} \tag{20}$$

for a geometric focus at $(r_a,0,0)$, where r_a is an adjustment chosen to produce a focal peak at (0, 0, 0). The width parameter σ_a in Eq. (19) is real and produces a real-valued weight of the aperture, while the length factor ℓ produces a phase shift in the aperture.

A relation analogous to that between the illumination pattern and the angular spectrum of the illumination given in Eq. (11) can be defined between the reception sensitivity and the angular spectrum of the reception sensitivity. The relation is

$$\psi_r(\mathbf{x},z) = \int_{\vartheta} \varphi_r(\vartheta) e^{j(2\pi f/c)\mathbf{u}(\phi,\vartheta)\cdot(\mathbf{x},z)} d\vartheta, \tag{21}$$

where $\varphi_r(\vartheta)$ is the reception sensitivity angular spectrum expressed as a function of the varying polar angle ϑ and $\mathbf{u}(\phi,\vartheta)$ is a unit vector in the plane-wave direction specified by a fixed azimuthal angle ϕ and the varying polar receive angle ϑ. The vector $\mathbf{u}(\phi,\vartheta)$ has an elevation-dimension projection of $\cos\vartheta$ and an x-y plane projection of $\sin\vartheta\, \mathbf{u}_{xy}(\phi)$, where $\mathbf{u}_{xy}(\phi)$ is a unit vector with angle ϕ in the x-y plane.

Expressions (11) and (21) are special forms for illumination patterns and receiver sensitivities. These forms consist of plane-wave components that all advance in the same horizontal direction, i.e., have the same azimuthal angle. Illumination patterns given by Eq. (11) are produced by exciting ring transducer elements with amplitudes specified in Eq. (17). Receiver sensitivities given by Eq. (21) require that the temporal-frequency components of the signals from transducer elements be accumulated with weights also given by Eq. (17). These patterns are well suited for scattering measurements because each scattering measurement can be written as a linear combination of far-field values determined by a plane-wave component from the illumination pattern together with a plane-wave component from the receiver sensitivity. This is an efficient representation of scattering measurements when illumination patterns and receiver sensitivities have few plane-wave components.

More general expressions for illumination patterns and receiver sensitivities can also be formed. This is accomplished by applying the Rayleigh–Sommerfeld integral to field values on a portion of the cylindrical surface specified by $\|\mathbf{x}\| = r_0$ for excitations other than those given by Eq. (17). The result is

$$\psi_t(\mathbf{x},z) = \int_{\zeta}\int_{z'} E(\mathbf{x}'(\zeta))\lambda(z')\frac{\partial}{\partial r_{xx'}} \times G_0^{(3\mathrm{D})}(\mathbf{x}-\mathbf{x}'(\zeta), z-z', f)\,dz'\,d\zeta, \tag{22}$$

where $r_{xx'} = \|\mathbf{x}-\mathbf{x}'\|$, $\|\mathbf{x}'(\zeta)\| = r_0$, and ζ is the azimuthal angle over which the integration is performed in the x-y plane. After a change in variables and simplification, Eq. (22) becomes

$$\psi_t(\mathbf{x},z) = \int_{\vartheta_i} (f\sin\vartheta_i/c)\lambda(f\sin\vartheta_i/c)e^{j(2\pi f/c)\cos\vartheta_i z} \times \left\{ -2\int_{\|\mathbf{x}'\|=r_0} E(\mathbf{x}'(\zeta))\frac{\partial}{\partial r_{xx'}} \times [G_0^{(2\mathrm{D})}(\mathbf{x}-\mathbf{x}'(\zeta), f\sin\vartheta_i)]\,d\zeta \right\} d\vartheta_i, \tag{23}$$

where the term in curly brackets may be identified as the two-dimensional Rayleigh–Sommerfeld integral for propagation of the ring element excitations $E(\mathbf{x}')$ at the temporal frequency $f\sin\vartheta_i$. If $E(\mathbf{x}')$ is chosen to have the plane-wave form $e^{j(2\pi f/c)\mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}'}$ in Eq. (17), then the propagated two-dimensional field given by the term in curly brackets may be approximated by the plane wave $e^{j(2\pi f/c)\sin\vartheta_i \mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}}$. Substitution of this approximation in Eq. (23) reproduces the expressions given by Eqs. (15)–(17) that represent the pattern of illumination.

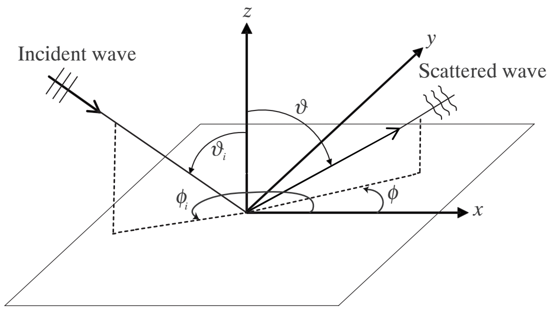


FIG. 1. Scattering geometry. The incident wave has azimuthal angle ϕ_i and polar angle ϑ_i, and the scattered wave has azimuthal angle ϕ and polar angle ϑ.

The illumination pattern or receiver sensitivity of a single ring transducer element at location $\mathbf{x}'(\phi_0)$ is obtained by choosing $E(\mathbf{x}'(\zeta)) = \delta(\zeta-\phi_0)$. For example, making this substitution in Eq. (23) yields

$$\psi_t(\mathbf{x},z,\phi_0) = \int_{\vartheta_i} (f\sin\vartheta_i/c)\lambda(f\sin\vartheta_i/c)e^{j(2\pi f/c)\cos\vartheta_i z} \times \left\{ -2\frac{\partial}{\partial r_{xx'}}[G_0^{(2\mathrm{D})}(\mathbf{x}-\mathbf{x}'(\phi_0), f\sin\vartheta_i)] \right\} d\vartheta_i, \tag{24}$$

in which the dependence of the illumination pattern ψ_t on the azimuthal angle ϕ_0 of the incident illumination is explicitly indicated. Other illumination patterns can be formed by choosing element excitations $E(\mathbf{x}'(\zeta))$ to weight the single-element patterns given in Eq. (24) and using

$$\psi_t(\mathbf{x},z,\phi_0) = \int_{\zeta} E(\mathbf{x}'(\zeta))\psi_t(\mathbf{x},z,\zeta)\,d\zeta. \tag{25}$$

Equations (24) and (25) simplify if the asymptotic form of the Hankel function[16] in $G_0^{(2\mathrm{D})}$ is used.

The scattering geometry and notation used throughout this paper are summarized in Fig. 1. The elevation-focused illumination pattern and reception sensitivity are depicted in Fig. 2. A configuration of the ring transducer and an offset sphere treated in the analysis below is diagramed in Fig. 3.

B. Scattering from a cylinder with focused illumination and reception

Consider a fluid cylinder that has radius a_{cyl}, sound speed c_{cyl}, and density ρ_{cyl} and is positioned with its axis along the z axis in a Cartesian coordinate system in which the coordinates in the x-y plane and elevation dimension are denoted by the vector $\mathbf{x}$ and scalar z, respectively, and the corresponding coordinates in a cylindrical coordinate system are denoted as (r_{xy}, ϕ, z). A cylinder is considered because it is an elementary two-dimensional object with an analytic solution for the scattered pressure, the cross section of a cylinder normal to its axis is circular, and a reconstruction of such a cross section can be compared to a reconstruction of

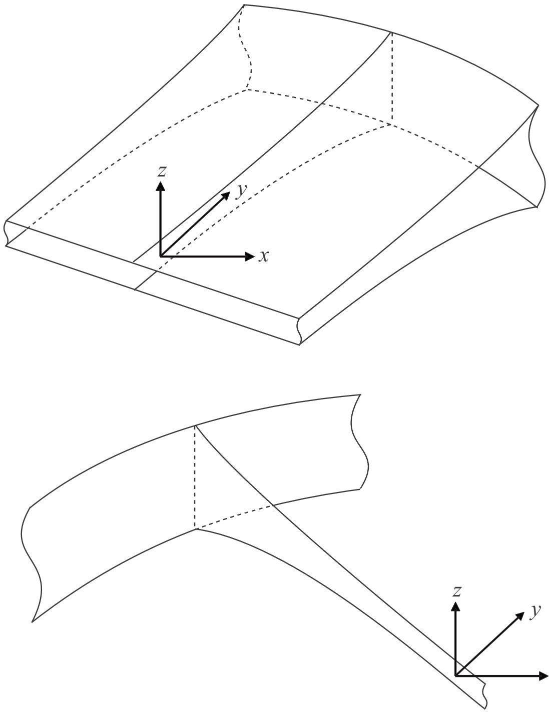


FIG. 2. Elevation-focused beams. The illumination pattern (upper panel) is focused in elevation by the lens factor λ(z) and extends uniformly over the lateral extent of the scattering object as a result of the excitation $E(\mathbf{x})$ applied to the elements. The reception sensitivity (lower panel) shown for a single element is also focused in elevation by the same lens factor and has a lateral sensitivity (not shown) that is also essentially uniform over the extent of the scattering object.

any cross section of a sphere for which all cross sections are circular. Assume that the cylinder is illuminated by a monochromatic plane wave that has time dependence $e^{-j2\pi ft}$ and travels in a direction defined by a unit vector $\mathbf{u}(\phi_i,\vartheta_i)$ with components $\sin\vartheta_i\mathbf{u}_{xy}(\phi_i)$ in the x-y plane and $\cos\vartheta_i$ in the elevation coordinate. By matching the conditions of continuous pressure and normal component of particle velocity at the boundary of the cylinder, the pressure scattered from this cylinder in a lossless fluid background in which the sound speed is c and density is ρ can be expressed using an orthogonal function expansion in cylindrical coordinates as[17,18]

$$p_s(r_{xy},\phi,z,\phi_i,\vartheta_i) = e^{jk_z z}\sum_{n=-\infty}^{+\infty} j^n D_n H_n^{(1)}(k_{xy}r_{xy})e^{jn(\phi-\phi_i)}, \quad r_{xy} > a_{\text{cyl}}. \tag{26}$$

In these expressions, r_{xy} is the radial distance in the x-y plane, ϕ is the azimuthal angle of the observation point, k_{xy} is the wavenumber of the incident wave component in the x-y plane, k_z is wavenumber of the incident wave component in the elevation coordinate, k_{cyl} is the wavenumber in the cylinder, D_n are coefficients depending on the conditions at the boundary of the cylinder, and $H_n^{(1)}$ is a cylindrical Hankel function of the first kind. The coefficients D_n in the orthogonal function expansion for the scattered pressure are found by matching boundary conditions and are the same as those in Ref. 17. The relation between the wavenumber k in the background and the wavenumber components k_{xy} and k_z in the background is

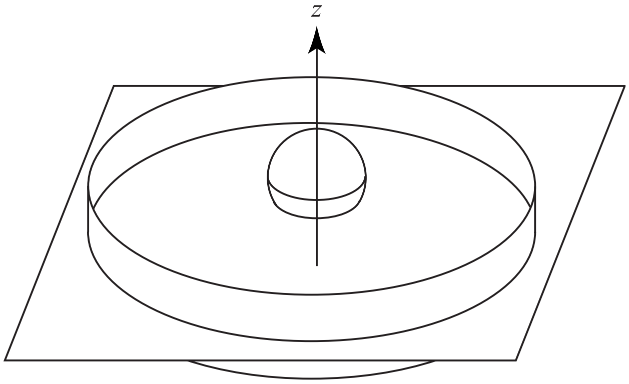


FIG. 3. A sphere with its center vertically offset from the center of the ring transducer imaging plane. Measurement of scattering in this configuration with an illumination pattern and reception sensitivity like those depicted in Fig. 2 is described.

$$k = \sqrt{k_{xy}^2 + k_z^2}. \tag{27}$$

The relation between the wavenumber component k_{xy} in the x-y plane and the polar angle ϑ_i is

$$k_{xy} = 2\pi\frac{f}{c}\sin\vartheta_i, \tag{28}$$

and the relation between the wavenumber component k_z in the elevation coordinate and the polar angle ϑ_i is

$$k_z = 2\pi\frac{f}{c}\cos\vartheta_i. \tag{29}$$

The relation between the wavenumber k_{cyl} in the cylinder and the wavenumber k in the background is

$$k_{\text{cyl}} = 2\pi\frac{f}{c}\sqrt{(c/c_{\text{cyl}})^2 + (\cos\vartheta_i)^2}. \tag{30}$$

The expression for D_n is also valid for a lossy cylinder with attenuation coefficient α_{cyl} in which case the effective planar wavenumber k_{cyl} corresponding to an incident plane wave with angle ϑ_i is

$$k_{\text{cyl}} = \sqrt{(2\pi f/c_{\text{cyl}} + j\alpha_{\text{cyl}})^2 - (2\pi f/c_{\text{cyl}})^2\cos^2\vartheta_i}. \tag{31}$$

The development of an expression for the measurement of scattering from a cylindrical object as a far-field (i.e., plane-wave) sensitivity pattern summed with a sequence of weights requires care because a cylindrical object is unbounded. However, such an accumulation is justified here by the elevation limits imposed by the elevation focus of the illumination. The scattered pressure given in Eq. (26) is, therefore, written as the Herglotz wave,[19]

$$p_s(\mathbf{x},z,\phi_i,\vartheta_i) = -k^2\int_{z'}\int_{\|\mathbf{x}'\|<r_{\text{cyl}}} G_0^{(3\text{D})}(\mathbf{x}-\mathbf{x}',z-z') \times \eta_{\text{cyl}}p(\mathbf{x}',z',\phi_i,\vartheta_i)d^2\mathbf{x}'dz', \tag{32}$$

where $\eta_{\text{cyl}}=1-(c/c_{\text{cyl}})^2$ and $p(\mathbf{x}',z',\phi_i,\vartheta_i)$ is the pressure inside the cylinder given by Eq. (26).

Since Eq. (32) gives the scattering produced by a single plane-wave component of the incident illumination, the scattered pressure produced by an angular spectrum of plane waves is obtained by summing all the angular spectrum components. This yields

$$p_s(\mathbf{x},z,\phi_i) = -k^2 \int_{z'} \int_{\|\mathbf{x}'\|<r_{\mathrm{cyl}}} G_0^{(3\mathrm{D})}(\mathbf{x}-\mathbf{x}',z-z') \times \left[\int_{\vartheta_i} \varphi_t(\vartheta_i)\, \eta_{\mathrm{cyl}} p(\mathbf{x}',z',\phi_i,\vartheta_i)\, d\vartheta_i \right] d^2\mathbf{x}'\, dz' \quad (33)$$

The Green's function in this integral can be interpreted as the sensitivity of a point receiver at the position $(\mathbf{x},z)$ to scattering that originates from a location $(\mathbf{x}',z')$ inside the scattering volume.

A measurement of scattering is the weighted accumulation of such point receivers with the angular spectrum sensitivity identified above. To model this measurement, the angular spectrum expansion for receiver sensitivity given in Eq. (21) is substituted for the Green's function in Eq. (33). This gives

$$M_{\mathrm{cyl}}(\phi,\phi_i) = -k^2 \int_{z'} \int_{\|\mathbf{x}'\|<r_{\mathrm{cyl}}} \left[\int_{\vartheta} \varphi_r(\vartheta) \times e^{j(2\pi f/c)\mathbf{u}(\phi,\vartheta)\cdot(\mathbf{x}',z')} d\vartheta \right] \times \left[\int_{\vartheta_i} \varphi_t(\vartheta_i)\, \eta_{\mathrm{cyl}} p(\mathbf{x}',z',\phi_i,\vartheta_i)\, d\vartheta_i \right] d^2\mathbf{x}'\, dz' . \quad (34)$$

Interchanging the order of integrations over ϑ_i and ϑ with the integrations over $\mathbf{x}'$ and z', factoring out the z component of the exponential term, and noting from Eq. (26) that $p(\mathbf{x}',z',\phi_i,\vartheta_i) = e^{jk_z z'} p(\mathbf{x}',0,\phi_i,\vartheta_i)$ allows the $\mathbf{x}'$ and z' integrals to be separated and the expression for the measurement to be written as

$$M_{\mathrm{cyl}}(\phi,\phi_i) = -k^2 \int_{\vartheta} \int_{\vartheta_i} \varphi_t(\vartheta_i)\varphi_r(\vartheta) \times \left[\int_{z'} e^{j(2\pi f \cos\vartheta/c + k_z)z'} dz' \right] \times \left[\int_{\|\mathbf{x}'\|<r_{\mathrm{cyl}}} e^{j(2\pi f/c)\sin\vartheta\, \mathbf{u}_{xy}(\phi)\cdot\mathbf{x}'}\, \eta_{\mathrm{cyl}} p(\mathbf{x}',0,\phi_i,\vartheta_i)\, d^2\mathbf{x}' \right] d\vartheta\, d\vartheta_i . \quad (35)$$

Since $k_z = 2\pi f \cos\vartheta_i/c$, the z integral evaluates to

$$\int_{z'} e^{j(2\pi f \cos\vartheta/c + k_z)z'} dz' = \delta(f(\cos\vartheta + \cos\vartheta_i)/c) = \frac{\delta(\vartheta - (\pi - \vartheta_i))}{f \sin\vartheta_i/c} . \quad (36)$$

When this result is substituted into Eq. (35), the ϑ variable is forced to assume the value $\pi - \vartheta_i$. However, symmetry in the angular spectra results in $\varphi_r(\pi - \varphi_i) = \varphi_r(\vartheta_i)$ so that the measurement of scattered pressure with elevation-focused illumination and reception can be expressed as

$$M_{\mathrm{cyl}}(\phi,\phi_i) = \int_{\vartheta_i} \frac{\varphi_t(\vartheta_i)\varphi_r(\vartheta_i)}{f \sin\vartheta_i/c} \times \left[-k^2 \int_{\|\mathbf{x}'\|<r_{\mathrm{cyl}}} e^{j(2\pi f/c)\sin\vartheta_i \mathbf{u}_{xy}(\phi)\cdot\mathbf{x}'} \times \eta_{\mathrm{cyl}} p(\mathbf{x}',0,\phi_i,\vartheta_i)\, d^2\mathbf{x}' \right] d\vartheta_i . \quad (37)$$

The integral over $\mathbf{x}'$ is the far-field evaluation of a Herglotz wave for two-dimensional scattering and, thus, can be equated to a radial limit of the expression for $p_s(\mathbf{x},0,\phi_i,\vartheta_i)$ given in Eq. (26) to obtain

$$-k^2 \int_{\|\mathbf{x}'\|<r_{\mathrm{cyl}}} e^{j(2\pi f/c)\sin\vartheta\, \mathbf{u}_{xy}(\phi)\cdot\mathbf{x}'}\, \eta_{\mathrm{cyl}} p(\mathbf{x}',0,\phi_i,\vartheta_i)\, d^2\mathbf{x}' = \lim_{r\to\infty} \left\{ j\sqrt{8\pi j k_{xy}(\vartheta_i) r}\, e^{-jk_{xy}(\vartheta_i) r} \times \sum_n j^n D_n H_n^{(1)}(k_{xy}(\vartheta_i) r) e^{jn(\phi-\phi_i)} \right\} \quad (38)$$

in which the dependency of the x-y plane background wavenumber k_{xy} on the polar angle ϑ_i is indicated explicitly. Use of the asymptotic expression for the Hankel function[15] permits evaluation of the limit in Eq. (38). The result is

$$-k^2 \int_{\|\mathbf{x}'\|<r_{\mathrm{cyl}}} e^{j(2\pi f/c)\sin\vartheta\, \mathbf{u}_{xy}(\phi)\cdot\mathbf{x}'}\, \eta_{\mathrm{cyl}} p(\mathbf{x}',0,\phi_i,\vartheta_i)\, d^2\mathbf{x}' = 4j \sum_n D_n(\vartheta_i) e^{jn(\phi-\phi_i)} \quad (39)$$

in which the dependency of the orthogonal function expansion coefficients D_n on the incident wave polar angle is also indicated explicitly. Substitution of Eq. (39) into Eq. (37) gives

$$M_{\mathrm{cyl}}(\phi,\phi_i) = \int_{\vartheta_i} \frac{\varphi_t(\vartheta_i)\varphi_r(\vartheta_i)}{f \sin\vartheta_i/c} \left[4j \sum_n D_n(\vartheta_i) e^{jn(\phi-\phi_i)} \right] d\vartheta_i . \quad (40)$$

To simplify the expression in Eq. (40) for the measured scattered pressure $M_{\text{cyl}}(\phi,\phi_i)$, the normalized far-field scattering pattern of a cylinder given by the right side of Eq. (39) is denoted using

$$A_{\text{cyl}}(\phi,\phi_i,\vartheta_i)=4j\sum_n D_n(\vartheta_i)e^{jn(\phi-\phi_i)}. \tag{41}$$

The expression for $M_{\text{cyl}}(\phi,\phi_i)$ in Eq. (40) can then be compactly written as

$$M_{\text{cyl}}(\phi,\phi_i)=\int_{\vartheta_i}\frac{\varphi_t(\vartheta_i)\varphi_r(\vartheta_i)}{f\sin\vartheta_i/c}A_{\text{cyl}}(\phi,\phi_i,\vartheta_i)d\vartheta_i. \tag{42}$$

If the angular spectrum is concentrated in the vicinity of the polar angle $\vartheta_i=\pi/2$, then the small angle approximation $\sin\vartheta_i=\cos(\pi/2-\vartheta_i)\approx 1$ may be used in Eq. (40). The result is

$$M_{\text{cyl}}(\phi,\phi_i)=\frac{1}{f/c}\left[\int_{\vartheta_i}\varphi_t(\vartheta_i)\varphi_r(\vartheta_i)d\vartheta_i\right]A_{\text{cyl}}(\phi,\phi_i,a_{\text{cyl}}) \tag{43}$$

in which the dependence of the far-field pattern A_{cyl} on the cylinder radius a_{cyl} is explicitly indicated. Thus, each scattering measurement is a product of a value from the two-dimensional far-field pattern for a cylinder of radius a_{cyl} and a weighting factor that is determined by the elevation overlap of the transmit focus and the receive focus.

The expression for $M_{\text{cyl}}(\phi,\phi_i)$ in Eq. (42) can be efficiently computed because $A_{\text{cyl}}(\phi,\phi_i,\vartheta_i)$, i.e., the bracketed quantity inside the integral in Eq. (40), is a Fourier series for which the coefficients can be obtained using a fast Fourier transform.

C. Scattering from a sphere with focused illumination and reception

Consider a fluid sphere that has radius a_{sph}, sound speed c_{sph}, and density ρ_{sph} and is positioned with its center at the origin of a Cartesian coordinate system in which, as in the case of the cylinder, the coordinates in the x-y plane and elevation dimension denoted by the vector $\mathbf{x}$ and scalar z, respectively, and the corresponding coordinates in a spherical coordinate system are denoted as (r,ϕ,ϑ). A sphere is considered because it is an elementary three-dimensional object with an analytic solution for scattered pressure, has, as noted, a circular cross section like the cross section normal to the axis of a cylinder, and reconstructions of cross sections can, as also noted, be compared to reconstructions of cross sections of a cylinder normal to its axis. Assume that the sphere is illuminated by a monochromatic plane wave that has the same form as the illumination of the cylinder, i.e., has time dependence $e^{-j2\pi ft}$ and travels in a direction defined by a unit vector $\mathbf{u}(\phi_i,\vartheta_i)$ with components $\sin\vartheta_i\mathbf{u}_{xy}(\phi_i)$ in the x-y plane and $\cos\vartheta_i$ in the elevation dimension. By matching the conditions of continuous pressure and normal component of particle velocity at the boundary of the sphere, the pressure scattered from this sphere in a lossless fluid background that is the same as in the case of a cylinder, i.e., in which the sound speed is c and the density is ρ, can be expressed using an orthogonal function expansion in spherical coordinates as[16,17]

$$p_s(r,\phi,\vartheta,\phi_i,\vartheta_i)=\sum_{n=0}^{\infty}\sum_{m=-n}^{+n}4\pi j^n d_n h_n^{(1)}(kr)\times Y_n^m(\vartheta,\phi)Y_n^m(\vartheta_i,\phi_i)^*,\quad r>a_{\text{sph}}. \tag{44}$$

In these expressions, r is the radial distance from the origin to the point of observation, ϕ and ϑ are the azimuthal angle and polar angle, respectively, of the observation point, d_n are coefficients that depend on the conditions at the boundary of the sphere, k_{sph} is the wavenumber in the sphere, $h_n^{(1)}$ is a spherical Hankel function of the first kind, Y_n^m are spherical harmonics, and $*$ denotes complex conjugate. The coefficients d_n in the orthogonal function expansion for the scattered pressure are found by matching boundary conditions and are the same as those in Ref. 17. The spherical harmonics Y_n^m are defined in terms of associated Legendre polynomials P_n^m and are written using the same convention as in Ref. 18. The relation between the wavenumber k_{sph} in the sphere and the wavenumber k in the background is

$$k_{\text{sph}}=\frac{c}{c_{\text{sph}}}k. \tag{45}$$

The expressions for d_n is also valid for a lossy sphere with attenuation coefficient α_{sph} in which case the wavenumber k_{sph} is

$$k_{\text{sph}}=2\pi f/c_{\text{sph}}+j\alpha_{\text{sph}}. \tag{46}$$

Development of an expression for the measurement of scattering from a spherical object as a far-field (i.e., plane-wave) sensitivity pattern summed with a sequence of weights is more straightforward than for a cylindrical object because a spherical object is bounded. If a sphere is illuminated with a plane wave traveling in the direction given by the unit vector $\mathbf{u}(\phi_i,\vartheta_i)$ and the scattered signal in direction given by the unit vector $\mathbf{u}(\phi,\vartheta)$ is measured with a receiver plane-wave sensitivity, then the measurement has a value that is the plane-wave sensitivity times the far-field pattern $A_{\text{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)$ of the sphere. Measurements made with illumination patterns and receiver sensitivities that are described by angular spectra of plane waves can then be found by summing over the measurements for all illumination and receive plane-wave combinations to obtain

$$M_{\text{sph}}(\phi,\phi_i)=\int_{\vartheta}\int_{\vartheta_i}\varphi_t(\vartheta_i)\varphi_r(\vartheta)A_{\text{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)d\vartheta_i d\vartheta. \tag{47}$$

The far-field pattern for the sphere results from taking the normalized radial limit of the scattered pressure given by

$$A_{\text{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)=\lim_{r\to\infty}[-4\pi re^{-jkr}p_s(r,\phi,\vartheta,\phi_i,\vartheta_i)]. \tag{48}$$

Substitution of the expression in Eq. (44) for the pressure scattered from a sphere illuminated by a plane wave traveling in the direction $\mathbf{u}(\phi_i,\vartheta_i)$ yields

$$A_{\mathrm{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)=\lim_{r\to\infty}\Bigg[-4\pi r e^{-jkr}\sum_{n=0}^{\infty}\sum_{m=-n}^{+n}4\pi j^n d_n \times h_n^{(1)}(kr)Y_n^m(\vartheta,\phi)Y_n^m(\vartheta_i,\phi_i)^*\Bigg]. \qquad (49)$$

The limit is found by using the asymptotic form for the spherical Hankel function.[15,19] The result is

$$A_{\mathrm{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)=\frac{j8\pi}{f/c}\sum_{n=0}^{\infty}\sum_{m=-n}^{+n}d_nY_n^m(\vartheta,\phi)Y_n^m(\vartheta_i,\phi_i)^*. \qquad (50)$$

Expressing the spherical harmonics in terms of the Legendre polynomials by using in Eq. (50) yields the relation

$$A_{\mathrm{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)=\frac{j8\pi}{f/c}\sum_{n=0}^{\infty}\sum_{m=-n}^{+n}d_n\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} \times P_n^m(\cos\vartheta)e^{jm\phi}\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} \times P_n^m(\cos\vartheta_i)e^{-jm\phi_i}, \qquad (51)$$

and substituting this expression for $A_{\mathrm{sph}}(\phi,\vartheta,\phi_i,\vartheta_i)$ in Eq. (47) for $M_{\mathrm{sph}}(\phi,\phi_i)$ results in

$$M_{\mathrm{sph}}(\phi,\phi_i)=\frac{j8\pi}{f/c}\sum_{n=0}^{\infty}\sum_{m=-n}^{+n}d_nT_n^mR_n^me^{jm(\phi-\phi_i)}, \qquad (52)$$

where the transmit sensitivity factor T_n^m is defined as

$$T_n^m=\int_{\vartheta_i}\varphi_t(\vartheta_i)\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\vartheta_i)d\vartheta_i \qquad (53)$$

and the receive sensitivity factor R_n^m is defined as

$$R_n^m=\int_{\vartheta}\varphi_r(\vartheta)\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\vartheta)d\vartheta. \qquad (54)$$

Interchanging the summation over n and the summation over m yields

$$M_{\mathrm{sph}}(\phi,\phi_i)=\frac{j8\pi}{f/c}\sum_{m=-\infty}^{+\infty}\left[\sum_{n=|m|}^{\infty}d_nT_n^mR_n^m\right]e^{jm(\phi-\phi_i)}. \qquad (55)$$

The expression for $M_{\mathrm{sph}}(\phi,\phi_i)$ in Eq. (55) for a spherical object can, as in the case of $M_{\mathrm{cyl}}(\phi,\phi_i)$ in Eq. (42) for a cylindrical object, be efficiently computed because the outer summation is a Fourier series for which the coefficients given by the bracketed quantity can be obtained using a fast Fourier transform. In this computation, additional efficiency is gained by evaluating the transmit sensitivity factors T_n^m and receive sensitivity factors R_n^m first and storing them. Also, since the sensitivity factors are independent of the sphere radius, once a set of factors have been computed, the factors may be used for a sphere with different acoustic properties (radius, sound speed, density, and position) as long as a sufficient number of orders for the different parameters have been computed.

The foregoing analysis has assumed that the sphere is located in the imaging plane at the center of the ring transducer, i.e., at (0,0,0). However, the analysis is easily modified to hold for a scattering object that is translated by an offset of $(\Delta\mathbf{x},\Delta z)$ from the origin because translation only changes the object far-field scattering pattern A_{obj} by the phase factor $e^{j(2\pi f/c)[\mathbf{u}(\phi_i,\vartheta_i)+\mathbf{u}(\phi,\vartheta)]\cdot(\Delta\mathbf{x},\Delta z)}$, i.e.,

$$\hat{A}_{\mathrm{obj}}(\phi,\vartheta,\phi_i,\vartheta_i)=A_{\mathrm{obj}}(\phi,\vartheta,\phi_i,\vartheta_i) \times e^{j(2\pi f/c)[\mathbf{u}(\phi_i,\vartheta_i)+\mathbf{u}(\phi,\vartheta)]\cdot(\Delta\mathbf{x},\Delta z)} \qquad (56)$$

in which $\hat{}$ denotes a shifted pattern. For a vertical offset from the center of the transducer ring, the phase factor assumes the simple form

$$e^{jk[\mathbf{u}(\phi_i,\vartheta_i)+\mathbf{u}(\phi,\vartheta)]\cdot(0,\Delta z)}=e^{jk(\cos\vartheta_i+\cos\vartheta)\Delta z}. \qquad (57)$$

Introduction of the factor given by Eq. (57) in foregoing analysis for a spherical scattering object only changes the final results by altering the expressions for the coefficients T_n^m and R_n^m given above by Eqs. (53) and (54), respectively, to have the forms

$$T_n^m=\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}\int_{\vartheta_i}\varphi_t(\vartheta_i)e^{j(2\pi f/c)\Delta z\cos\vartheta_i} \times P_n^m(\cos\vartheta_i)d\vartheta_i \qquad (58)$$

and

$$R_n^m=\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}\int_{\vartheta}\varphi_r(\vartheta)e^{j(2\pi f/c)\Delta z\cos\vartheta} \times P_n^m(\cos\vartheta)d\vartheta. \qquad (59)$$

These expressions also describe the effect of vertically translating (by $-\Delta z$) the origin of the plane-wave components in the angular spectrum expansions for the illumination pattern and the sensitivity of reception when the center of the sphere is at the origin of the coordinate system.

D. Equal angles

When the scattered pressure is needed at a number of temporal frequencies, such as for the extended eigenfunction method of image reconstruction used in this paper, an angular spectrum calculation is required at multiple temporal harmonic frequencies. If the transmit angular spectrum is sampled at the Nyquist rate for each temporal frequency, a different set of angles can result in the angular spectrum. However, different angles for each temporal frequency creates a bottleneck that slows the calculation of scattering at multiple frequencies because the associated Legendre polynomials in the definition of the spherical harmonics must be recomputed for each temporal frequency. This recalculation is avoided by choosing the number of samples in the aperture to be sufficiently large to prevent spatial aliasing at the highest temporal frequency and by varying the sampling rate to permit use of the same angles for all temporal frequencies.

In the technique, the aperture is first spatially sampled at the Nyquist rate determined by the highest temporal frequency f_h. This spatial sampling rate ν_h is

$$\nu_h = 2\frac{f_h}{c}. \tag{60}$$

Next, the number of points N_a in the spatial Fourier transform is chosen to avoid aliasing at the highest temporal frequency. For an aperture of length L, the number of points is

$$N_a = L\nu_h. \tag{61}$$

Then, in order to maintain the same set of angles for each temporal frequency, this number of samples is fixed and used to define the constant angular spacing $\Delta\vartheta$ given by

$$\Delta\vartheta = \arccos(2/N_a). \tag{62}$$

This results in the set of angles

$$\{\vartheta_n\} = \left\{\arccos\left(n\frac{c}{Lf_h}\right)\right\}, \quad n = 0, 1, \ldots, N_a - 1 \tag{63}$$

that are the same for the calculation of elevation-focused illumination and reception at each temporal frequency.

This technique does not require an extraordinarily large number of points in the Fourier transform. For example, the extreme case in which f_h=5.0 MHz, L=60 mm, and c=1.5 mm/μs results in ν_h=6.7 cycle/mm and N_a=400 samples or Fourier transform points. Thus, although the aperture is always sampled at a rate determined by the highest temporal frequency, the so-called equal-angle approach still requires fewer computations than the approach of recomputing the associated Legendre polynomials at each temporal frequency.

E. Verification in the far field by using a weak-scattering approximation

The scattered pressure for elevation-focused illumination and reception calculated for a cylinder using Eq. (42) and for a sphere using Eq. (55) can be compared to an independent calculation of scattered pressure obtained under a weak-scattering approximation in the far field to instill confidence that the calculations are correct. For this comparison, the normalized limit used in Eq. (48) is applied to the Herglotz expression for scattering. This yields the far-field pattern

$$A(\phi,\vartheta,\phi_i,\vartheta_i) = k^2\int_{\mathbf{x}}\int_z \eta(\mathbf{x},z)p(\mathbf{x},z,\phi_i,\vartheta_i) \times e^{j(2\pi f/c)\mathbf{u}(\phi,\vartheta)\cdot(\mathbf{x},z)}dz d\mathbf{x}, \tag{64}$$

where $p(\mathbf{x},z,\phi_i,\vartheta_i)$ is the total pressure and where the z integration is restricted to the range of elevations over which the product of the transmit focus and the receive focus is appreciable.

Using a weak-scattering approximation, i.e., $p(\mathbf{x},z,\phi_i,\vartheta_i) \approx p_i(\mathbf{x},z,\phi_i,\vartheta_i)$, in Eq. (64) gives

$$A(\phi,\vartheta,\phi_i,\vartheta_i) = k^2\int_{\mathbf{x}}\int_z \eta(\mathbf{x},z) \times e^{j(2\pi f/c)[\mathbf{u}(\phi_i,\vartheta_i)+\mathbf{u}(\phi,\vartheta)]\cdot(\mathbf{x},z)}dz d\mathbf{x}. \tag{65}$$

The far-field scattered pressure with elevation-focused illumination and reception described by angular spectra can then be written as

$$\tilde{A}(\phi,\vartheta,\phi_i,\vartheta_i) = \int_{\vartheta}\int_{\vartheta_i} A(\phi,\vartheta,\phi_i,\vartheta_i)\varphi_t(\vartheta_i)\varphi_r(\vartheta)d\vartheta_i d\vartheta \tag{66}$$

in which $\tilde{}$ denotes focusing in elevation. Substitution of Eq. (65) into Eq. (66) for elevation-focused illumination and reception and use of Eqs. (11) and (21) for $\psi_t(\mathbf{x},z)$ and $\psi_r(\mathbf{x},z)$, respectively, yields

$$\tilde{A}(\phi,\vartheta,\phi_i,\vartheta_i) = k^2\int_{\mathbf{x}}\int_z \eta(\mathbf{x},z)\psi_t(\mathbf{x},z)\psi_r(\mathbf{x},z)dz d\mathbf{x}. \tag{67}$$

The expressions for $\psi_t(\mathbf{x},z)$ and $\psi_r(\mathbf{x},z)$ in Eq. (67) can be approximated in the scattering region by the product of an amplitude envelope and a plane wave propagating in the direction of the illumination for $\psi_t(\mathbf{x},z)$ and in the direction of reception for $\psi_r(\mathbf{x},z)$, i.e., by the expressions

$$\psi_t(\mathbf{x},z) = |\psi_t(z)|e^{j(2\pi f/c)\mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}} \tag{68}$$

and

$$\psi_r(\mathbf{x},z) = |\psi_r(z)|e^{j(2\pi f/c)\mathbf{u}_{xy}(\phi_i)\cdot\mathbf{x}}, \tag{69}$$

respectively, in which $(2\pi f/c)\mathbf{u}_{xy}(\phi_i)$ and $(2\pi f/c)\mathbf{u}_{xy}(\phi)$ are the wave vectors of the incident illumination and the scattered wave, respectively, in the x-y plane. Using these expressions in Eq. (67), the normalized far-field scattered pressure for elevation-focused illumination and reception under a weak-scattering approximation becomes

$$\tilde{A}(\phi,\phi_i) = k^2\int_{\mathbf{x}} \eta_p(\mathbf{x})e^{j(2\pi f/c)[\mathbf{u}_{xy}(\phi_i)+\mathbf{u}_{xy}(\phi)]\cdot\mathbf{x}}d\mathbf{x} \tag{70}$$

in which

$$\eta_p(\mathbf{x}) = \int_z \eta(\mathbf{x},z)|\psi_t(z)||\psi_r(z)|dz. \tag{71}$$

Equation (70) shows that the normalized far-field pattern in the presence of elevation-focused illumination and reception is the Fourier transform of the illuminated volume of $\eta(\mathbf{x},z)$ weighted and projected onto the x-y plane according to Eq. (71). When the variation in the medium is essentially a two-dimensional quantity, i.e., $\eta(\mathbf{x},z) \approx \eta(\mathbf{x})$, Eq. (70) simplifies to

$$\tilde{A}(\phi,\phi_i) = Wk^2\int_{\mathbf{x}} \eta(\mathbf{x})e^{j(2\pi f/c)[\mathbf{u}_{xy}(\phi_i)+\mathbf{u}_{xy}(\phi)]\cdot\mathbf{x}}d\mathbf{x} \tag{72}$$

in which

$$W = \int_z |\psi_t(z)||\psi_r(z)|dz \tag{73}$$

and is a scale factor that accounts for use of elevation-focused illumination and reception.

Equations (72) and (70) provide expressions to verify the scattered pressure given for a cylinder by Eq. (42) and for a sphere by Eq. (55), respectively, in the presence of elevation focusing.

TABLE I. Pulse, focus, ring, and background parameters used in calculations of scattering.

Parameter (symbol)	Value
Pulse-width parameter (σ_t)	0.220 μs
Center frequency (f_0)	2.50 MHz
Aperture width parameter (σ_a)	4.231 mm
Ring radius (r_0)	75.0 mm
Geometric focus adjustment (r_a)	17.224 mm
Background sound speed (c)	1.590 mm/μs
Background density (ρ)	0.997 g/mm^3

III. COMPUTATIONAL METHODS

The temporal envelope of elevation-focused transmit-receive beam patterns was calculated for a Gaussian-shaped bandpass pulse given by Eq. (10) and transmit and receive apertures with identical Gaussian amplitude elevation weighting given by Eq. (19). In the calculations, the transmit and the receive beams given by Eqs. (16) and (18) were found at each temporal frequency in the pulse, corresponding temporal-frequency patterns were multiplied, and an inverse Fourier transform was taken to obtain the temporal beam as a function of time. Values of parameters in the calculations are given in Table I. The peak of the temporal envelope was then found in elevation cross sections to provide a description of the elevation-focused beams used in calculations of scattering by cylinders and by spheres.

To instill confidence that the scattered pressure for elevation-focused illumination and reception calculated using Eq. (40) for a cylinder is correct, the far-field scattered pressure under a weak-scattering approximation for a cylinder was computed using the Fourier transform relation given in Eq. (72) for a two-dimensional scattering object. To instill confidence also that the scattered pressure for elevation-focused illumination and reception calculated using Eq. (55) for a sphere is correct, the far-field scattered pressure under a weak-scattering approximation was computed for a sphere using the Fourier transform relation given in Eq. (70) for a projection of a three-dimensional scattering object and compared to the corresponding far-field normalized expressions used in the two-dimensional reconstruction process. Values of parameters in the two calculations are given in Table II.

The pressure scattered by a cylinder and by a sphere each with elevation-focused illumination and reception was computed at multiple temporal frequencies by using Eqs. (40) and (55), respectively, with the acoustic parameters of the spheres and cylinders having the values shown in Table III. In the computations, the axis of the cylinder was along the z axis of the coordinate system, the center of the sphere was 0, 3, or 5 mm above the center of the focus that was at the origin of the coordinate system, the diameter of the sphere was 20 mm, and the diameter of the cylinder was the same as the diameter of the cross section of the sphere in the x-y plane at $z=0$; i.e., the cylinder diameter was 20, 19.08, and 17.32 mm, respectively. The computations were performed at 512 receive azimuthal angles and for 512 incident wave azimuthal angles each equally spaced between 0 and 2π around the circumference of the ring. This number of incident wave angles and receive angles was chosen to provide a sufficient number of angular samples for the size objects being reconstructed to avoid aliasing in the reconstruction process.[13] The elevation-focused angular spectra that modeled the described elevation focusing in the ring transducer system were calculated using the equal-angle method to reduce computation time. The illumination and reception apertures included a Gaussian-shaped amplitude weight given by Eq. (19) and a length factor given by Eq. (20) to create a focus with a peak at the center of the 150-mm diameter ring transducer.

TABLE II. Scattering object parameters for comparison of analytic solution with weak-scattering far-field approximation.

Parameter (symbol)	Value
Radius (r_{cyl}, r_{sph})	10.0 mm
Sound speed (c_{cyl}, c_{sph})	1.50901 mm/μs
Attenuation slope (β_{cyl}, β_{sph})	0.0 dB/(cm MHz)
Density (ρ_{cyl}, ρ_{sph})	0.997 g/mm^3
Frequency (f)	1.40625 MHz

TABLE III. Cylinder and sphere parameters used in elevation-focused calculations of scattering.

Parameter	Value
Sound speed	1.570 mm/μs
Attenuation slope	0.462 dB/(cm MHz)
Density	0.997 g/mm^3

After the scattered pressure for elevation-focused illumination and reception was calculated at each temporal frequency, the pressure was normalized by the factor W given by Eq. (73) to compensate for the gain from elevation-focused illumination and reception. The resulting scattered pressure was then extrapolated to the far field using a two-dimensional extrapolation procedure described in Ref. 14 and normalized to remove the cylindrical-wave factor weighting the far-field pattern.

Reconstructions were obtained using the extended eigenfunction method described in Ref. 14. In the extension, scattering by a model object is calculated to form the difference between the scattering by the true object and the scattering by the model object. The scattering potential defined by this difference is expanded in a basis of products of acoustic fields. These fields are defined by eigenfunctions of the scattering operator associated with the estimate.

Computations that implement the reconstruction process were performed in five steps that are detailed in Ref. 14 and summarized here for convenient reference. In the first step, inner products needed for the computation of the coefficients in the expansion of the scattering potential are found by a two-dimensional Fourier transform of the normalized far-field pattern that is the difference between the calculated scattering by the true object and the calculated scattering by the model object. In the second step, the coefficients in the expansion were computed using symmetries that reduce the amount of computation for radial objects like those considered in this study. In the third step, a potential that is the

TABLE IV. Reconstruction parameters used in eigenfunction method of image reconstruction.

Parameter	Value
Initial frequency	1.40625 MHz
Frequency increment	78.125 kHz
Final frequency	3.4375 MHz

lowpass difference between the potential of the true scattering object and the model scattering object was found using a linearization that also exploits symmetries that exist in relations for radial scattering objects. In the fourth step, a weighted average of the potentials reconstructed at each temporal frequency was computed after the potentials were phase compensated using a phase that maximizes the coherence of the reconstructions as a function of frequency. In the fifth step, iteration was used to refine the estimate of the scattering potential until the difference potential between the model and the true object is sufficiently small.

Since the eigenfunction method utilizes *a priori* knowledge, an estimate of the scatterer shape and acoustic parameters is necessary for the reconstructions. A circular shape, i.e., the cross section of a sphere and the cross section of a cylinder normal to its axis, was used in the model for the reconstruction of a sphere and a cylinder. In the case of a cylinder, the radius of the cylinder was assumed known. In the case of a sphere, the estimated radius was the average radius of the volume isolated by the elevation-focused illumination and reception. The value of sound speed in the model for each iteration was the average of the sound speed in a 2-mm diameter region in the center of the previous estimate. The value of attenuation in the model was found frequency by frequency by minimizing the norm of the difference between the far-field scattering by the model and the far-field scattering obtained from the object over the range of 0°–10° using elevation-focused transmit-receive beams. Since the calculations were designed to investigate the use of two-dimensional algorithms with three-dimensional objects, reconstructions of spheres using elevation-focused illumination and reception were compared to reconstructions of cylinders. For the comparisons, cross sections of a sphere were reconstructed for three offsets of the center of the sphere above the center of the focus. Parameters used in the reconstruction process are listed in Table IV. The calculations were repeated using a 4-mm diameter sphere that mostly fits within the half-amplitude (i.e., −6 dB) elevation width of the elevation-focused illumination and reception.

IV. NUMERICAL RESULTS

Cross sections and profiles of the peak temporal envelope of the elevation-focused transmit-receive beam patterns for the parameter values in Table I and depictions of the geometry for which the envelopes were computed are shown in Fig. 4. For these parameter values that were chosen to model the pulse and the focusing in the ring transducer system described in Ref. 3, the −3 dB width in elevation is 2.1 mm and remains the same as the transmit-receive angle varies, and the −3 dB focal length increases monotonically from 72.1 mm at a transmit-receive angle of 180° to 125.6 mm at a transmit-receive angle of 90°. The corresponding values for the −6 dB width in elevation and the range of the −6 dB focal length are 3.2 mm and 124.9–150.0 mm, respectively.

The analytic solution for scattering from a cylinder and from a sphere each with elevation-focused illumination and reception and the corresponding far-field weak-scattering approximations are shown in Fig. 5 for the parameter values in Table II. At the 1.40625 MHz frequency of the calculations, the combined illumination-reception elevation focusing resulted in a −6-dB beam width of 5.4 mm in elevation. The

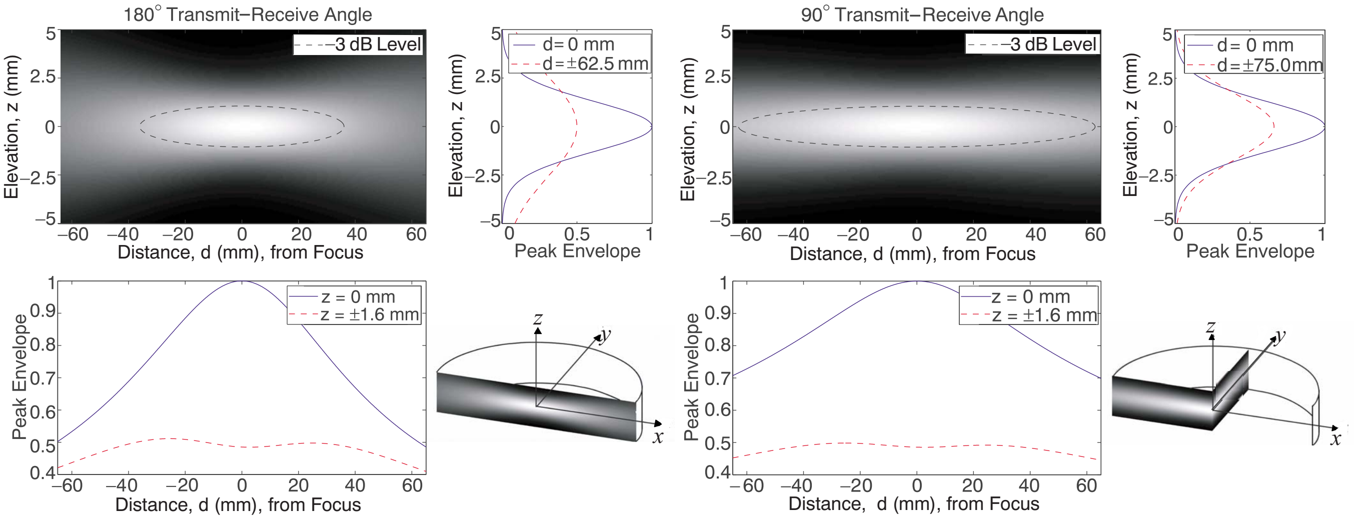


FIG. 4. (Color online) Peak temporal envelope of the elevation-focused transmit-receive beam pattern for transmit and receive apertures with identical Gaussian amplitude weighting in elevation and a bandpass pulse with a Gaussian envelope. The grayscale plots are cross sections of the envelope normalized to 1.0 at $(x,z)=(0,0)$ and shown on a linear scale. The Cartesian plots are profiles of the envelope in the cross section where the envelope in orthogonal grayscale-plot coordinate is a maximum (i.e., 0 dB) and is one-half the maximum (i.e., −6 dB relative to the maximum) in that coordinate. The three-dimensional diagrams show the spatial orientation of grayscale cross sections. For a transmit-receive angle of 180°, the cross sections are the same at every value of the coordinate orthogonal to the plane of the cross section. For a transmit-receive angle of 90° and the focus parameters in this study, the cross sections are essentially constant in the $x-y$ plane when $-20 \leqslant x, y \leqslant +20$ mm.

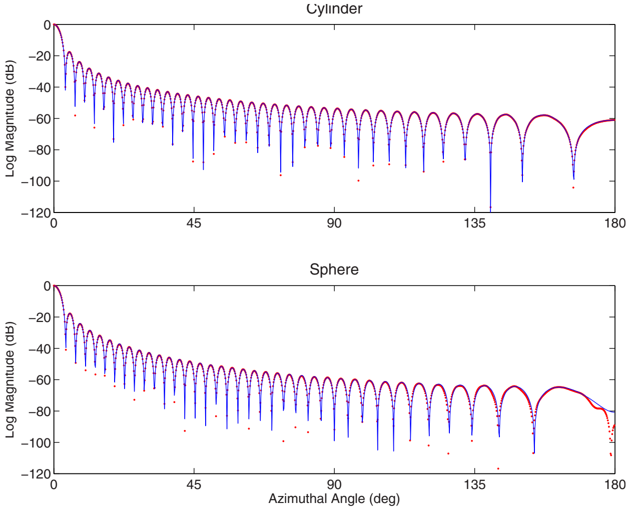


FIG. 5. (Color online) Comparison of the analytic solution for scattering from a lossy cylinder with elevation-focused illumination and reception and the corresponding far-field weak-scattering approximation and comparison of the analytic solution for scattering from a lossy sphere with elevation-focused illumination and reception and the corresponding far-field weak-scattering approximation. The scattered pressure in each case is normalized to 1.0 in the forward (0° azimuthal angle) direction. The center of the sphere is at the center of the focus in elevation. —, far-field analytic solution; ····, far-field weak-scattering solution.

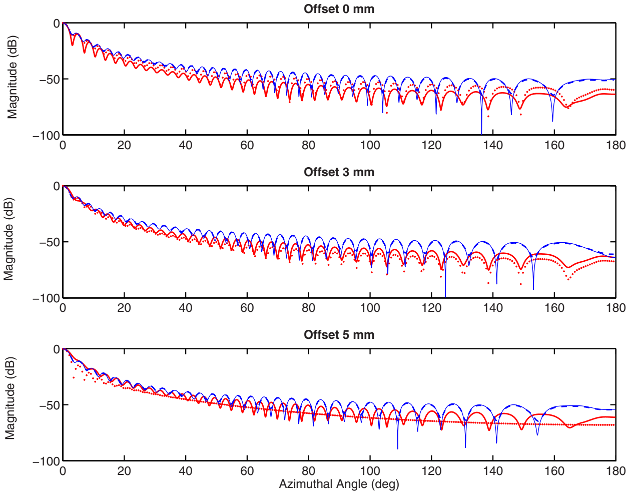
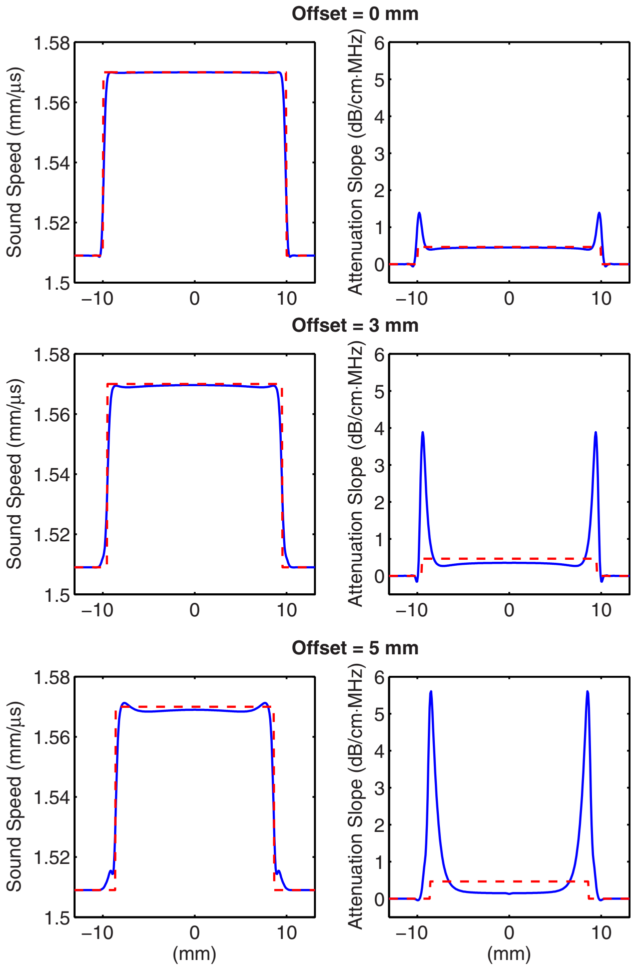


FIG. 6. (Color online) Comparison of plane-wave illumination and point reception with elevation-focused illumination and reception for a lossy cylinder and for a lossy sphere. The scattered pressure in each case is evaluated in the far field and normalized to 1.0 in the forward (0° azimuthal angle) direction. In the calculations, the center of the sphere was vertically offset from the center of the focus by the distance given above each panel, and the radius of the cylinder was the radius of the circular intersection of the sphere with the ring plane through the center of the focus in elevation, i.e., 10, 9.54, and 8.66 mm in the upper, middle, and lower panels, respectively. —, cylinder: plane-wave illumination and point reception; – – –, cylinder: elevation-focused illumination and reception; **—**, sphere: plane-wave illumination and point reception; ····, sphere: elevation-focused illumination and reception.

close agreement of the two plots for the cylinder shows that the two independent calculations of scattered pressure from a cylinder produce essentially the same values, and the corresponding close agreement of the two plots for the sphere shows that the two independent calculations of scattered pressure from a sphere produce essentially the same values. This agreement instills confidence that the calculations of scattering using elevation-focused illumination and reception are correct. The similarity in the scattering from the cylinder and from the sphere supports the expectation that scattering produced by elevation-focused illumination and measured using elevation-focused reception can be treated as two-dimensional when a three-dimensional scattering object has an essentially constant cross section throughout the extent of the elevation-focused beams as is the case in the calculations shown.

Scattering from a cylinder and from a sphere observed using elevation-focused illumination and reception and observed using plane-wave illumination and point reception is compared in Fig. 6 for the parameter values in Table III in which the offsets of the sphere are the vertical distance above the center of the focus. As noted above, the combined illumination-reception elevation focus at the frequency of 1.40625 MHz resulted in a 5.4-mm and 6-dB width in elevation. For the scattering object parameter values that were chosen to be clinically relevant for imaging two-dimensional and three-dimensional objects with the previously noted ring transducer system, the plot for elevation-focused illumination and elevation-focused reception is essentially identical to the plot for plane-wave illumination and point reception for the cylinder. The corresponding plots for a sphere show a difference that increases as the center of the sphere is more offset from the imaging plane and is due to the elevation-focused illumination and reception. For each of the three offsets with elevation focusing, the high amplitude scattering from the sphere in the forward direction (i.e., at low azimuthal angles) is similar to the scattering amplitude from the corresponding cylinder, but the low amplitude scattering in the backward direction (i.e., high azimuthal angles) lacks similarity except for the zero-offset case. The similarity in the forward direction implies that low spatial-frequency components in reconstructions made for a spherical object by using elevation-focused illumination and elevation-focused reception are well represented by a two-dimensional model, while the lack of similarity in the backward direction implies that high spatial-frequency components in reconstructions made for a sphere by using elevation-focused illumination and reception may not be well represented by a two-dimensional model.

Radial profiles through reconstructions obtained by using transmit-receive elevation-focused beams and the two-dimensional form of eigenfunction method are shown in Fig. 7 for a 20-mm diameter sphere with an offset of the center 0, 3, and 5 mm from the imaging plane. The profiles of sound speed and attenuation slope show differences between the ideal radial profiles formed by intersection of the sphere and the imaging plane and the radial profiles in reconstructions obtained using transmit-receive elevation-focused beams and a two-dimensional reconstruction algorithm. The differences between the ideal and reconstructed profiles of sound speed and attenuation slope are larger close to the edges of the reconstructed spheres. The differences are small when the portion of the sphere located in the focused beam is relatively constant in elevation, but the differences increase as the center of the sphere is offset further from the plane of the



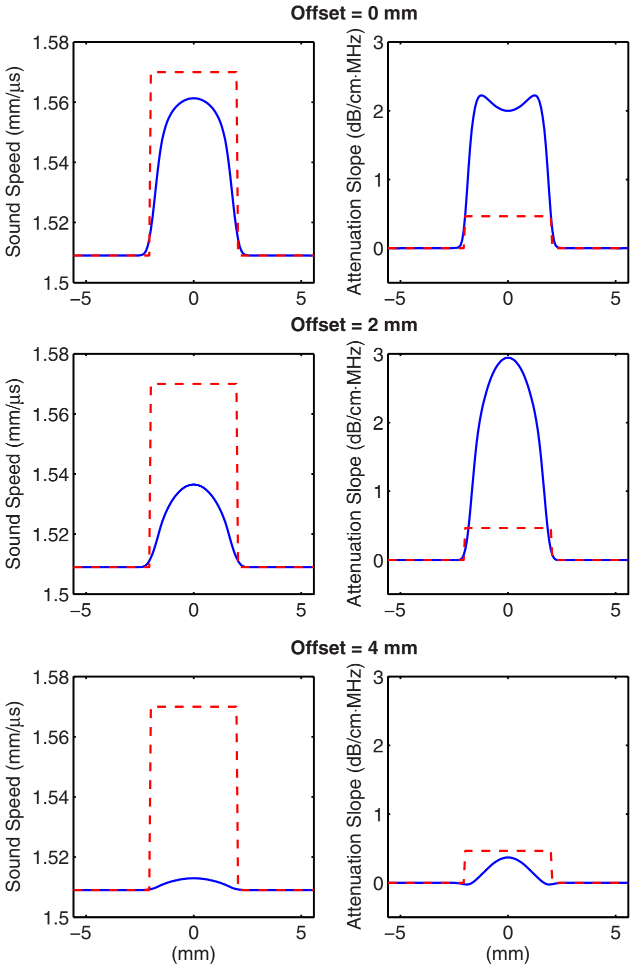


FIG. 7. (Color online) Radial profiles of sound speed and attenuation slope for a 20-diameter lossy sphere with center offset of 0, 3, or 5 mm from the imaging plane that is in the center of the elevation-focused transmit-receive beam. The solid lines are profiles through reconstructions using scattering from the sphere. The dashed lines are profiles through a cylinder with the diameter of the disk formed by the intersection of the sphere with the imaging plane. —, sphere; – – –, cylinder.

FIG. 8. (Color online) Radial profiles of sound speed and attenuation slope for a four-diameter lossy sphere with center offset of 0, 2, or 4 mm from the imaging plane that is in the center of the elevation-focused transmit-receive beam. The solid lines are profiles through reconstructions using scattering from the sphere. The dashed lines are profiles through a cylinder with the diameter of the sphere. —, sphere; – – –, cylinder.

focused beam. Large artifacts at the profile edges and large value differences in the center of the profiles indicate that three-dimensional scattering is not being adequately included in the elevation-focused beams.

Analogous radial profiles through reconstructions obtained using the same beams and two-dimensional form of the eigenfunction method are shown in Fig. 8 for a 4-mm diameter sphere with an offset of the center 0, 2, and 4 mm from the imaging plane. Because the −6 dB width of elevation focus in the calculations was 3.2 mm, the 4-mm diameter sphere with zero offset was essentially contained within the elevation-focused beam in zero-offset case, partially encompassed by the elevation-focused beam in the 2-mm offset case, and at the edge of the elevation-focused beam in the 4-mm offset case. These beam-object relations result in less effect on the overall propagation time of the incident wave and in substantial scattered energy outside the receive aperture. The consequence of the propagation time effect is that the sound speed profiles are generally lower than ideal values defined by a central cross section of the sphere, and the consequence of the unmeasured scattering energy is that the attenuation slope profiles can be much larger than ideal values again defined by a central cross section of the sphere.

V. DISCUSSION

The theoretical development that starts with a wave equation in three dimensions and yields expressions for measurements using elevation-focused illumination and reception has a special structure that merits note. The approach yields a two-dimensional wave equation with a term that describes the effect of scattering object variation in the width of the elevation-focused beams. The development uses expressions for the transmit beam pattern and the receive sensitivity pattern in forms that show the reciprocal nature of transmission and reception. The angular spectrum representation of transmit illumination and reception sensitivity facilitates the use of known relations for scattering by regular objects illuminated by monochromatic plane waves and for far-field reception of the scattering. Efficient calculation of scattering by elevation-focused transmit-receive beams temporal frequency by temporal frequency is enabled by varying the number of samples in the aperture to allow use of the same angles for angular spectrum beam components at each temporal frequency. The consideration of cylindrical and spherical scattering objects that have the same cross sections permits direct comparison of images formed using elevation-focused beams and two-dimensional reconstruction methods. Overall, the theoretical formulation is general but facilitates computational elucidation of phenomena influencing the accuracy of using a two-dimensional reconstruction method to image objects having a variable cross section within the width of elevation-focused transmit-receive beams.

The two-dimensional reconstruction method used in these studies is an iterative multifrequency inverse scattering algorithm that relies on successive improvement of a model for the scattering object. The algorithm converges to an accurate representation of the object intrinsic parameters when the initial model is chosen to be a reasonably good approximation to the scattering object. A good approximation, however, does not necessarily mean that the model has intrinsic parameters approximating those of the scattering object but, rather, that the two-dimensional scattering by the model approximates the scattering from the cross section of the object in the elevation-focused transmit-receive beams. Good approximations in this study were obtained using model objects that include additional frequency-dependent attenuation to compensate for energy transported away from the plane of the cross section by vertical components of wave propagation.

The attenuation assigned to the homogeneous cylinders that were used to model the sphere cross sections can be divided into two additive components. One component is the intrinsic attenuation resulting from absorption of energy by the scattering medium and is an important characteristic of the medium. The other component is the extrinsic attenuation resulting from energy loss caused by spreading of the elevation-focused beam and by out-of-plane reflection and refraction at the surface of an object. Energy reflected into the plane of the cross section is included in the scattering measurements and, therefore, is not included in the energy loss associated with extrinsic attenuation. Extrinsic attenuation that is not uniform across an object with homogeneous intrinsic attenuation could be modeled more accurately by a spatially varying attenuation parameter. The model could also include boundary conditions that extend over a spatial region rather than being abrupt. However, these refinements increase computational cost and were not needed to secure convergence of the reconstructions in this study.

The attenuation for the two-dimensional model cylinders was chosen in this study frequency by frequency to minimize the norm of the residual error between the two-dimensional scattering from the cylindrical model and the scattering from the section of the sphere in the elevation-focused beam. Since the final rendering of each cross section is formed as a sum of the model parameters and the residual reconstruction, the extrinsic attenuation must be estimated and subtracted from the final attenuation in order to assign a value to the intrinsic attenuation. A portion of extrinsic attenuation that is due to the spreading of the focus may be conveniently estimated by simulation of elevation-focused scattering from a three-dimensional cylinder with the same radius, sound speed, and attenuation as in the two-dimensional model. However, the contribution to extrinsic attenuation that is due to interaction with an object boundary that varies within the elevation-focused beam requires *a priori* knowledge of the boundary variation in elevation. Lack of this information may preclude accurate estimation of intrinsic attenuation but is not an obstacle to high-fidelity reconstruction of sound speed.

Out-of-plane reflection and refraction at the boundary of an object produce a loss even when the object has uniform intrinsic attenuation. A cylinder model is only able to compensate for this energy by means of a uniform increase in attenuation throughout the cylinder. However, the residual reconstruction in the final iteration of the inverse scattering computation depicts how this attenuation should be distributed. This residual term corrects the uniform attenuation of the model by modestly reducing attenuation in the interior region and substantially increasing attenuation near the boundary. Since extrinsic attenuation is small in regions far from the boundary, values of reconstructed attenuation approximate intrinsic attenuation near the center of the cross section.

The effect of unmeasured scattering on image reconstruction using a two-dimensional method is illustrated by calculations of scattering from which the reconstruction proceeds. First, however, the excellent agreement between the magnitude of the scattered amplitude obtained with a far-field weak-scattering approximation and the magnitude of the scattered amplitude obtained using a direct evaluation of the orthogonal function expansion for the scattered amplitude with elevation-focused beams in Fig. 5 is noted for both the scattering from a lossy cylinder and the scattering from a lossy sphere. This agreement gives confidence that other calculations of scattering using orthogonal function expansions with elevation-focused transmit-receive reception for a cylinder and for a sphere are correct. The comparison of plane-wave illumination and point reception with elevation-focused illumination and reception in Fig. 6 shows that calculations using plane-wave illumination and point reception can accurately describe illumination and reception with elevation-focused transmit-receive beams for a cylinder. The analogous comparison for a sphere shows an increasing lack of correspondence in the magnitude of scattered amplitude for a sphere as the center of the sphere is more offset from the imaging plane and the cross section of the sphere consequently varies more in the width of the elevation-focused beam. The comparison for a sphere centered in the imaging plane shows, however, that plane-wave illumination and point reception can reasonably approximate elevation-focused transmit-receive beams when the cross section of the sphere is relatively constant within the elevation beam width.

Features of sound speed and attenuation slope profiles in Fig. 7 for a 20-mmdiameter sphere with a center at different offsets from the imaging plane merit comment because the features also illustrate points noted above. The radial profile of sound speed in the reconstruction for no elevation offset closely matches the ideal profile of sound speed in the disk formed by the intersection of the sphere and the imaging plane. The corresponding radial profile of attenuation slope also closely matches the ideal profile of attenuation slope in the disk formed by the intersection of the sphere and the imaging plane except at the edges. The profiles of sound speed and attenuation slope in the reconstruction for a 3-mm offset have similar but larger edge artifacts and larger differences of the central values from the corresponding ideal values again defined by the intersection of the sphere and the imaging plane. The profiles of sound speed and attenuation slope in the reconstruction for a 5-mm offset have still larger edge artifacts and differences of the central values from the corresponding ideal values defined by the intersection of the sphere and the imaging plane. Overall, the results show that profiles of sound speed can reasonably approximate ideal profiles of sound speed when the portion of the sphere in the elevation-focused beam is not constant, but profiles of attenuation slope can be larger than ideal profiles when the center of the sphere is offset from the imaging plane and appreciable scattering is not collected by the receive aperture.

Features of sound speed and attenuation slope profiles in Fig. 8 for a 4-mm diameter sphere with a center at different offsets from the imaging plane merit comment because the features are further illustrative. The profiles of sound speed in reconstructions of the 4-mm diameter sphere that was either mostly encompassed by, partially encompassed by, or at the edge of the elevation-focused beam are more smoothed and have smaller values than ideal values defined by a central cross section of the sphere. This is a result of partial volume effects. The profiles of attenuation slope have larger values than similarly defined ideal profiles. This is a result of energy scattered outside the area of the elevation-focused receive aperture dominating partial volume effects.

Insight about attenuation artifacts can be gained by expressing the three-dimensional wave equation as a wave equation in two dimensions plus a term in the third dimension. To begin, consider a pressure p that satisfies the wave equation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)p(\mathbf{x},z)+k^2p(\mathbf{x},z)=k^2\eta(\mathbf{x},z)p(\mathbf{x},z). \tag{74}$$

Integrating both sides of this equation in the elevation or z direction with weighting $w(z)$ gives

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz+\int_{-\infty}^{+\infty}\frac{\partial^2 p(\mathbf{x},z)}{\partial z^2}w(z)dz$$
$$+k^2\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz=k^2\int_{-\infty}^{+\infty}\eta(\mathbf{x},z)p(\mathbf{x},z)w(z)dz. \tag{75}$$

If $p(\mathbf{x},z)$ slowly varies in the z direction, then the second integral on the left side may be neglected and Eq. (75) can be written as

$$(\Delta_{x,y}+k^2)\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz$$
$$=k^2\int_{-\infty}^{+\infty}\eta(\mathbf{x},z)p(\mathbf{x},z)w(z)dz \tag{76}$$

in which

$$\Delta_{x,y}(\cdot)=\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)(\cdot). \tag{77}$$

Finally, to complete the reduction to two dimensions, the integral on the right side is written as

$$\int_{-\infty}^{+\infty}\eta(\mathbf{x},z)p(\mathbf{x},z)w(z)dz=\bar{\eta}(\mathbf{x})\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz, \tag{78}$$

where

$$\bar{\eta}(\mathbf{x})=\int_{-\infty}^{+\infty}\eta(\mathbf{x},z)p(\mathbf{x},z)w(z)dz\bigg/\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz. \tag{79}$$

If $\eta(\mathbf{x},z)=\eta(\mathbf{x})$, i.e., the medium variations are independent of the elevation dimension z, then $\eta(\mathbf{x})=\eta(\mathbf{x})$ regardless of the pressure characteristics. However, if $\eta(\mathbf{x},z)$ varies in the z direction, then values of $\bar{\eta}(\mathbf{x})$ differ when the right side of Eq. (79) is evaluated for different pressures. To ensure a consistent value of $\bar{\eta}(\mathbf{x})$ in two-dimensional measurements of scattering, the pressure that is incident pressure under the Born approximation should have the same elevation profile throughout the range of z values over which $w(z)$ is appreciable. Then, Eq. (77) becomes

$$(\Delta_{x,y}+k^2)\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz=k^2\bar{\eta}(\mathbf{x})\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz, \tag{80}$$

i.e., a two-dimensional reduced wave equation for the two-dimensional pressure that is a weighted z-axis projection of the three-dimensional pressure. The weights $w(z)$ in the calculations reported here were chosen to be the complex amplitudes of the lens that produced the elevation focus in the ring transducer system because this choice ensures that the calculations correspond to scattering measurements made using the system.

Worthy of special comment is that the term

$$\int_{-\infty}^{+\infty}\frac{\partial^2 p(\mathbf{x},z)}{\partial z^2}w(z)dz \tag{81}$$

was assumed negligible in the above analysis. To study this term, the pressure $p(\mathbf{x},z)$ is factored into the product $p(\mathbf{x},z)=\psi(\mathbf{x})\zeta(z)$ so that the term in Eq. (81) can be written as

$$\int_{-\infty}^{+\infty}\frac{\partial^2 p(\mathbf{x},z)}{\partial z^2}w(z)dz=\Delta\bar{\eta}k^2\int_{-\infty}^{+\infty}p(\mathbf{x},z)w(z)dz \tag{82}$$

in which

TABLE V. Power loss calculations and power loss ratios for various lossless and lossy cylinders and spheres illuminated by a plane wave. In the case of the cylinders, the plane wave travels normal to the axis of the cylinder.

	Radius (mm)	P_{loss} (nW)	P_{scat} (nW)	P_{abs} (nW)	$(P_{\text{abs}})/P_{\text{loss}}$ (%)
Lossless cylinder	10.00	10.2732	10.2732	0.8882×10^{-14}	0.864×10^{-13}
	9.54	8.1582	8.1582	0.3375×10^{-12}	-0.414×10^{-12}
	8.66	6.0006	6.0006	-0.3642×10^{-12}	-0.606×10^{-12}
Lossless sphere	10.00	0.1550	0.1550	0.4331×10^{-9}	2.79×10^{-7}
	9.54	0.1365	0.1365	0.2938×10^{-9}	2.15×10^{-7}
	8.66	0.1200	0.1200	0.3052×10^{-9}	2.54×10^{-7}
Lossy cylinder	10.00	10.8788	8.7190	2.1598	19.9
	9.54	9.0971	7.1125	1.9845	21.8
	8.66	7.1201	5.4543	1.6658	23.4
Lossy sphere	10.00	0.1661	0.1377	0.0284	17.1
	6.54	0.1478	0.1229	0.0249	16.8
	8.66	0.1279	0.0190	0.0189	14.8

$$\Delta\bar{\eta} = \int_{-\infty}^{+\infty} \frac{\partial^2\zeta(z)}{\partial z^2} w(z)dz/k^2 \int_{-\infty}^{+\infty} \zeta(z)w(z)dz \tag{83}$$

and $\Delta\bar{\eta}$ represents a scattering potential variation when the pressure is not constant in elevation. A wave field with variations in elevation implies some of the wave is radiating out of the image plane and is not collected by the receive aperture. The resulting loss is manifested as additional attenuation in the reconstructions.

To quantify the observed loss in the reconstructions, a version of the so-called extinction theorem or optical theorem[b,20,21] is used. A general expression of this theorem for a monochromatic acoustic pressure wave is

$$P_{\text{loss}} = \frac{k^2}{2\rho_0\omega} \operatorname{Im}\left[\int_V \eta(\mathbf{x})p_i^*(\mathbf{x})p(\mathbf{x})d^3\mathbf{x}\right], \tag{84}$$

where P_{loss} is the total power lost and p_i and p are the incident and total pressures, respectively. For plane-wave illumination, the total lost power is proportional to the imaginary part of the through-transmission normalized far-field pressure, i.e.,

$$P_{\text{loss}} = \frac{1}{2\rho_0\omega} \operatorname{Im}[A(\mathbf{u}_i,\mathbf{u}_i)], \tag{85}$$

where A is given in three dimensions by Eq. (48) with the angular arguments represented using the unit vector $\mathbf{u}_i$ that is in the direction of the incident wave. Equations (84) and (85) can also be normalized to express the power loss in terms of a cross section by including a scale factor that results in units of area but are written here without a normalizing factor so the units are those of power. For the object sizes, the loss characteristics, and the wavelengths considered in this report, the various lost powers are on the order of nanowatts.

The power lost due to scattering alone can be calculated using the surface integral

$$P_{\text{scat}} = -\frac{1}{2\rho_0\omega} \operatorname{Im}\left[\int_{\partial V} p_s^*(\mathbf{x})\,\nabla\, p_s(\mathbf{x})\cdot\mathbf{n}(\mathbf{x})d\sigma(\mathbf{x})\right] \tag{86}$$

in which $p_s(\mathbf{x})$ is the scattered pressure field and $\mathbf{n}(\mathbf{x})$ is a unit vector normal to the surface of the integration. This expression for the scattered power simplifies when the incident illumination is a plane wave and the integration surface is in the far field. In two dimensions, the simplified expression is

$$P_{\text{scat}}^{\text{2d}} = \frac{1}{16\pi kc\rho_0} \int |A(\mathbf{u},\mathbf{u}_j)|^2 d\varsigma(\mathbf{u}), \tag{87}$$

where $\mathbf{u}$ is a unit vector in the direction of the observation line ς. In three dimensions, the corresponding expression is

$$P_{\text{scat}}^{\text{3d}} = \frac{1}{32\pi^2 c\rho_0} \int |A(\mathbf{u},\mathbf{u}_i)|^2 d\sigma(\mathbf{u}), \tag{88}$$

where $\mathbf{u}$ is a unit vector in the direction of the observation surface σ. The total power lost due to absorption can then be found from the total lost power and the total scattered power by using

$$P_{\text{abs}} = P_{\text{loss}} - P_{\text{scat}} \tag{89}$$

in which P_{abs} and P_{scat} are the power lost due to absorption and scattering, respectively.

To give confidence that the calculations of lost power by using the extinction theorem and by using the surface integrals are correctly performed in both two and three dimensions, the lost power was computed for a lossless cylinder and a lossless sphere. The calculations were performed at 2.5 MHz and used the values of the background parameters in Table I and the values of object properties in Table III except that the attenuation slope was 0 dB/(cm MHz). The results of the calculations are shown in Table V. For comparison, the corresponding calculations for a lossy cylinder and sphere with an attenuation slope of 0.462 dB/(cm MHz) are also shown in Table V. The very small values of absorbed power for lossless objects are attributed to numerical

TABLE VI. Power scattered into a Gaussian-weighted band of an enclosing spherical surface and into the remainder of the enclosing surface by a lossy sphere centered in illumination by an elevation-focused beam.

Offset (mm)	P_{loss} (nW)	P_{scat} (nW)	$P^{\text{in}}_{\text{scat}}$ (nW)	$P^{\text{out}}_{\text{scat}}$ (nW)	$P^{\text{out}}_{\text{scat}}/P_{\text{loss}}$ (%)	$P^{\text{2D}}_{\text{loss}}$ (nW)
0	0.1574	0.1256	0.0824	0.0432	27.5	10.5035
3	0.1342	0.1052	0.0621	0.0431	32.1	8.9585
5	0.1152	0.0915	0.0425	0.0490	42.5	7.6785

TABLE VII. Power loss calculations and power loss ratios for a lossy and lossless 20-mm diameter sphere with elevation-focused illumination and reception having various offsets from the center of the sphere.

	Offset (mm)	$P^{\text{2D}}_{\text{loss}}$ (nW)	$P^{\text{2D}}_{\text{scat}}$ (nW)	$P^{\text{2D}}_{\text{abs}}$ (nW)	$P^{\text{2D}}_{\text{abs}}/P^{\text{2D}}_{\text{loss}}$ (%)
	0	9.8270	9.6081	0.2188	2.2
Lossless sphere	3	8.0430	6.3658	1.6772	20.9
	5	6.7677	3.5179	3.2498	48.0
	0	10.5035	8.2117	2.2918	21.8
Lossy sphere	3	8.9585	5.7281	3.2304	36.1
	5	7.6785	3.5630	4.1155	53.6

effects. The values of absorbed power for the lossy objects are, however, appreciable. The ratios of absorbed power to total lost power that appear in the last column of these tables are noteworthy because these ratios are dimensionless quantifications of the rates at which energy is absorbed. These ratios are dominated by the intrinsic attenuation of the scattering object although they are also influenced by the geometry and size of the scattering object as shown by the variations in this column of table entries. Since these ratios vanish for lossless objects, they cannot be used to assess extrinsic attenuation.

Two calculations were carried out to quantify loss due to extrinsic attenuation. In the first calculation, the total power lost from an elevation-focused beam illuminating a lossless sphere was found by direct application of the two-dimensional form of the extinction theorem. In the second calculation, the power losses due to scattering and absorption were found by separating the surface integral expression in Eq. (86) into an aperture surrounding the equatorial plane of the sphere and the complementary regions above and below the aperture. The power scattered into the aperture was weighted by a Gaussian that assumed a maximum value of 1 in the center of the aperture and declined with elevation above and below the aperture. The result of this calculation is a value $P^{\text{in}}_{\text{scat}}$ that measures the scattered energy in the aperture of the already noted ring transducer system. The absorbed power was then computed by the difference expression in Eq. (89). These computations were performed for a 20-mm diameter sphere with the same parameters as used in the reconstruction except that a lossless sphere was considered to eliminate the influence of absorption. In the calculations given by Eq. (86), a focused transmit beam and a 75-mm spherical collection surface were used to obtain power losses for the geometry of the measurements reported in Ref. 14.

The results of the calculations described in the above paragraph are given in Table VI. The ratios in the last two columns of the table are dimensionless quantities that are easiest to interpret. The ratios in the second-to-last column represent fractions of energy lost to measurement according to the surface integral calculation. The ratios in the last column of the table are determined from the two-dimensional extinction theorem. These show substantially smaller losses. This discrepancy is attributed to receive focusing that is included in two-dimensional extinction theorem calculations but is absent from the surface integral formulation. Receive focusing causes a reduction of both P_{loss} and P_{scat}, but the portion of power loss attributed to absorption is less because P_{loss} is reduced more.

The forgoing discussion suggests that power loss ratios obtained from the two-dimensional extinction theorem are mostly determined by extrinsic attenuation that appears in the two-dimensional reconstructions. Table VII gives values of this ratio for cross sections of a lossless and lossy 20-mm diameter sphere at offsets of 0, 3, and 5 mm. The loss fractions for the cross sections with 3 and 5 mm offsets are substantially greater than those for the cross section with no offset and are comparable for lossless and lossy media. This shows that the extrinsic attenuation in the two-dimensional reconstructions of offset cross sections can be greater than the intrinsic attenuation. The power distributions are visualized as grayscale variations over the surface of the sphere of observation in Fig. 9 for a lossless sphere.

Power loss fractions can also be computed for the two-dimensional cylindrical models used during the reconstruction process to adjust model cylinders that are unable to mimic the power loss fractions for offset cross sections because the abrupt boundary of the cylinder produces significant in-plane scattering across a wider range of scattering angles than the cross section of an object with a boundary that varies in the width of the elevation focus. The effective boundary of such varying cross sections is distributed over a range of radii and results in less in-plane scattering. This suggests that scattering from an elevation-varying cross section might be better modeled by a cylinder with a graduated boundary as may be produced by adding a matching layer to the perimeter of the cylinder. Then, comparison of power loss ratios might provide a numerical guide for adjusting these boundary conditions.

Cross-sectional imaging techniques that use two-dimensional methods to reconstruct sections of three-dimensional objects are more efficient and require less comprehensive measurement systems than reconstructions that are fully three-dimensional. However, significant artifacts and distortion can appear in reconstructions formed in this manner if three-dimensional effects are not taken into account. The presented simulations of three-dimensional scattering measurements using elevation-focused transmit-receive beams have explored these corrupting influences and have illustrated ways that these influences can be compensated. The simulations show that two-dimensional inverse scattering methods can be used successfully to reconstruct cross sections of three-dimensional scattering objects, provided that the residuals used in the reconstruction are made

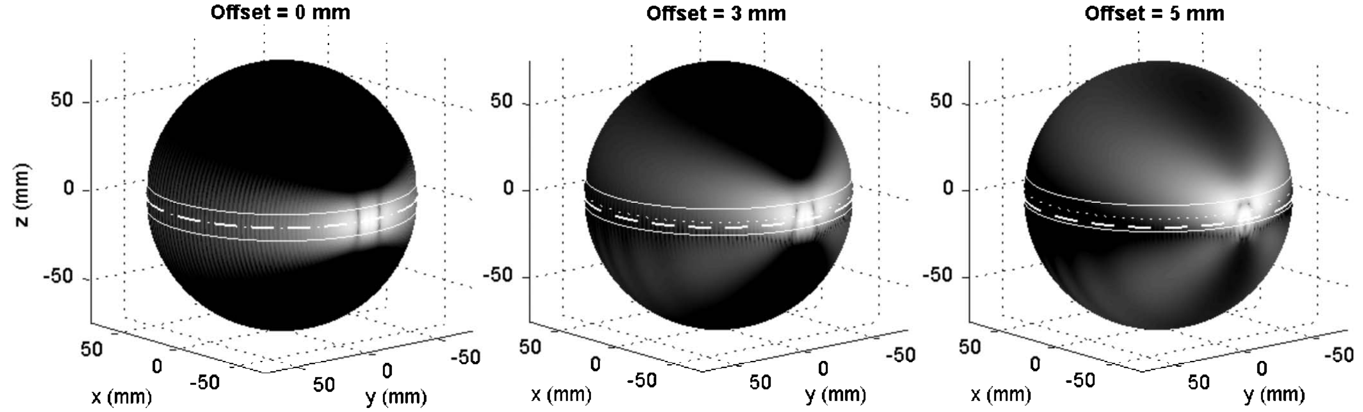


FIG. 9. Intensity of scattering by a 20-mm diameter lossless sphere illuminated with an elevation-focused transmit beam offset of 0, 3, or 5 mm below the center of the sphere. The solid lines show the effective width of the receive aperture. The radius of the surface on which the intensity is shown is 75 mm. The dotted line denotes the center of the receive aperture. The dashed line is the intersection of the imaging plane with the sphere. The grayscale is logarithmic and spans a range of 60 dB.

sufficiently small by use of models that compensate for energy loss in the direction that is orthogonal to the plane of the cross section.

VI. CONCLUSION

A detailed analysis of using a two-dimensional method of image reconstruction with transmit-receive elevation-focused beams to image a cross section of a three-dimensional object has been performed. In the analysis, scattering in three dimensions is written as the sum of two-dimensional scattering that results from each plane wave or spatial-frequency in the elevation dimension. An angular spectrum form for the illumination pattern and a corresponding form for reception sensitivity lead to expressions for measured scattered pressure when the illumination pattern and receive sensitivity are focused in elevation. These expressions are used to calculate scattering by a cylinder and by a sphere with focused illumination and reception. The results show that two-dimensional reconstructions of scattering objects using elevation-focused beams can give accurate results when the scattering object is nearly the same throughout the width of the elevation focus. In situations where the three-dimensional scattering is not entirely captured by the receive aperture, two-dimensional reconstructions of sound speed are relatively unaffected, but corresponding reconstructions of attenuation slope can have elevated values at the object boundary. These artifacts can be compensated by endowing a background model with additional loss that accounts for unmeasured scattering.

ACKNOWLEDGMENTS

Jason C. Tillett is thanked for his suggestions, helpful comments, and useful discussions throughout the duration of this study. Makoto Tabei and Li Sun are thanked for their contributions during the initial development of the forward-scattering and equal-angle code. Michele M. Foster is thanked for help with the preparation of figures. The participation of Jing Jin and Gheorghe Salahura in reviews of this work as it progressed is acknowledged with appreciation. This research was funded by NIH Grant Nos. HL 50855, CA 74050, and EB 00280, DARPA Grant No. N00014-96-0749, and the University of Rochester Diagnostic Ultrasound Research Laboratory Industrial Associates.

APPENDIX: FOURIER TRANSFORM IN ELEVATION OF THE THREE-DIMENSIONAL OUT-GOING FREE-SPACE GREEN'S FUNCTION

The Fourier transform in elevation of the three-dimensional out-going free-space Green's function is given by

$$\begin{aligned} G_0^{(3D)}(x,\nu,f) &= \int_{-\infty}^{+\infty} G_0^{(3D)}(x,z,f)e^{-j2\pi\nu z}dz \\ &= \int_{-\infty}^{+\infty} \frac{e^{j(2\pi f/c)\sqrt{r^2+z^2}}}{4\pi\sqrt{r^2+z^2}}e^{-j2\pi\nu z}dz. \end{aligned} \tag{A1}$$

Since $G_0^{(3D)}(\mathbf{x},z,f)=G_0^{(3D)}(\mathbf{x},-z,f)$, the integration can be reduced to a half line and written as

$$G_0^{(3D)}(x,\nu,f) = \frac{1}{2\pi}\int_0^{\infty} \frac{e^{j(2\pi f/c)\sqrt{r^2+z^2}}}{\sqrt{r^2+z^2}}\cos(2\pi\nu z)dz. \tag{A2}$$

This integral can be evaluated using Formulas 3.876.1 and 3.876.2 in Ref. 22 with $p=2\pi f/c$, $a=r$, and $b=2\pi\nu$. The result is

$$G_0^{(3D)}(x,\nu,f) = \begin{cases} \dfrac{j}{4}H_0(2\pi\sqrt{(f/c)^2+\nu^2 r}), & f/c > \nu \\ \dfrac{1}{2\pi}K_0(2\pi\sqrt{\nu^2-(f/c)^2 r}), & f/c < \nu. \end{cases} \tag{A3}$$

Since $K_0(\rho)=(\pi j/2)H_0(j\rho)$, the H_0 term on the right side can be used for all values of ν. This allows the Fourier transform in elevation of the three-dimensional free-space Green's function $G_0^{(3D)}(\mathbf{x},\nu,f)$ to be expressed in terms of the two-dimensional free-space Green's function $G_0^{(3D)}(\mathbf{x},f_e)$ as

$$G_0^{(3D)}(x,\nu,f) = G_0^{(2D)}(x,f_e) \quad \text{(A4)}$$

in which f_e is the effective temporal frequency given by $\sqrt{f^2 - c^2\nu^2}$.

[a] In Eq. (4) and in following analysis, the notation uses the quantum mechanics convention in which a symbol represents a conceptual object and the argument indicates the domain in which an object is evaluated. Thus, $p_n(\mathbf{x},z,f)$ is the scattered pressure in the space domain as a function of the spatial-position vector $(\mathbf{x},z)$ and $p_n(\mathbf{x},\nu,f)$ is the one-dimensional Fourier transform as a function of the spatial-frequency ν.

[b] For a history of the optical theorem, see R. G. Newton, "The optical theorem and beyond," *Am. J. Phys.*, **44**, 639–642 (1976).

[1] S. L. Hagen-Ansert, *Textbook of Diagnostic Ultrasonography*, 6th ed. (C. V. Mosby, St. Louis, 2006), Vol. **1**.

[2] C. Rumack, S. Wilson, J. W. Charboneau, and J. Johnson, *Diagnostic Ultrasound*, 3rd ed. (C. V. Mosby, St. Louis, 2004), p. 2004.

[3] R. C. Waag and R. J. Fedewa, "A ring transducer system for medical ultrasound research," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1707–1718 (2006).

[4] R. Mueller, M. Kaveh, and G. Wade, "Reconstructive tomography and applications to ultrasonics," Proc. IEEE **67**, 567–587 (1979).

[5] N. Duric, P. Littrup, A. Babkin, D. Chambers, S. Azevedo, R. Pevzner, M. Tokarev, E. Holsapple, O. Rama, and R. Duncan, "Development of ultrasound tomography for breast imaging: Technical assessment," Med. Phys. **32**, 1375–1386 (2005).

[6] A. J. Devaney, "A filtered back-propagation algorithm for diffraction tomography," Ultrason. Imaging **4**, 336–350 (1982).

[7] M. Kaveh, M. Soumekh, and J. F. Greenleaf, "Signal processing for diffraction tomography," IEEE Trans. Sonics Ultrason. **31**, 230–239 (1984).

[8] W. C. Chew and Y. M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," IEEE Trans. Med. Imaging **9**, 218–225 (1990).

[9] C. Lu, J. Lin, W. Chew, and G. Otto, "Image reconstruction with acoustic measurement using distorted Born iteration method," Ultrason. Imaging **18**, 140–156 (1996).

[10] S. A. Johnson and M. L. Tracy, "Inverse scattering solutions by a sinc basis, multiple source, moment method—Part I: Theory," Ultrason. Imaging **5**, 361–375 (1983).

[11] S. A. Johnson, D. T. Borup, J. W. Wiskin, F. Natterer, F. Wubeling, Y. Zhang, and S. C. Olsen, "Apparatus and method for imaging with wavefields using inverse scattering techniques," U.S. Patent No. 6,005,916 (21 December 1999).

[12] T. D. Mast, A. I. Nachman, and R. C. Waag, "Focusing and imaging using eigenfunctions of the scattering operator," J. Acoust. Soc. Am. **102**, 715–725 (1997).

[13] F. Lin, A. I. Nachman, and R. C. Waag, "Quantitative imaging using a time-domain eigenfunction method," J. Acoust. Soc. Am. **108**, 899–912 (2000).

[14] R. C. Waag, F. Lin, T. K. Varslot, and J. P. Astheimer, "An eigenfunction method for reconstruction of large-scale and high-contrast objects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1316–1332 (2007).

[15] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. (Dover, New York, 1972), p. 364, Eqs. (9.2.1) and (9.2.2).

[16] A. Lowan, P. Morse, H. Feshbach, and M. Lax, "Scattering and radiation from circular cylinders and spheres—Tables of amplitudes and phase angles," Report No. 62.1R, U.S. Navy Department, Office of Research and Inventions, Arlington, VA, 1945.

[17] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8.

[18] K. Chadan, D. Colton, L. Paivarinta, and W. Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems* (Society for Industrial and Applied Mathematics, Philadelphia, 1997), p. 40.

[19] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. (Dover, New York, 1972), p. 437, Eq. (10.1.1).

[20] H. C. van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957), p. 39.

[21] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light*, 7th ed. (Cambridge University Press, Cambridge, UK, 2000), Chap. XIII, p. 720.

[22] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), p. 472.

Forward projection of transient sound pressure fields radiated by impacted plates using numerical Laplace transform

Jean-François Blais and Annie Ross[a)]
Department of Mechanical Engineering, CREPEC, École Polytechnique de Montréal, CP 6079 Station Centre-ville, Montréal, Québec H3C 3A7, Canada

(Received 16 June 2008; revised 18 February 2009; accepted 18 February 2009)

Forward propagation of the transient sound pressure radiated by an impacted plate is presented. It is shown that direct and inverse time domain discrete Fourier transforms, involved in Fourier transform based near-field acoustical holography (NAH), lead to aliasing errors in the reconstructed time signals. Adding trailing zeros to the initial time signals is an inefficient way to reduce time aliasing errors. Hence, the numerical Laplace transform is introduced and a Fourier transform based transient NAH (TNAH) approach is formulated. An error measure is introduced to compare both NAH and TNAH with respect to the propagation distance and the location of the observation point in the projection plane. The percentage of error with TNAH is reduced by more than a factor of 10 without adding trailing zeros to the initial signals. Simulation results are validated experimentally using a free Plexiglas plate impacted at its center.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097765]

PACS number(s): 43.60.Sx, 43.20.Px, 43.40.Kd, 43.40.Dx [EJS] Pages: 3120–3128

I. INTRODUCTION

Sounds radiated by impacted structures are an important source of noise in industrial applications such as riveting and hammering and can cause hearing impairment. In these sound fields, transient phenomena such as the radiation due to the reflection of flexural waves in the structure (at edges, holes, or any local changes in the structure's mechanical impedance) have time and space properties which are of interest for the development of impact noise control devices.[1] Accordingly, a time and space visualization technique such as near-field acoustical holography (NAH)[2,3] could be used as a tool to gain a better understanding of these transient acoustic phenomena and thus lead to better control devices.

Forward propagation of transient sound fields has been investigated by de La Rochefoucauld *et al.*[4,5] using four formulations of time domain holography. They simulated the sound pressure radiated by a baffled piston and carried out a parametric study with respect to the bandwidth of the time signals, geometric characteristics of the source and of the measurement aperture, and the propagation distance. They also used the four formulations to propagate experimental sound pressure fields radiated by loudspeakers and impacted plates. Fourier transform based NAH (i.e., the standard NAH formulation in the frequency and wave vector domains) leads to better results than time/space or frequency/space formulations. It has also been applied successfully by Clement *et al.*[6] to propagate the sound pressure radiated by a piezoelectric transducer.

The noise produced by impacted plates is of interest because plates are simple structures encountered in industry. However, by examining closely the results of de La Rochefoucauld *et al.* relative to plate radiation, one could note that propagated signals are non-causal, i.e., sound pressure is not zero at time $t=0$ corresponding to the beginning of impact. This can be attributed to the use of direct and inverse time domain discrete Fourier transforms (DFTs) involved in the standard NAH formulation, leading to time aliasing errors in the propagated signals if the initial data are not properly padded with zeros.[7] Aliasing errors are negligible for sound pressure signals much shorter than the observation time window. The sound pressure radiated by the loudspeakers or the transducer mentioned above are examples of very short signals. On the other hand, time aliasing errors become more significant both in amplitude and duration when initial time signals are truncated, which is the case of the sound pressure radiated by an impacted plate.

Wu *et al.*[8] derived a transient formulation of the Helmoltz equation least-squares (HELS)[9] method to study transient acoustic sources. This approach makes use of the Laplace transform, which is a common tool for transient problems. In the present paper, a Fourier transform based transient NAH (TNAH) formulation, also using the Laplace transform, is presented. While the inverse Laplace transform in the transient HELS formulation is calculated using the residue theory, the TNAH approach is based on direct and inverse time domain numerical Laplace transforms (NLTs). Consequently, the k-space Green's function involved in the Fourier transform based NAH is reformulated. Standard DFT algorithms for time and space transformations can still be used, without requiring time domain zero padding.

In Sec. II, general equations for the transient radiation of an impacted plate and the basics of the standard NAH formulation are summarized. Section III presents the time aliasing phenomenon and how it can affect forward propagation. The sound pressure radiated by an impacted plate is simulated in a measurement plane. These data are propagated using the standard NAH formulation to illustrate time aliasing errors. The use of time domain zero padding as a way to

[a)] Author to whom correspondence should be addressed. Electronic mail: annie.ross@polymtl.ca

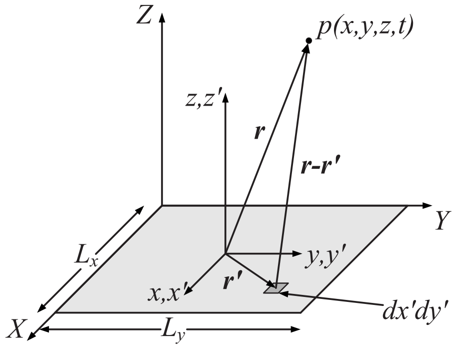


FIG. 1. Rectangular plate and coordinate systems.

reduce aliasing errors is also discussed. The NLT and its application to NAH are introduced in Sec. IV. The Gibbs phenomenon in the time domain, which is not an issue in standard NAH, must be considered in the TNAH formulation and is discussed in this section. In Sec. V, an error measure is introduced to compare both NAH and TNAH with respect to the propagation distance and the location of the observation point in the projection plane. Finally, in Sec. VI, simulation results are validated experimentally with an impacted, free Plexiglas plate.

II. SOUND RADIATION AND PROPAGATION

A. Transient sound radiation of an impacted plate

Let us consider the rectangular plate of dimensions $L_x \times L_y$ and thickness h shown in Fig. 1. The Cartesian coordinates (X,Y,Z) originate at the corner of the plate whereas coordinates (x,y,z) are located at its center.

Suppose that a sphere strikes the plate at point (X_0,Y_0) at time $t=0$ s. The amplitude and the duration of the impact force $F(t)$ are functions of the sphere radius (r_s), the impact velocity (v_s), and other physical and mechanical properties of both the plate (subscript p) and the sphere (subscript s) such as Young's modulus (E), mass density (ρ), and Poisson's ratio (ν).[10,11]

The transverse displacement of an undamped simply supported plate due to an impact is given in terms of a time convolution $(*)$ by[11]

$$w(X,Y,t) = \frac{1}{\rho_p h}\sum_m \sum_n \frac{\Phi_{mn}(X_0,Y_0)\Phi_{mn}(X,Y)}{\omega_{mn}} \times \{F(t) * \sin(\omega_{mn}t)\}, \qquad (1)$$

where Φ_{mn} are the mode shapes and ω_{mn} are the natural frequencies.[3] The transverse acceleration is obtained from the double time derivation of Eq. (1) and is given by

$$\ddot{w}(X,Y,t) = \frac{1}{\rho_p h}\sum_m \sum_n \Phi_{mn}(X_0,Y_0)\Phi_{mn}(X,Y) \times \{F(t) * [-\omega_{mn}\sin(\omega_{mn}t) + \delta(t)]\}, \qquad (2)$$

where $\delta(t)$ is a Dirac impulse.

For a plate in an infinite baffle, the free field radiated sound pressure is given by the Rayleigh surface integral in the Cartesian coordinates (x,y,z) shown in Fig. 1, i.e.,

$$p(x,y,z,t) = \frac{\rho_0}{2\pi}\int\int \ddot{w}\left(x',y',t-\frac{R}{c}\right)\frac{dx'dy'}{R}, \qquad (3)$$

with

$$R = |\boldsymbol{r} - \boldsymbol{r}'|,$$

where c and ρ_0 are the sound velocity in the ambient medium and the density of the medium, respectively. Vectors $\boldsymbol{r}$ and $\boldsymbol{r}'$ shown in Fig. 1 represent the position of a sound pressure measurement point and that of an infinitesimal surface element $dx'dy'$ on the plate.

B. NAH

Since the basics of NAH are widely described in the literature,[3] only a brief summary is presented here. Let us first introduce the Fourier transform pairs applied to acoustic pressure

$$P(k_x,k_y,z,\omega) = \int_0^{+\infty}\int\int_{-\infty}^{+\infty} p(x,y,z,t) \times e^{-i(k_x x + k_y y - \omega t)} dxdydt \qquad (4)$$

and

$$p(x,y,z,t) = \frac{1}{8\pi^3}\int\int\int_{-\infty}^{+\infty} P(k_x,k_y,z,\omega) \times e^{i(k_x x + k_y y - \omega t)} dk_x dk_y d\omega, \qquad (5)$$

where k_x and k_y are the trace wavenumbers in the x and y directions.

Considering a measurement plane (hologram) located at $z=z_0$, propagation of the sound field to a plane $z>z_0$ can be written as[3]

$$P(k_x,k_y,z,\omega) = P(k_x,k_y,z_0,\omega)G_{pp}(k_x,k_y,d,\omega), \qquad (6)$$

where G_{pp} is the k-space Green's function in the frequency domain, and $d=z-z_0$. For short propagation distances d in both transient cases[5] and stationary cases,[12] it is preferable to sample G_{pp} directly in the k-space domain, i.e.,

$$G_{pp}(k_x,k_y,d,\omega) = e^{ik_z d}, \qquad (7)$$

where $k_z^2 = k^2 - k_x^2 - k_y^2$ and $k=\omega/c$.

Tukey windows can be applied in both space and time domains to avoid spectral leakage in NAH calculations. The Tukey window is defined as[3]

$$\Gamma(\xi) = \begin{cases} 1 & |\xi| < \xi_m - \xi_w \\ \frac{1}{2}\left[1 - \cos\left(\frac{\pi(|\xi| - \xi_m)}{\xi_w}\right)\right] & \xi_m - \xi_w \le |\xi| \le \xi_m, \end{cases} \qquad (8)$$

where ξ_m is the maximum value taken by ξ (corresponding to x, y, or t), and ξ_w is the width of the tapered rim.

III. TIME ALIASING IN NAH

Time aliasing appears when a time function [Fig. 2(a)] is truncated at $t=T$ and is subjected to a delay τ_0 through a transfer function applied in the frequency domain. The use of an inverse time DFT replicates the resulting time signal with

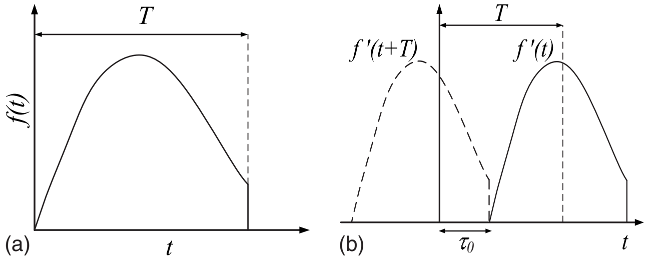


FIG. 2. Application of a delay τ_0 to a function using a DFT. (a) Initial function. (b) Result of delay.

a period T, causing an error at time $t<\tau_0$, as shown in Fig. 2(b). In NAH, the sound pressure signal at each point (x_0,y_0,z_0) of the hologram is subjected to a different attenuation and time delay caused by the k-space Green's function in Eq. (7). As a result, individual time aliasing errors can overlap the actual signal at a given observation point (x,y,z). The amplitude of the sound pressure at the observation point and that of the time aliasing error due to NAH propagation can be of the same order of magnitude. The duration of aliasing errors mainly depends on the dimensions of the hologram and on the position of the observation point in the propagation plane.

A. Illustration of time aliasing in NAH

To illustrate the importance of time aliasing errors, forward propagation of a simulated hologram was carried out with Fourier transform based NAH. The transverse acceleration of the plate was simulated considering a central impact force $(X_0=L_x/2, Y_0=L_y/2)$. Impact parameters are listed in Table I. Summations in Eq. (2) were computed over the 40 first modes in each direction $(m,n\leq 40)$. Experimental validations were carried out using an instrumented hammer with a spherical metal tip.[13] In Table I, m_i is the total mass of the impactor. The impact force (– – –) and the initial transverse acceleration of the plate at the impact point (——) are shown in Fig. 3.

The acoustical hologram was simulated in the plane $z_0=5$ cm using a double trapezoidal integration scheme to evaluate Eq. (3). Time and space properties of the simulated measurement plane including signal duration (T), time step (Δt), number of measurements (N_x,N_y), total number of points including the spatial zero padding $(\bar{N}_x,\bar{N}_y)$, and the distance between these points $(\Delta x,\Delta y)$ are also summarized in Table I. The hologram was tapered over a duration t_w and over widths x_w and y_w using Tukey windows [Eq. (8)].

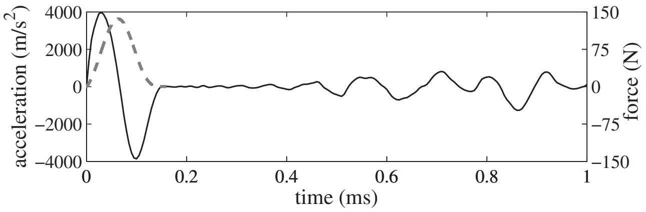


FIG. 3. Time signals of the transverse acceleration of the plate at the impact point (——) and the impact force (– – –).

The initial transient radiation of the impacted plate is illustrated in Fig. 4 for two different locations at $z=z_0$. In Fig. 4(a), the sound pressure signal is calculated on the impact axis. The initial sound pressure peak (from 0.15 to 0.25 ms) is clearly visible and is followed by ringing. At 28 cm from the impact axis [Fig. 4(b)], sound propagation delay and apparent dispersion of the initial sound pressure peak[13] can be observed. At both locations, because the initial sound pressure amplitude is null for a fair amount of time, any disturbance due to time aliasing would be observable in NAH results.

Let us now consider the propagation of the sound field simulated at $z_0=5$ cm to the plane $z=35$ cm. Initial time signals were increased to a duration $\bar{T}=2T$ by adding trailing zeros. This way, time aliasing causes the zeros to be continued at the beginning of the propagated signal, thereby reducing time aliasing error. Moreover, adding a sufficient number of zeros allows to separate the time-aliased part from the actual signal.

Figure 5(a) represents the signal simulated at $z=35$ cm over a duration $2T$ along the impact axis. The amplitude of the initial pressure peak is reduced by about 12.9 dB compared to that at $z_0=5$ cm. The peak is also delayed by approximately 0.9 ms, which corresponds to the propagation time d/c. The ringing wave form is also observed. Figure 5(b) shows the propagation over a distance $d=30$ cm of the zero padded time signals simulated at $z_0=5$ cm. In Fig. 5(c), propagation was obtained without padding in the time domain, which explains why the resulting signal lasts half the duration of that in Fig. 5(b). To facilitate the observations and the discussion, the graphs in Fig. 5 were segmented in three zones, numbered from I to III. Borderlines are located at $t=T=2.56$ ms and $t=T+d/c-t_w=3.2$ ms. Zones II and III correspond to the second half of the signal, that is, to the zero padding on the original signals.

First, it can be observed that the time signal obtained using a zero pad [Fig. 5(b)] is very similar to the simulated

TABLE I. Properties of the numerical simulation.

Aluminum plate		Steel sphere		Ambient air		Hologram	
$L_x\times L_y$	60.9×91.4 cm^2	r_s	3.2 mm	ρ_0	1.29 kg/m^3	T	2.56 ms
h	4.8 mm	m_i	26.8 g	c	343 m/s	Δt	5 μs
ρ_p	2700 kg/m^3	ρ_s	7800 kg/m^3			t_w	0.25 ms
E_p	71 GPa	E_s	200 GPa			$N_x\times N_y$	127×127
ν_p	0.33	ν_s	0.28			$\bar{N}_x\times\bar{N}_y$	147×147
		v_s	0.23 m/s			$\Delta x,\Delta y$	2 cm
						x_w, y_w	20 cm

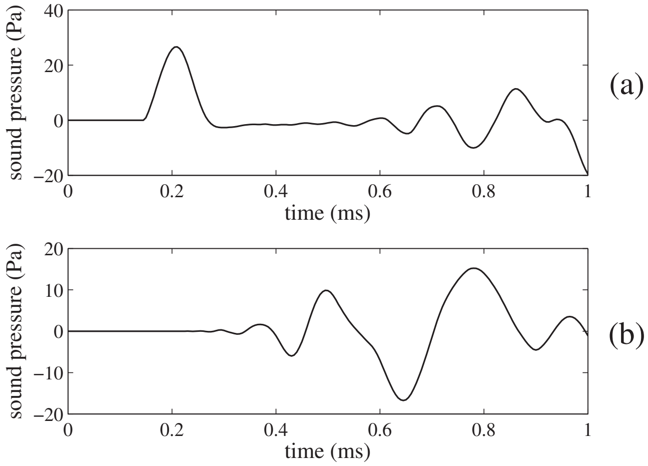


FIG. 4. Transient radiation of the impacted plate in the plane $z_0=5$ cm: (a) on the impact axis and (b) at $x=y=20$ cm.

signal in zones I and II. Each point of the hologram contributes to the propagated signal over duration $T+R_0/c-t_w$, where R_0 is the distance between a measurement point located at (x_0,y_0,z_0) and the observation point. Therefore, the contribution of hologram point $(0,0,z_0)$ closest to the reconstruction point located at $(0,0,z)$ stops at $t=3.2$ ms. Thereafter, in zone III, contributions from the other hologram points (farther from the reconstruction point) stop gradually as the propagation distance increases, causing an attenuated signal.

Second, the time signal obtained using standard NAH with no trailing zeros [Fig. 5(c)] is quite different from the simulated signal [Fig. 5(a)]: the amplitude of the time aliasing error is greater than that of the initial pressure peak in Fig. 5(a). Consequently, this peak and its time of arrival at around 1.1 ms cannot be clearly identified from the signal. In fact, it can be shown that superimposing zones II and III over zone I in Fig. 5(b) gives the signal in Fig. 5(c). Better agreement with the simulation is obtained after $t=1.5$ ms due to the fact that, in this particular case, the amplitude ratio of the folded signal over the actual signal becomes small (less than 3%). This is a typical example of the time aliasing phenomenon when sound fields are propagated using Fourier transform based NAH.

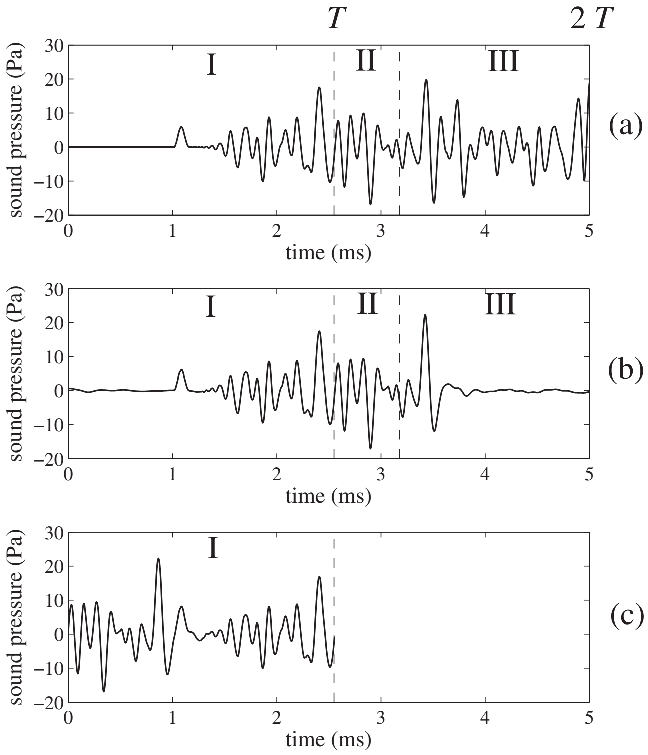


FIG. 5. Sound pressure on the impact axis at $z=35$ cm. (a) Simulated signal. Propagation using Fourier transform based NAH from field simulated at $z_0=5$ cm: (b) with trailing zeros added to the initial signals and (c) without trailing zeros.

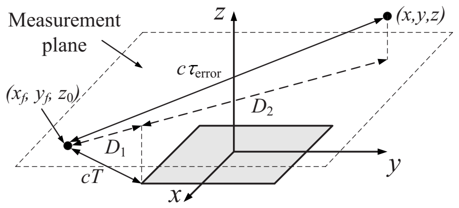


FIG. 6. Geometry used in estimating the error duration.

B. Duration of the error

The duration of the time aliasing error τ_{error} is defined as the minimum duration of the zero pad $(\bar{T}-T)$ required to avoid time aliasing. This duration can be estimated considering the point in the hologram (x_f,y_f,z_0) that is the farthest from the observation point and for which the sound pressure is not null at the end of the interval $t\leq T$. Generally, this point corresponds to the farthest corner of the measurement aperture. For a large hologram or a short observation window, as in the simulated hologram presented above, the error duration, shown in Fig. 6, is given by

$$\tau_{\text{error}}(x,y,z)=\frac{1}{c}\sqrt{(D_1+D_2)^2+d^2}, \qquad (9)$$

where

$$D_1^2=(cT)^2-z_0^2,$$

$$D_2^2=(L_x/2+|x|)^2+(L_y/2+|y|)^2.$$

In order to entirely avoid time aliasing in NAH, the zero pad must last at least $\bar{T}-T=\tau_{\text{error}}(x,y,z)$, that is, in the above example, 4.2 ms. In Fig. 5(b), since this condition is not satisfied, some ripples, which are the continuation of those in zone III, are mainly visible at $t<1$ ms. Even though their amplitude is relatively weak and they do not alter the initial pressure peak significantly, these ripples still cause sizable errors and more zeros would have been required to obtain a better reconstruction of the propagated signal. In addition, the farther the sound field is propagated, the more zeros are required, which also results in huge computing requirements. The use of the NLT instead of the DFT is thus proposed to obtain a better representation of the propagated signal without zero padding.

IV. NLT

The Laplace transform is a common tool to study transient phenomena. However, functions in the Laplace domain either may be difficult to invert or are only available in a discrete form. The direct and inverse NLTs have been introduced by Wilcox.[14] The NLT has been successfully applied to several fields such as electrical transmission lines,[14–16] vibrations,[17–19] and acoustics.[20,21]

A. Formulation of the NLT

The analytical Laplace transform of a function $f(t)$ is given by

$$F(s) = \int_{0}^{+\infty} f(t)e^{-st}dt, \tag{10}$$

where s is the time domain Laplace variable. Its inverse can be expressed in a complex form by the Mellin–Fourier integral[22]

$$f(t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} F(s)e^{st}ds \tag{11}$$

and is possible only if the system is stable relative to σ.

Replacing s by $\sigma+i\omega$, where σ is a constant damping factor and ω is the frequency used in the Fourier transform, and considering a causal system for which $f(t<0)=0$, Eqs. (10) and (11) can be rewritten as

$$F(\sigma + i\omega) = \mathcal{F}_t\{f(t)e^{\sigma t}\} \tag{12}$$

and

$$f(t) = e^{-\sigma t}\mathcal{F}_t^{-1}\{F(\sigma + i\omega)\}, \tag{13}$$

where $\mathcal{F}_t$ is the time DFT operator and $\mathcal{F}_t^{-1}$ is its inverse.[14] Depending on the sign of the exponential in the definition of the time domain Fourier transform, σ is positive or negative. It always acts as *damping*, i.e., the exponential factor in Eq. (12) must decrease sufficiently so that the amplitude at the end of the time signal becomes negligible. In our case, referring to Eqs. (4) and (5), σ is chosen to be negative.

B. Signal processing issues

As previously seen in Sec. III, the error at the beginning of sound pressure signals $p(t)$ is roughly equal to $p(T)$. So, one could think that the damping factor should be as large as possible to minimize the initial error. However, signal processing issues such as the Gibbs oscillations for which amplitudes are generally small compared to that of the signals of interest in NAH applications have to be treated carefully when applying the NLT.

The Gibbs phenomenon appears at sharp discontinuities of a given time signal when its spectrum is truncated. Since the use of the inverse time domain DFT leads to a periodization of the time function with $p(T^+)=p(0^+)\neq p(T^-)$, ripples mainly occur at the beginning and at the end of the signal. Therefore, the exponential amplification factor in Eq. (13) ($e^{-\sigma t}$, keeping in mind that σ is negative) can overamplify the Gibbs oscillations.

A first way to reduce the impact of the Gibbs phenomenon is to increase the sampling frequency so that truncation of the spectrum is less significant. However, this is not necessarily possible experimentally since sampling frequency is limited by the acquisition device. That is why windowing in the frequency domain has been introduced.[23] Several windows such as Hanning, Blackman, Lanczos, and Riesz have been studied by Ramirez *et al.*[16] However, the amplitudes of such windows decrease rapidly in the spectrum and this is not suitable for signals with large bandwidths compared to half of the sampling frequency $(F_s/2)$. It would lead to a significant attenuation of high frequency components which would be detrimental to the recovery of the initial time signal. In this paper, a Tukey window [Eq. (8)] is used as a trade-off between Hanning and rectangular windows. The width of the tapered rim on both sides is chosen to be $f_w=F_s/4$ so that half of the window is rectangular.

Values of σ have been proposed in the literature.[14,17] Parameters T and N_t (the number of points in the signal) and the bandwidth of the signal have been studied by Inoue *et al.*[24] The value of σ suggested by Wedepohl[15] was chosen for the current application. It is given with the appropriate sign by

$$\sigma = -\frac{2\ \ln(N_t)}{T}. \tag{14}$$

C. Application to NAH

The NLT can be applied to the standard NAH formulation to obtain a Fourier transform based TNAH method. Let us first consider a monochromatic wave

$$p(x,y,z,t) = P_0e^{i(k_xx+k_yy+k_zz)-st}, \tag{15}$$

which is artificially amplified with the factor σ in $s=\sigma+i\omega$. Similar to the Fourier domain development of NAH, Eq. (15) can be combined with the wave equation

$$\nabla^2p(x,y,z,t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(x,y,z,t) = 0 \tag{16}$$

to get

$$k_x^2 + k_y^2 + k_z^2 + \frac{s^2}{c^2} = 0.$$

It is then possible to write k_z as

$$k_z = k_z' + i\kappa, \tag{17}$$

with

$$k_z' = \left(\frac{a + \sqrt{a^2+b^2}}{2}\right)^{1/2},$$

$$\kappa = \left(\frac{-a + \sqrt{a^2+b^2}}{2}\right)^{1/2},$$

$$a = \frac{\omega^2 - \sigma^2}{c^2} - k_x^2 - k_y^2 \quad \text{and} \quad b = -\frac{2\omega\sigma}{c^2}.$$

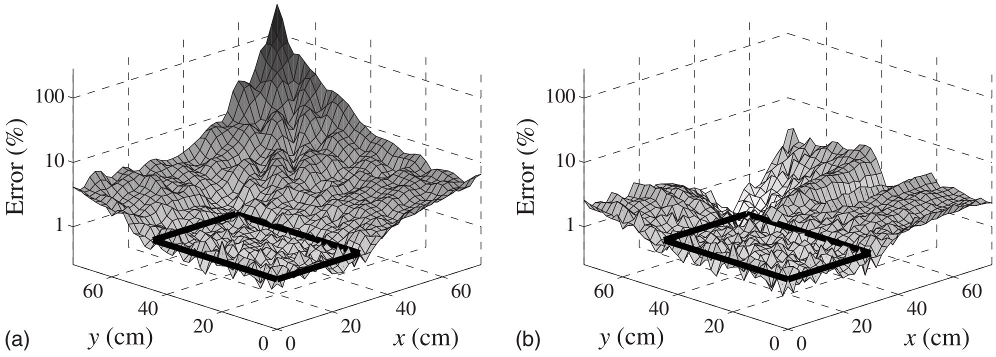


FIG. 7. Distribution of the relative error in the plane $z=10$ cm after a propagation of $d=5$ cm with (a) standard NAH and (b) TNAH.

By combining Eqs. (7) and (17), the frequency dependence of the k-space Green's function can be formulated in the Laplace domain, i.e.,

$$G_{pp} = e^{ik_z d} = e^{-\kappa d} e^{ik'_z d}, \tag{18}$$

where κ and k'_z are both real. The first term represents artificial damping and evanescence, and the second term corresponds to propagation. Accordingly, this transfer function attenuates all angular frequencies, even propagative waves, when sound fields are forward propagated, and amplifies frequencies when fields are propagated back toward the source.

Fourier transform based TNAH is formulated by using Eqs. (12) and (13) in Eqs. (4) and (5), respectively, and by applying the k-space Green's function in Eq. (18). It is now proposed to compare TNAH to the standard NAH formulation for forward propagation of transient sound fields.

V. COMPARISON OF TNAH WITH STANDARD NAH

To test the efficiency of the new transient approach and to compare it to Fourier transform based NAH, the sound pressure field simulated in the plane $z_0=5$ cm and studied in Sec. III A was propagated in both the Fourier and the Laplace domains. The main objective of this section is to compare both methods for same sources and same measurement aperture properties. A complete parametric study is not considered here since it has been done by de La Rochefoucauld *et al.* in the standard NAH case. Only the propagation distance (d) and the location of the observation point in the propagation planes are studied. Both techniques behave in a similar way with other parameters such as the distance between microphones and the number of measurements since they both involve the computation of spatial two dimensional DFTs.

To ensure minimally acceptable results in the Fourier domain, the duration of initial time signals was doubled by adding trailing zeros, as in Sec. III A. The relative error is calculated with

$$\mathrm{Er}(x,y) = \sqrt{\frac{\sum_{n=1}^{N_t^*} (P_{\mathrm{prop}}(x,y,t_n) - P_{\mathrm{ref}}(x,y,t_n))^2}{\sum_{n=1}^{N_t^*} P_{\mathrm{ref}}(x,y,t_n)^2}}, \tag{19}$$

where P_{prop} is the sound pressure signal propagated using one of the Green's functions [Eq. (7) or Eq. (18)] and P_{ref} is the pressure signal simulated at the propagation plane. Because the field propagated in the Fourier domain was tapered with a Tukey window, the relative error is calculated on the first $N_t^*=462$ points, which corresponds to a duration of 2.3 ms.

The error distribution over one-fourth of the calculation plane for a propagation distance $d=z-z_0=5$ cm is presented in Fig. 7. A projection of the plate is indicated with black lines. For both methods, errors are generally more significant near the boundaries of the propagation plane: there is less energy in these signals due to a longer time of arrival, so the denominator in Eq. (19) is very small. The overall relative error over the plate is slightly larger (≈1.3%) with the use of NAH, even with trailing zeros. For greater distances along the plane, the relative error in NAH increases exponentially whereas it remains relatively constant in TNAH.

In Fig. 8, relative errors are shown as functions of the propagation distance for three coordinates ($x=y=0$, $x=y=20$, and $x=y=40$ cm). The results are presented on a semi-log scale. The relative error in NAH with a Green's function sampled in the k-space domain (– – –) increases with the propagation distance, as it has been pointed out by de La Rochefoucauld *et al.*[5] The Green's function used in the TNAH formulation (——) is also sampled in the k-space domain; the relative error, however, remains relatively low and constant with distance.

A time domain example is presented in Fig. 9 on the impact axis and at $x=y=40$ cm for a propagation distance of 50 cm, from $z_0=5$ cm to $z=55$ cm. In this figure, both methods are compared to a signal simulated at $z=55$ cm.

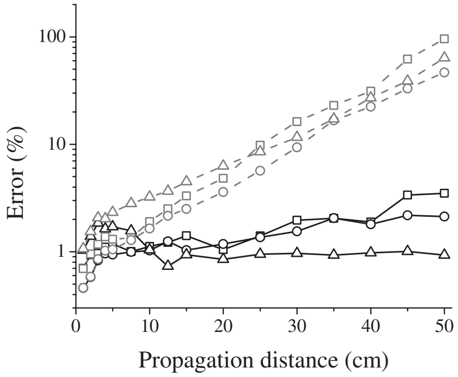


FIG. 8. Relative error as a function of the propagated distance: TNAH (——) and NAH (– – –). Impact axis (□), $x=y=20$ cm (○), and $x=y=40$ cm (△).

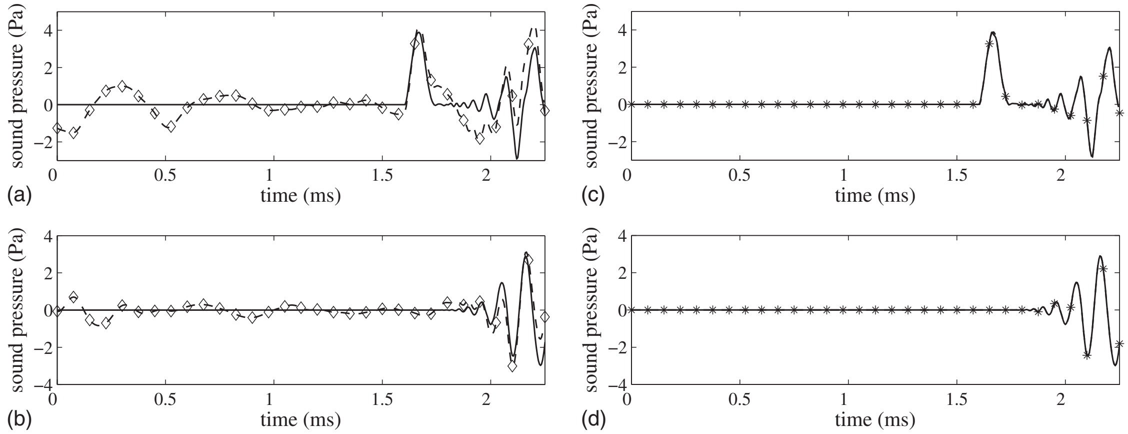


FIG. 9. Propagation from $z_0=5$ cm over a distance of 50 cm. Simulated signal (——), propagated signal using NAH (-◇-), and propagated signal using TNAH (– * –): (a) and (c) at (0, 0, 55) cm and (b) and (d) at (40, 40, 55) cm.

With TNAH, the initial pressure peak from 1.6 to 1.75 ms on the impact axis [Fig. 9(c)] and the dispersed signal [Fig. 9(d)] are easily identified. The simulated and propagated signals are clearly superimposed. With NAH, the time of arrival of the initial pressure peak [Fig. 9(a)] is not as clear as in TNAH; the time of arrival of the dispersed signal in Fig. 9(b) cannot be determined.

In summary, the percentage of error using the TNAH formulation is relatively constant and small compared to that of standard NAH, regardless of the propagation distance or the location of the observation point in the projection plane. Moreover, time domain zero padding is not required. In Sec. VI, experimental sound pressure data are forward propagated using TNAH and NAH. Reconstructed time signals are compared to measured sound pressure signals.

VI. EXPERIMENTAL VALIDATION

The experimental setup in Fig. 10 was used in an anechoic chamber to validate the potential of TNAH. A plate hung at its four corners was impacted at its center by a pendulum. The properties of the two bodies are listed in Table II. A two-axis manual translation stage was used to displace the 4×2 microphone array. The 130A Acousticel electret microphones were calibrated in amplitude and phase and were connected to an NI PXI-4472 acquisition device. Acquisition was triggered by another microphone located at 2.3 cm from the impact axis and at 2 cm from the plate.

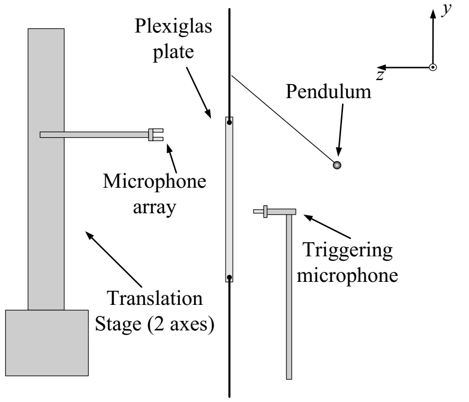


FIG. 10. Experimental setup.

The microphone array was moved over one quadrant of the plate in the plane $z_0=2.2$ cm. Symmetries were used to generate the entire hologram defined in Table II. Impacts were repeated ten times at each array location for signal averaging.

Signals measured at z_0 were propagated over a distance $d=6.9$ cm to the plane $z=9.1$ cm using the Green's functions in Eqs. (7) and (18). Prior to the NAH propagation only, the time signal was tapered using a Tukey window on a width of 0.05 ms. To keep a constant number of points for both methods, no trailing zeros were added in the time domain.

In Fig. 11, signals propagated with both approaches (– – –) are compared to measured signals (——) at two locations (x,y) on the reconstruction plane $z=9.1$ cm: (8.4, 0.0) cm and (15.4, 22.4) cm. The sound pressure signals are strongly damped and their amplitude is reduced to less than 1% of their maximum in less than 20 ms. Hence, one should expect small time aliasing errors with the standard NAH method. However, since initial signals were truncated at only 2.5 ms, these errors are significant enough to hide the propagation delay in Figs. 11(a) and 11(b). Such errors are not present in Figs. 11(c) and 11(d) for which TNAH was used and hence, propagation delays are in very good agreement with those observed experimentally. After 0.6 ms, both methods lead to propagated signals that are almost identical to each other and are quite similar to the measured signals. Differences with the experimental results are mainly due to

TABLE II. Properties of the experimental setup.

Impacting bodies		Hologram	
	Plexiglas plate	T	2.5 ms
$L_x \times L_y$	50.6×29.4 cm^2	F_s	102.4 kHz
h	5.8 mm	$N_x \times N_y$	55×51
	Steel pendulum	$\bar{N}_x \times \bar{N}_y$	109×101
r_s	5.5 mm	$\Delta x, \Delta y$	1.4 cm
v_s	2 m/s	x_w, y_w	7 cm

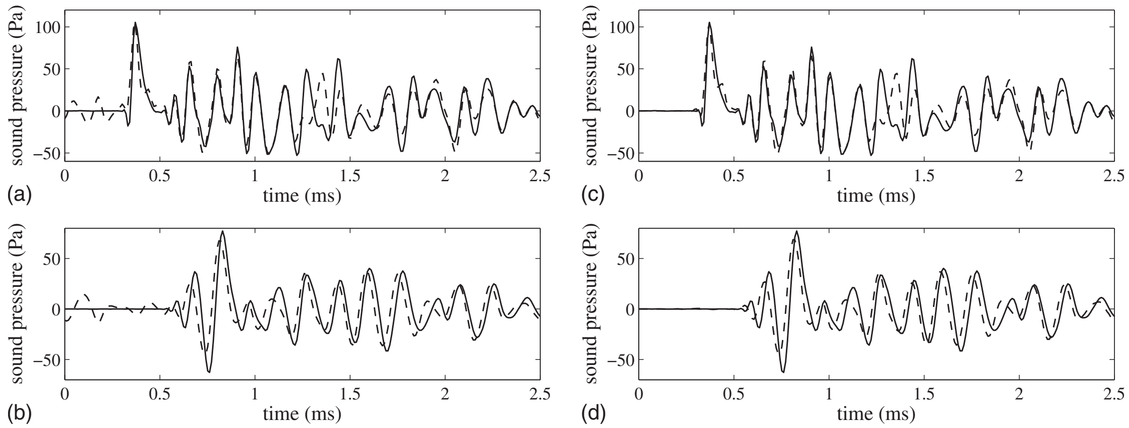


FIG. 11. Propagation of signals measured at $z_0=2.2$ cm over a distance of 6.9 cm. Signal measured on the reconstruction plane (——), signal propagated using NAH [(a) and (b)] or using TNAH [(c) and (d)] (– – –): (a) and (c) at $x=8.4$, $y=0.0$ cm and (b) and (d) at $x=15.4$, $y=22.4$ cm.

the spatial precision of the experimental setup that can cause phase shifts in the higher frequency components of the measured initial signals.

VII. CONCLUSION

Forward propagation of the initial sound pressure radiated by an impacted plate was presented using a *k*-space Green's function in the frequency domain. It was shown from numerical simulations that truncation of the time signals leads to significant time aliasing errors with Fourier transform based NAH. These errors mainly occur at the beginning of the time signals, thus masking the initial transient pressure in the calculation plane.

To reduce these errors, a Fourier transform based TNAH formulation in the Laplace domain was developed. The NLT is computed using a standard DFT algorithm. An error measure was introduced to compare both standard NAH and TNAH with respect to the propagation distance and the location of the observation point in the projection plane. TNAH significantly reduces aliasing errors without requiring trailing zeros in the initial time signals.

Forward propagation of the transient noise of an impacted plate was tested from experimental data. The quality of propagated transient signals is significantly improved with the use of TNAH. Initial transient time signals can be recovered without loss of precision on the amplitudes or propagation delay, provided that the initial data are accurately measured.

The use of TNAH for backward propagation still needs to be investigated. The application of the new formulation in the Laplace domain could also be extended to non-planar coordinates or other acoustical reconstruction methods to visualize the transient sound pressure radiated by more complex structures.

ACKNOWLEDGMENTS

This work was supported by the Fonds québécois de la recherche sur la nature et les technologies, the Institut de recherche Robert-Sauvé en santé et en sécurité au travail, and the Natural Sciences and Engineering Research Council of Canada. Sincere thanks to Sylvie Nguene and Valérie Lalonde for their help with experimental validation.

[1] S. Schedin, C. Lambourge, and A. Chaigne, "Transient sound fields from impacted plates: Comparison between numerical simulations and experiments," J. Sound Vib. **221**, 471–490 (1999).

[2] J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).

[3] E. G. Williams, *Fourier Acoustics, Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, CA, 1999).

[4] O. de La Rochefoucauld, "Résolution du problème inverse spatio-temporel en imagerie acoustique de champ proche: Application au rayonnement de sources industrielles instationnaires (Resolution of the space/time inverse problem in near field acoustical holography: Application to the radiation of non stationary industrial sources)," Ph.D. dissertation, Université du Maine, Le Mans, France (2002).

[5] O. de La Rochefoucauld, M. Melon, and A. Garcia, "Time domain holography: Forward projection of simulated and measured sound pressure fields," J. Acoust. Soc. Am. **116**, 142–153 (2004).

[6] G. T. Clement, R. Liu, S. V. Letcher, and P. R. Stepanishen, "Forward projection of transient signals obtained from a fiber-optic pressure sensor," J. Acoust. Soc. Am. **104**, 1266–1273 (1998).

[7] J. A. Mann III, E. G. Williams, K. Washburn, and K. Grosh, "Time-domain analysis of the energy exchange between structural vibrations and acoustic radiation using near-field acoustical holography measurements," J. Acoust. Soc. Am. **90**, 1656–1664 (1991).

[8] S. F. Wu, H. Lu, and M. S. Bajwa, "Reconstruction of transient acoustic radiation from a sphere," J. Acoust. Soc. Am. **117**, 2065–2077 (2005).

[9] Z. Wang and S. F. Wu, "Helmholtz equation–least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).

[10] W. Heitkämper, "Näherungsweise Berechnung der Schallabstrahlung von storßartig angeregten Platten (Numerical approximation of sound radiation of impact-excited plates)," Acustica **58**, 141–148 (1985).

[11] W. Goldsmith, *Impact: The Theory and Physical Behaviour of Colliding Solids* (Edward Arnold Ltd., London, England, 1960).

[12] W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).

[13] A. Ross and G. Ostiguy, "Propagation of the initial transient noise from an impacted plate," J. Sound Vib. **301**, 28–42 (2007).

[14] D. J. Wilcox, "Numerical Laplace transformation and inversion," Int. J. Electr. Eng. Educ. **15**, 247–265 (1978).

[15] L. M. Wedepohl, "Power systems transients: Errors incurred in the numerical inversion of the Laplace transform," in Proceedings of the Midwest Symposium on Circuits and Systems, Puebla, Mexico (1983), pp. 174–178.

[16] A. Ramirez, P. Gomez, P. Moreno, and A. Gutierrez, "Frequency domain analysis of electromagnetic transients through the numerical Laplace transforms," in Proceedings of the IEEE Power Engineering Society General Meeting, Denver, CO (2004), Vol. **1**, pp. 1136–1139.

[17] W. Krings and H. Waller, "Contribution to the numerical treatment of partial differential equations with the Laplace transformation—An application of the algorithm of the fast Fourier transformation," Int. J. Numer. Methods Eng. **14**, 1183–1196 (1979).

[18] D. E. Beskos and A. Y. Michael, "Solution of plane transient elastodynamic problems by finite elements and Laplace transform," Comput. Struct. **18**, 695–701 (1984).

[19] H. Inoue, K. Kishimoto, T. Shibuya, and K. Harada, "Regularization of numerical inversion of the Laplace transform for the inverse analysis of impact force," JSME Int. J., Ser. A **41**, 473–480 (1998).

[20] D. Lévesque and L. Piché, "A robust transfer matrix formulation for the ultrasonic response of multilayered absorbing media," J. Acoust. Soc. Am. **92**, 452–467 (1992).

[21] T. Nishigaki, T. Ohyama, and M. Endo, "Study of the transient sound radiated by impacted solid bodies based on the boundary element method," JSME Int. J., Ser. C **39**, 218–224 (1996).

[22] J. L. Schiff, *Laplace Transform: Theory and Applications* (Springer-Verlag, New York, 1999).

[23] S. J. Day, N. Mullineux, and J. R. Reed, "Developments in obtaining transient response using Fourier transforms. I. Gibbs phenomena and Fourier integrals," Int. J. Electr. Eng. Educ. **3**, 501–506 (1965).

[24] H., Inoue, M., Kamibayashi, K., Kishimoto, T., Shibuya, and T., Koizumi, "Numerical Laplace transformation and inversion using fast Fourier transform," JSME Int. J., Ser. I **35**, 319–324 (1992).

Generalized shot noise model for time-reversal in multiple-scattering media allowing for arbitrary inputs and windowing

Kevin J. Haworth
Department of Radiology and the Applied Physics Program, University of Michigan, Ann Arbor, Michigan 48109

J. Brian Fowlkes and Paul L. Carson
Department of Radiology, Department of Biomedical Engineering, and the Applied Physics Program, University of Michigan, Ann Arbor, Michigan 48109

Oliver D. Kripfgans[a]
Department of Radiology and the Applied Physics Program, University of Michigan, Ann Arbor, Michigan 48109

(Received 21 May 2008; revised 26 January 2009; accepted 25 February 2009)

A theoretical shot noise model to describe the output of a time-reversal experiment in a multiple-scattering medium is developed. This (non-wave equation based) model describes the following process. An arbitrary waveform is transmitted through a high-order multiple-scattering environment and recorded. The recorded signal is arbitrarily windowed and then time-reversed. The processed signal is retransmitted into the environment and the resulting signal recorded. The temporal and spatial signal and noise of this process is predicted statistically. It is found that the time when the noise is largest depends on the arbitrary windowing and this noise peak can occur at times outside the main lobe. To determine further trends, a common set of parameters is applied to the general result. It is seen that as the duration of the input function increases, the signal-to-noise ratio (SNR) decreases (independent of signal bandwidth). It is also seen that longer persisting impulse responses result in increased main lobe amplitudes and SNR. Assumptions underpinning the generalized shot noise model are compared to an experimental realization of a multiple-scattering medium (a time-reversal chaotic cavity). Results from the model are compared to random number numerical simulation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106133]

PACS number(s): 43.60.Tj, 43.60.Cg [DRD] Pages: 3129–3140

I. INTRODUCTION

Parvulescu and Clay[1] performed the first time-reversal acoustics (TRA) experiments in 1965. In the following two decades, however, relatively little work was done in TRA. Beginning approximately 15 years ago, however, there has been a burst of activity, led in large part by Professor Mathias Fink. Over this time, nearly all of the experiments can be described by the general process of sending a short pulse into a medium and recording the resulting wave $s(t)$. The recorded signal is time-reversed $s(t) \rightarrow s(-t)$ (and possibly processed) and then retransmitted into the medium. The resulting time-reversal focused signal is recorded and analyzed.

One area of TRA that has shown both surprising and fruitful results is the time-reversal of waves that have traveled through random high-order multiple-scattering environments. Traditionally, random multiple-scattering environments display chaotic behaviors that prohibit focusing through them. Propagating waves through a "forest-of-needles" to a point receiver and then time-reversing this signal and showing that it would refocus at its origin was the first demonstration of the ability to focus through this type of medium.[2] Surprisingly, it was found that the focused signals were even more well defined after going through the multiple-scattering medium than if focusing was performed in a homogenous medium![2] These results were extended to other high-order multiple-scattering environments such as a chaotic cavity, where the boundaries of a reverberant material scattered the sound.[3] Various approaches have been taken to describe the unexpected results of TRA in multiple-scattering media. This paper is concerned with generalizing the approach first developed by Derode *et al.*[4] to explain this phenomenon. This approach treats the multiple-scattering events via a shot noise model. The shot noise approach allowed Derode *et al.*[4] to successfully model 1-bit time-reversal. It has also been used to determine the impact of windowing a signal before performing time-reversal.[5] Others have also extended the shot noise model to incorporate additional phenomena, such as scattering dependencies in forest-of-needle experiments.[6] In addition to the shot noise approach other approaches have used scattering theory,[7–9] eigenmode decomposition,[3] and Green's functions.[10] Each of which makes different assumptions and elucidates different effects (e.g., coherent backscatter,[9] noise emission in time-reversal,[11] etc.).

The primary motivation for the development of this generalized model is to predict the signal-to-noise ratio (SNR) for future time-reversal chaotic cavity (TRCC) experiments

[a] Author to whom correspondence should be addressed. Electronic mail: oliver.kripfgans@umich.edu

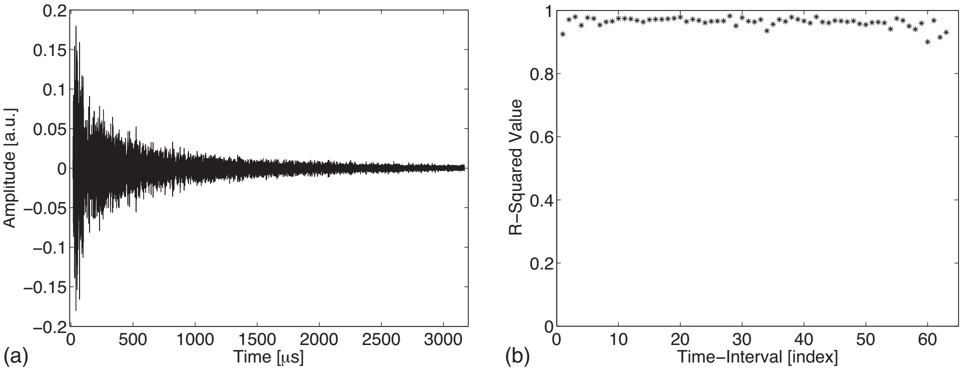


FIG. 1. (a) Experimentally obtained IR for a TRCC. (b) R^2 values indicating that the amplitudes in each 50-μs time-interval of the IR are well modeled as normal random variables.

(though other applications may be found, as illustrated below). In particular, for cases where one is interested in transmitting signals besides delta functions, the long tone-bursts are associated with high-intensity focused ultrasound for heating or acoustic radiation force experiments. TRCCs were first introduced by Fink and co-workers.[3,12,13] TRCC experiments work by having a transducer transmit an acoustic/elastic pulse into a solid (typically a metal). The sound reverberates within the cavity, reflecting off of the solid's walls. The acoustic signal at any point within the cavity can quickly become a diffuse wave. If the diffuse wave is recorded, time-reversed, and retransmitted, the waves will approximately retrace their paths and focus at the transducer that originally transmitted the pulse. The initial experiments by Draeger and Fink measured elastic waves in a two dimensional silicon wafer. Quieffin *et al.*[14] showed that this concept could be extended to three dimensional (3D) solids and more importantly that if the solid were put into contact with a water bath, signal would leak out of the solid and could be recorded with a hydrophone in the water. Then using spatial reciprocity, they showed that if the signal recorded by the hydrophone was time-reversed and retransmitted by the original transducer, a pulse would focus on the hydrophone's location outside of the cavity. The location of the hydrophone could be varied and thus it was found that TRCCs allowed focusing throughout a 3D volume with as few as one ultrasound transducer. Additional work by Fink and co-workers has led to prototypes for imaging devices[15] and high amplitude ultrasound therapy devices[16] among other applications. Since the initial development of TRCCs, Sarvazyan and co-workers[17–19] have also made significant progress in understanding and utilizing TRCCs.

Modeling of high-order multiple-scattering TRA has found other applications. These include biomedical engineering,[20] non-destructive testing and evaluation,[21,22] geophysics,[23–25] underwater acoustics,[26] imaging,[27,28] and (wireless) communication.[29–31] In many of these applications, one may be interested in sending not just a short pulse through the multiple-scattering medium but an extended pulse that could be used to contain extensive information or induce an effect. In this paper, the model initiated by Derode *et al.*[4] is generalized to account for arbitrary input functions and arbitrary windowing. The goal of the model will be to compute the expectation value and variance of a time-reversal focused signal through a multiple-scattering medium.

This article is organized as follows. Section II outlines the problem in greater detail (Sec. II B), derives the expectation value (Sec. II C), variance (Sec. II D), and directivity pattern (Sec. II F), and provides some physical (Sec. II A) and numerical support (Sec. II E) for the model. In Sec. III a set of common parameters is then applied to these general results, allowing the expectation value (Sec. III A) and variance (Sec. III B) to be simplified. The SNR under these conditions is also derived and discussed (Sec. III C).

II. GENERAL THEORY

Derode *et al.*[4] provided an explanation for using a shot noise model to describe multiple-scattering events that is briefly repeated here for completeness. For any high-order random multiple-scattering process, the majority of the signal is composed of a diffuse wave that is the result of interference of a large number of multiple-scattering paths. Since the scattering is random (e.g., random variations in the impedance of the medium) the arrival time from each path can be described as a series of Poisson impulses. For an impulse sent through the medium, the output can be described by the convolution of the acousto-electrical impulse response (IR) with the Poisson impulses corresponding to every path. By definition, this is a shot noise process.[32] Since the number of paths is large, one can assume that the density of Poisson impulses per acousto-electrical IR is large. This then allows the shot noise process to be approximated as a normal random process with mean zero (assuming that the input signal is also mean zero) and variance $\sigma^2(t)$.[32]

A. Experimental validation of shot noise assumptions

To further motivate the use of a shot noise model, the IR was experimentally obtained from a TRCC (Imasonic SAS, Besançon, France) constructed to be the same as the one used by Montaldo *et al.*[16] The IR was obtained by driving one of the elements on the TRCC with a step function generated from an HP33120A function generator (Agilent, Palo Alto, CA) and amplified by an ENI A-300 rf power amplifier (Rochester, NY). The face opposite the elements was placed in a water tank, and a hydrophone (PVF$_2$, Raytheon Co., Waltham, MA) recorded the signal transmitted into the water. The IR [Fig. 1(a)] of the cavity was recorded and broken into 50 μs time-intervals. If the IR can be modeled as a normal random variable, then the amplitudes within each time-interval should have a normal distribution. To determine if this is true, the amplitudes in each interval were plotted in a histogram. The histogram was then fitted to a Gaussian function $Ae^{(x-\mu)^2/2\sigma^2}$ and the R-squared value (i.e., square of the Pearson product-moment correlation coefficient) computed to determine the goodness-of-fit. Figure 1(b) shows that the computed R-squared values are indeed all close to 1, indicating a good fit.

While it has been demonstrated that the assumptions used for the shot noise model are valid for a specific case, the derivations that follow should hold for any high-order multiple-scattering process that satisfy the assumptions made above. In fact, the derivation is not necessarily specific to acoustic waves and could be applied to any wave phenomena where phase coherent detectors with reversible signals exist or are developed (e.g., radio frequency electromagnetic waves). The results of the derivations will be compared to a random number numerical simulation, rather than a physical experiment or a numerical simulation based on the wave equation. Others have shown that the shot noise model does describe physical results of time-reversal focusing under various particular conditions.[5,6]

B. Time-reversal focused signal

The generic time-reversal experiment that will be modeled occurs as follows. An input signal $g(t)$ is transmitted into the scattering medium, which has an IR $h(t)$. The resulting output signal $g(t)\otimes_t h(t)$ is recorded (where $\otimes_t$ is a convolution over time t). A window function $W(t)$ is applied to the output signal, which is then normalized by its maximum value $\mathbf{M}$:

$$\frac{1}{\mathbf{M}} W(t)\cdot(g(t)\otimes_t h(t)). \tag{1}$$

The window function selects the desired portion of the total output signal that will be used in the time-reversal experiment. The windowed function is then time-reversed ($t \rightarrow -t$) with a temporal shift of T to ensure causality, yielding

$$\frac{1}{\mathbf{M}} W(T-t)\cdot(g(-t)\otimes_t h(T-t)). \tag{2}$$

The time-reversed function is now retransmitted into the medium yielding an output

$$r(t)=\left(\frac{1}{\mathbf{M}} W(T-t)\cdot(g(-t)\otimes_t h(T-t))\right)\otimes_t h(t). \tag{3}$$

The expectation value and variance of this output will be computed.

C. Expectation value

As Derode *et al.*[4] showed, even though $\mathbf{M}$ is a random variable, because of its origins in $g(t)\otimes h(t)$, it is approximately constant and thus can be pulled out of the expectation integral. As such it will now be denoted as M. Haworth[33] provided an approximation for M. To compute the expectation value of Eq. (3), the convolutions are written as integrals. The integration variables for the convolutions are θ_t and τ_θ. The functional variables associated with each convolution are t and θ, respectively,

$$E\{r(t)\} \approx \frac{1}{M} E\left\{ \int_{\theta_t=-\infty}^{\infty} \left(W(T-\theta_t) \times \int_{\tau_\theta=-\infty}^{\infty} g(-\tau_\theta) \times h(T-(\theta_t-\tau_\theta))\, d\tau_\theta \right) h(t-\theta_t)\, d\theta_t \right\}. \tag{4}$$

Since $h(t)$ defines a stochastic process with a normal probability distribution function $f(h(t))$, the expectation value of $r(t)$ can be computed as

$$E\{r(t)\} = \int_{h(t)=-\infty}^{\infty} [r(t;h(t))\cdot f(h(t))]\, dh(t). \tag{5}$$

Using this notation, Eq. (4) is first rewritten expressing the expectation value as an integral. The integrals are then reordered, making the integral over $h(t)$ the innermost integral, and finally rewriting that integral in the $E\{\cdot\}$ notation

$$E\{r(t)\} \approx \frac{1}{M} \int_{\tau_\theta=-\infty}^{\infty} \int_{\theta_t=-\infty}^{\infty} g(-\tau_\theta) W(T-\theta_t) \times E\{h(T-(\theta_t-\tau_\theta)) h(t-\theta_t)\}\, d\theta_t\, d\tau_\theta. \tag{6}$$

Now all constants with respect to the stochastic process $h(t)$ have been removed from the expectation value. As was described above, $h(t)$ is a normal random variable with mean zero and variance $\sigma^2(t)$. The expectation value of two mean zero, jointly normal random variables multiplied together is[32]

$$E\{\mathbf{x}(t_1)\cdot\mathbf{y}(t_2)\} = \rho(t_1-t_2)\cdot\sigma_x(t_1)\cdot\sigma_y(t_2), \tag{7}$$

where $\rho(t_1-t_2)=\rho(t_2-t_1)$ is the correlation coefficient of the normal random variables $\mathbf{x}(t_1)$ and $\mathbf{y}(t_2)$, and $\sigma_x(t_1)$ and $\sigma_y(t_2)$ are the standard deviations of $\mathbf{x}(t_1)$ and $\mathbf{y}(t_2)$, respectively. Applying this and rewriting the integrals in traditional convolution notation yields

$$E\{r(t)\} \approx \frac{1}{M} g(-t) \otimes_t [\rho(T-t)((W(T-t)\cdot\sigma(T-t))\otimes\sigma(t))]. \tag{8}$$

Equation (8) is the statistical approximation for the expectation value of the time-reversal focused signal. Note that it is non-zero only for a duration approximately as long as the input function $g(t)$. This time with non-zero amplitude will be referred to as the main lobe of the signal and all times outside this as side lobes. Looking at the square bracket term, one sees that a rapidly oscillating function $\rho(T-t)$ is multiplied by a relatively slow changing function $(W(T-t)\cdot\sigma(T-t))\otimes_t\sigma(t)$ (assuming that the window function is not rapidly changing). As a result, the term in the square brackets will look approximately like a scaled version of $\rho(T-t)$. Therefore, how large this term is, and thus how large $E\{r(t)\}$ is, depends directly on the amplitude and decay of the envelope and how large the window is. Therefore the result found by others that the expected signal increases as the amplitude of the IR (or time-reversed signal in the case of 1-bit time-reversal) increases is confirmed.[4,5,34] Equation (8) is easily shown to simplify to the corresponding result of Derode *et al.*[4] [term 1 of Eq. A2] by letting the window extend to positive and negative infinities [i.e., $W(t)=1\,\forall\, t$], setting $T=0$, and letting $g(t)\rightarrow\delta(t)$.

D. Variance

Since the variance can be computed from

$$\mathrm{VAR}\{r(t)\}=E\{r^2(t)\}-E^2\{r(t)\},$$

the process is similar to what was done in Sec. III C. However, the mathematics are more tedious due to the squaring of $r(t)$.[33] The result is

$$\begin{aligned}&\left(\frac{1}{M^2}\int_{\theta_1=-\infty}^{\infty}\int_{\theta_2=-\infty}^{\infty}W(T-\theta_1)W(T-\theta_2)\sigma^2(t-\theta_1)\sigma^2(t-\theta_2)\right.\\&\quad\times[g(-\theta_1)\otimes_{\theta_1}\rho(T-\theta_1-(t-\theta_2))]\\&\quad\left.\times[g(-\theta_2)\otimes_{\theta_2}\rho(T-\theta_2-(t-\theta_1))]d\theta_2 d\theta_1\right)\\&\quad+\left(\frac{1}{M^2}\int_{\theta_1=-\infty}^{\infty}W(T-\theta_1)\sigma(t-\theta_1)\right.\\&\quad\times[\sigma(t)\otimes_t\{W(T-t)\rho(t-\theta_1)\\&\quad\left.\times[g(-\theta_1)\otimes_{\theta_1}\{\sigma^2(T-\theta_1)[g(-t)\otimes_t\rho(t-\theta_1)]]\}\}]d\theta_1\right).\end{aligned}\qquad(9)$$

Each of these two terms can now be related to the physical processes occurring in multiple-scattering. The nature of the symmetry of the square brackets in the first term indicates that this term only contributes at the main lobe and is zero outside of it. This term comes from variations in the total energy contained in the time-reversed signal that is transmitted into the medium. That is, it originates from the variance in the total energy in the time-reversed signal over different realizations of multiple-scattering processes (i.e., a reordering of the scatterers or placement of the sound source). Note that on average the energy will be the same for a particular interval, and conservation of energy dictates that the total energy in $h(t)$ over all time will be constant. However, over different realizations of the multiple-scattering environment, the distribution of energy will change, even for the same time-interval. Due to this, the first term will be referred to as the coherent or correlated variance. To minimize this variance, one might ensure that the wave is completely diffuse or try to include a larger portion of the signal when time-reversing. Further changes that can be made to reduce this term are outlined in Sec. III.

The second term is due to the interference of the time-reversed signal and $h(t)$ at all times when the two waveforms are uncorrelated. Thus one can refer to this term as being the incoherent or uncorrelated variance. To maximize the main lobe to side lobe ratio (a possible SNR definition) one would be interested in minimizing the second term while maximizing the expectation value. Alternatively, for a more consistent main lobe amplitude, the first term should be minimized.

E. Comparison of derived model with numerical simulation

In addition to verifying the equations by showing their reduction to previously obtained results, the equations were also compared to ensembles of simulated data, as was done by Derode *et al.*[4] The simulated data were created from a normally distributed random array [created using the randn function in MATLAB (The Mathworks, Inc., Natick, MA), which is based on the Ziggurat algorithm[35] that has a period of approximately 2^{64}] convolved with a normalized 3.5-cycle sine-wave to simulate $h(t)$, where the 3.5-cycle sine-wave models an acousto-electrical IR. $h(t)$ was then convolved with $g(t)$ (chosen as a delta function for this simulation), windowed, and time-reversed. The result was then convolved with the original $h(t)$ to give a simulated realization of $r(t)$. The mean and standard deviation of an ensemble of 500 simulated $r(t)$ [each with a unique $h(t)$] were plotted against the analytical solution for $E\{r(t)\}$ [Eq. (8)] and the square root of the $\mathrm{VAR}\{r(t)\}$ [Eq. (9)] (Fig. 2). The figure demonstrates that the model accurately predicts the numerical simulation. R^2 values quantifying the goodness-of-fit are shown for each plot.

Based on the equations derived for the expectation value and variance and Fig. 2, one can draw conclusions for how various parameters will impact the time-reversal focusing. Initially, if one focuses on Figs. 2(b) and 2(c) it is possible to see the impact of the decay constant of the envelope of the IR. As one would physically expect, both the variance and the expectation value increase as the decay constant increases. This is due to the fact that there is more energy in the time-reversed signal. Also, one can see that the side lobes fall off more slowly as the time constant increases. The extrapolation of this result has been seen for 1-bit time-reversal, where the side lobes are approximately constant in amplitude[4] and is consistent with the physical explanation of the incoherent variance given earlier. Next, the impact of shifting the window to later times can also be seen [Figs. 2(c) and 2(d)]. Looking at the expectation value, one sees that its maximum amplitude has decreased for Fig. 2(d). This is expected for two physical reasons. First, since the signal is windowed, the signal that is retransmitted has less total energy. Additionally, this signal correlates with the IR at a later time when the IR has decreased in amplitude. Thus the amplitude of the correlation of the signals (which is essentially what the time-reversal process does) is smaller. Analogous reasoning also leads to explaining why the coherent variance is seen to be smaller. For Figs. 2(b)–2(d), the shape of the incoherent portion of the side lobes initially increases in all cases as more of the time-reversed signal is transmitted and incoherently interferes with the IR. Once the entire time-reversed signal has been transmitted and no more energy is being injected into the system, the incoherent interference decreases as the magnitude of the IR falls off with time. Thus for a delayed window, such as Fig. 2(d), the incoherent interference peaks and begins to fall off before the time-reversed signal has lined up and coherently interferes with the IR. Hence, the variance peaked around 600 μs while the expectation value peak did not occur until 1050 μs. This

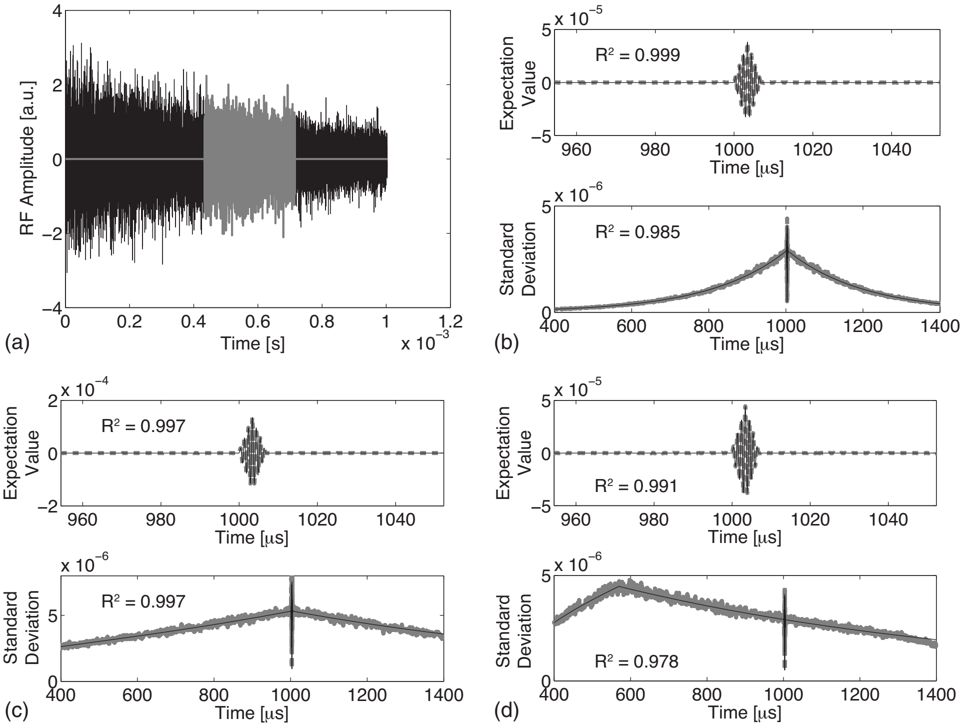


FIG. 2. (a) Sample of a numerically simulated IR with the gray section corresponding to the windowed portion used for (d). (b)–(d) compare the expectation/mean value (top plots) and standard deviation (i.e., square root of the variance) (lower plots) for the statistical model (black line) and numerical simulation (gray line). (b) Time-reversal focusing of the full IR shown in (a). (c) Time-reversal focusing of a full IR with a slower decay than the one shown in (a). (d) Time-reversal focusing of the windowed (gray) portion of the IR in (a). R^2 values are given for each plot demonstrating that the model is a good fit to the numerical simulation.

should always be seen whenever the window zeros out the initial part of the recorded signal $g(t)\otimes_t h(t)$. Note that the shape of the side lobes after the peak should fall off like $\sigma(t)$, while the shape before the peak will be more complicated and depend on both $g(t)$ and $\sigma(t)$. Further trends for changing parameters will be investigated more closely in Sec. III.

F. Directivity pattern

To estimate the directivity pattern, the same process as above will be used; however, the IR will be specific to a particular location x_o. As a result, Eq. (2) is written as

$$\frac{1}{\mathbf{M}} W(T-t)\cdot(g(-t)\otimes_t h(x_o,T-t)). \tag{10}$$

This windowed and time-reversed signal is then retransmitted into the cavity and recorded at a different location x_1. Thus Eq. (10) is convolved with an IR $h(x_1,t)$ from a different location x_1. Assuming that the change in distance is large enough for $h(x_o,t)$ and $h(x_1,t)$ to decorrelate, which can be determined from the van Cittert–Zernike theorem,[4,36] the expectation value and variance are easy to compute. The expectation value goes to zero since $\rho(t_1-t_2,x_o-x_1)$ in Eq. (7) goes to zero as x_o-x_1 increases. Similarly, the coherent variance component will also go to zero, leaving only the incoherent variance component. Assuming that each of the paths has the same scattering statistics [i.e., $h(x_o,t)$ and $h(x_1,t)$ both decay as $\sigma(t)$] then, the total variance is

$$\mathrm{VAR}\{r(x_1,t;x_o)\} = \Bigg(\frac{1}{M^2}\int_{\theta_1=-\infty}^{\infty} W(T-\theta_1)\sigma(t-\theta_1) \times[\sigma(t)\otimes_t\{W(T-t)\rho(t-\theta_1,x_1-x_1) \times[g(-\theta_1)\otimes_{\theta_1}\{\sigma^2(T-\theta_1) \times[g(-t)\otimes_t\rho(t-\theta_1,x_o-x_o)]\}]\}]d\theta_1\Bigg). \tag{11}$$

Just as Derode *et al.*[4] found previously in the simplified case, it is found here that the −6 dB width of the directivity pattern is a measure of the correlation length of the scattered waves. For applications in imaging, this term determines the background noise, above which all signals must be observed. Also note that as the amplitude of $g(t)$ increases, the noise floor will increase.

If $h(t)$ and $h'(t)$ decay with different decay envelopes [$\sigma(t)$ and $\sigma'(t)$ respectively], then the result becomes

$$\mathrm{VAR}\{r(x_1,t;x_o)\} = \Bigg(\frac{1}{M^2}\int_{\theta_1=-\infty}^{\infty} W(T-\theta_1)\sigma'(t-\theta_1 \times)[\sigma'(t)\otimes_t\{W(T-t)\rho(t-\theta_1,x_1-x_1) \times[g(-\theta_1)\otimes_{\theta_1}\{\sigma^2(T-\theta_1) \times[g(-t)\otimes_t\rho(t-\theta_1,x_o-x_o)]\}]\}]d\theta_1\Bigg). \tag{12}$$

III. APPLICATION TO A COMMON SET OF PARAMETERS

While it was possible to ascertain some qualitative physical insights from the above equations, they do not lend themselves to easily determining the quantitative impact of various parameters (such as changing the window placement or input function). In this section certain conditions for the envelope $\sigma(t)$, window function $W(t)$, and input function that are commonly seen in experimental work will be assumed

[Eq. (13)]. This will make it possible to simplify $E\{r(t)\}$ [Eq. (8)] and $\mathrm{VAR}\{r(t)\}$ [Eq. (9)] to the point where trends can be surmised. This will result in a better understanding of the above equations and show the results for commonly seen experimental parameters. If any of these assumptions are violated, one can always return to the original equations from Sec. II.

As others have noted, for high-order multiple-scattering events, the envelope of the scattered signal, which is proportional to the standard deviation, often decays exponentially[4,5,37] [e.g., Fig. 1(a)]. Therefore it will be assumed that $\sigma(t)=u(t)e^{-\alpha t}$, where $u(t)$ is the Heaviside function. Next, a rect-window will be used for windowing the time-reversed signal. It will also be assumed that the duration of $g(t)$, t_g, is small compared to the envelope decay time constant ($\tau_\sigma=1/\alpha$). While there is interest in choosing long input functions, this assumption is necessary to simplify the expectation value and variance. As will be seen, as the assumption $t_g \ll \tau_\sigma$ is initially violated (i.e., $t_g < \tau_\sigma$ holds but $t_g \ll \tau_\sigma$ does not hold), the trends found for the expectation value and variance below will still hold approximately, though the exact equations will not. This is because this condition is merely used to assure that the envelopes of functions do not change over time-intervals specified below. As $\sigma(t)$ does change, it is slow and smooth so the impact is not dramatic. Of course when $t_g \ll \tau_\sigma$ is strongly violated (i.e., $t_g \geq \tau_\sigma$), one must return to the equations derived in Sec. II. Finally, it will be assumed that the duration of the acousto-electric IR $\rho(t)$, t_ρ, is small compared to the t_g. Summarizing these assumptions,

$$\sigma(t) = u(t)e^{-\alpha t}, \tag{13a}$$

$$W(t) = \begin{cases} 1 & \text{if } t_{\text{on}} \leq t \leq t_{\text{off}} \\ 0 & \text{otherwise}, \end{cases} \tag{13b}$$

$$t_\rho \ll t_g \ll \frac{1}{\alpha} = \tau_\sigma. \tag{13c}$$

A. Expectation value

Recalling the expectation value [Eq. (8)],

$$E\{r(t)\} \approx \frac{1}{M} g(-t) \otimes_t [\rho(T-t)((W(T-t)\sigma(T-t)) \otimes \sigma(t))].$$

Applying the assumptions outlined in Eq. (3) and beginning with the innermost portion,

$$(W(T-t)\cdot\sigma(T-t)) \otimes \sigma(t)$$

$$= \int_{-\infty}^{\infty} W(T-\tau)\sigma(T-\tau)\sigma(t-\tau)d\tau \tag{14}$$

$$= \int_{T-t_{\text{off}}}^{T-t_{\text{on}}} u(T-\tau)e^{-\alpha(T-\tau)}u(t-\tau)e^{-\alpha(t-\tau)}d\tau \tag{15}$$

$$= \frac{e^{-\alpha t}e^{\alpha T}}{2\alpha} e^{-2\alpha t_{\text{on}}}(1-e^{-2\alpha\Delta t}), \tag{16}$$

where $\Delta t = t_{\text{off}} - t_{\text{on}}$ is the window width. Equation (16) is then multiplied by $\rho(T-t)$, which is non-zero only for $t \approx T$. Additionally, since Eq. (16) is a slowly changing function (on the order of $1/\alpha$), the following approximation can be made:

$$\rho(T-t)((W(T-t)\cdot\sigma(T-t)) \otimes \sigma(t))$$

$$\approx \delta(T-t)\left(\frac{e^{-\alpha t}e^{\alpha T}}{2\alpha}e^{-2\alpha t_{\text{on}}}(1-e^{-2\alpha\Delta t})\right) \tag{17}$$

$$= \frac{1}{2\alpha}e^{-2\alpha t_{\text{on}}}(1-e^{-2\alpha\Delta t})\,\delta(T-t). \tag{18}$$

The application of the assumptions [Eq. (13)] to the expectation value [Eq. (8)] is[33]

$$E\{r(t)\} \approx \frac{1}{2M\alpha}e^{-2\alpha t_{\text{on}}}(1-e^{-2\alpha\Delta t})\cdot g(T-t). \tag{19}$$

First, it is seen that the expectation value is a scaled version of the input function and does not increase in amplitude as t_g increases (as will be the case for the variance). Second, as the window shifts to later times (i.e., t_{on} increases), the peak amplitude of the expectation value drops off exponentially. This is a result of the following. First, the waveform transmitted into the system, $(1/M)W(T-t)(g(T-t)\otimes_t h(-t))$, is always the same magnitude since it is normalized by M. It is also always the same shape since the exponential decay function is self-similar. Second, a signal is not obtained until this waveform is correlated with the portion of $h(t)$ from which it came, the magnitude of this portion being $e^{-\alpha t_{\text{on}}}$. Therefore, one would expect the convolution of $(1/M)W(T-t)(g(T-t)\otimes_t h(-t))$ and $h(t)$ to scale as $(1/M)e^{-\alpha t_{\text{on}}}$. As the window width increases, the peak amplitude of the expectation value grows. This being due to the fact that more signal is included in the pulse-compression that occurs during time-reversal. Since the IR falls off exponentially, the contribution naturally saturates. Finally, as the decay time increases, the peak increases due to more energy being included for a given window width. These dependences can be seen in Fig. 3.

B. Variance

1. Coherent variance

The simplification of the coherent variance based on the assumptions in Eq. (13) for $t_g > t_{\text{on}}$ is[33]

$$\mathrm{VAR}_{\text{coherent}} \approx \frac{\kappa_1}{4\alpha M^2}\begin{cases} 0 & \text{if } t < T-t_g \quad \text{or} \quad T < t \\ e^{4\alpha(T-t)}e^{-4\alpha t_{\text{on}}}(e^{-4\alpha(T-t_{\text{on}}-t)} - e^{-4\alpha\Delta t}) & \text{if } T-t_g \leq t \leq T-t_{\text{on}} \\ e^{4\alpha(T-t)}e^{-4\alpha t_{\text{on}}}(1-e^{-4\alpha\Delta t}) & \text{if } T-t_{\text{on}} \leq t < T, \end{cases} \tag{20}$$

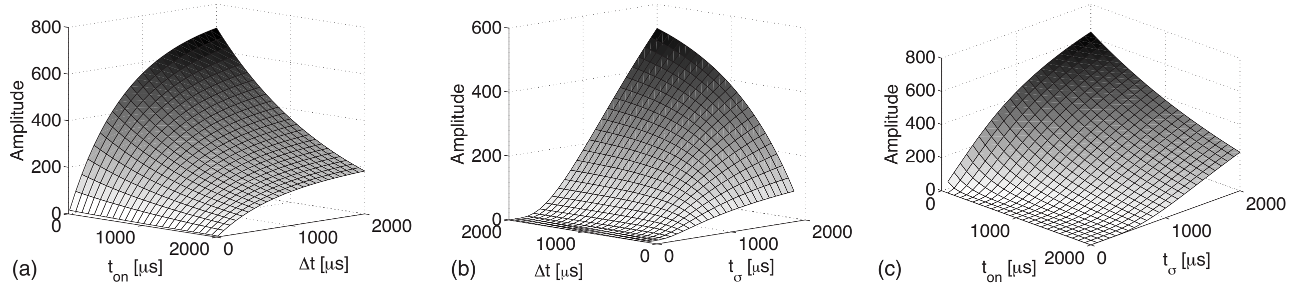


FIG. 3. The dependence of the expectation value on (a) window width Δt and placement t_{on} (choosing a decay time of $t_\sigma = 1500\ \mu$s), (b) window width Δt and decay constant t_σ (choosing a window placement of $t_{\text{on}} = 1000\ \mu$s), and (c) window placement t_{on} and decay constant t_σ (choosing a window width of $\Delta t = 1000\ \mu$s).

where $\kappa_1 = \int_{-t_g/2}^{t_g/2} g(T-(t+\tau))g(T-(t-\tau))d\tau$. If the duration of the input function is less than the window turn on time ($t_g \le t_{\text{on}}$), then[33]

$$\mathrm{VAR}_{\text{coherent}} \approx \frac{\kappa_1}{4\alpha M^2}\begin{cases} 0 & \text{if } t < T - t_g \quad \text{or} \quad T < t \\ e^{4\alpha(T-t)}e^{-4\alpha t_{\text{on}}}(1-e^{-4\alpha\Delta t}) & \text{if } T - t_g \le t < T. \end{cases} \tag{21}$$

From this result, it can clearly be seen that the coherent variance term only contributes at the main lobe. It is also seen that the amplitude saturates as the window width grows for the same reason as the expectation value, though the exact shape of this saturation varies depending on whether t_g or t_{on} is larger and is not the same as the expectation value. The amplitude also increases linearly with the decay constant $\tau_\sigma = 1/\alpha$. The square root of the coherent variance (i.e., the coherent standard deviation) decays similarly to $E\{r(t)\}$ as t_{on} increases. Finally, it is important to note that the coherent variance scales with κ_1, which increases as the pulse duration lengthens, independent of bandwidth.

2. Incoherent variance

The simplification of the incoherent variance based on the assumptions in Eq. (13) is[33]

$$\mathrm{VAR}_{\text{incoherent}} \approx = \frac{\kappa_2(0)}{4\alpha M^2}e^{2\alpha(T-2t_{\text{on}}-t)}\begin{cases} (e^{-4\alpha(T-t_{\text{on}}-t)} - e^{-4\alpha\Delta t}) & \text{if } t < T - t_{\text{on}} \\ (1-e^{-4\alpha\Delta t}) & \text{if } t \ge T - t_{\text{on}}, \end{cases} \tag{22}$$

where $\kappa_2(0)$ is approximated by the pulse-intensity integral (PII). Thus the incoherent variance scales with the PII. Noting that M will go as the maximum of $\sigma(t)$, which is approximately $e^{-\alpha t_{\text{on}}}$, it can be seen that the location of the window (i.e., t_{on}) does not change the peak magnitude of the incoherent variance term but rather shifts where it occurs. Specifically, as t_{on} increases, the time of the peak shifts in a linear manner. This contrasts with the expectation value and coherent variance, which do not shift in time as t_{on} changes but rather stay at the same location and decrease exponentially in amplitude. In general for $t < T - t_{\text{on}}$ the incoherent variance increases with t as a result of more energy being transmitted into the system as described earlier. It then peaks at $t = T - t_{\text{on}}$ and decays exponentially for larger t since no additional energy is being transmitted into the system at this point. It is also clear that the incoherent variance grows as both the window width and decay constant increase in the same manner as the coherent variance.

3. Total variance

The total variance can now be obtained from

$$\mathrm{VAR} = \mathrm{VAR}_{\text{coherent}} + \mathrm{VAR}_{\text{incoherent}}. \tag{23}$$

When $t_{\text{on}} < t_g$,

$$\mathrm{VAR} \approx \frac{1}{4\alpha M^2}\begin{cases} \kappa_2(0)e^{2\alpha(T-2t_{\text{on}}-t)}(e^{-4\alpha(T-t_{\text{on}}-t)} - e^{-4\alpha\Delta t}) & \text{if } t \le T - t_g \\ (\kappa_2(0) + \kappa_1 e^{2\alpha(T-t)})e^{2\alpha(T-2t_{\text{on}}-t)}(e^{-4\alpha(T-t_{\text{on}}-t)} - e^{-4\alpha\Delta t}) & \text{if } T - t_g \le t \le T - t_{\text{on}} \\ (\kappa_2(0) + \kappa_1 e^{2\alpha(T-t)})e^{2\alpha(T-2t_{\text{on}}-t)}(1-e^{-4\alpha\Delta t}) & \text{if } T - t_{\text{on}} \le t < T \\ \kappa_2(0)e^{2\alpha(T-2t_{\text{on}}-t)}(1-e^{-4\alpha\Delta t}) & \text{if } T < t \end{cases} \tag{24}$$

and when $t_g \le t_{\text{on}}$,

$$\mathrm{VAR} \approx \frac{1}{4\alpha M^2}\begin{cases} \kappa_2(0)e^{2\alpha(T-2t_{\mathrm{on}}-t)}(e^{-4\alpha(T-t_{\mathrm{on}}-t)} - e^{-4\alpha\Delta t}) & \text{if } t < T - t_{\mathrm{on}} \\ \kappa_2(0)e^{2\alpha(T-2t_{\mathrm{on}}-t)}(1 - e^{-4\alpha\Delta t}) & \text{if } T - t_{\mathrm{on}} \leq t < T - t_g \\ (\kappa_2(0) + \kappa_1 e^{2\alpha(T-t)})e^{2\alpha(T-2t_{\mathrm{on}}-t)}(1 - e^{-4\alpha\Delta t}) & \text{if } T - t_g \leq t < T \\ \kappa_2(0)e^{2\alpha(T-2t_{\mathrm{on}}-t)}(1 - e^{-4\alpha\Delta t}) & \text{if } T < t. \end{cases} \quad (25)$$

Figure 4 shows when each of the conditions in the piecewise function contributes. The relative magnitude of the coherent and incoherent terms is determined by $(\kappa_2(0) + \kappa_1 e^{4\alpha(T-t)})$. When $T-\tau_\sigma/4\ \ln(\kappa_2(0)/\kappa_1) < t$ the incoherent term is larger. Recalling that there will only be a contribution from coherent term for $T-t_g < t$ and that $t_g \ll t_\sigma$, the condition $T-\tau_\sigma/4\ \ln(\kappa_2(0)/\kappa_1) < t$ will always be satisfied. Thus the contribution from the incoherent term will always be larger by approximately $\kappa_2(0)/\max_{\mathrm{time}}\{\kappa_1\}$. For tone-bursts greater than 5-cycles this ratio is approximately 4. The change in the total variance as a function of the decay constant, window width, window placement, and length of a tone-burst input function can be seen in Fig. 5. The broad portion of the plot being due to the incoherent variance and the sharp peak around 3000 μs being the coherent variance. The dependence of both the coherent and incoherent variances on t_g is of particular note since it increases rapidly, independent of the input signal's bandwidth.

C. SNR

The SNR is defined as the ratio of the main lobe peak-to-peak amplitude to the side lobe standard deviation. Outside the main lobe only the incoherent term contributes.

$$\mathrm{SNR} = \frac{\max\{E\{r(t)\}\} - \min\{E\{r(t)\}\}}{\sqrt{\mathrm{VAR}_{\mathrm{incoherent}}\{r(t)\}}}. \quad (26)$$

Based on the nature of the directivity pattern, the above equation describes both the SNR as a function of time at the focus and SNR associated with the directivity pattern. Applying Eq. (13) and simplifying,

$$\mathrm{SNR}(t) = \frac{(\max\{g(T-t)\} - \min\{g(T-t)\})}{\sqrt{\alpha\kappa_2(0)}e^{\alpha(T-t)}}\begin{cases} \dfrac{1 - e^{-2\alpha\Delta t}}{\sqrt{(e^{-4\alpha(T-t_{\mathrm{on}}-t)} - e^{-4\alpha\Delta t})}} & \text{if } t < T - t_{\mathrm{on}} \\ \dfrac{1 - e^{-2\alpha\Delta t}}{\sqrt{(1 - e^{4\alpha\Delta t})}} & \text{if } t \geq T - t_{\mathrm{on}}. \end{cases} \quad (27)$$

As a function of time, it is seen that the SNR has a minimum at $t=T-t_{\mathrm{on}}$ when the functional dependence on t switches from $e^{-4\alpha(T-t_{\mathrm{on}}-t)}$ to 1. Thus, it should be noted that this minimum shifts with the window placement, while the center of the main lobe is always at $t=T$ (i.e., it does not change as the window placement changes). The window width also impacts the SNR. Both $(1-e^{-2\alpha\Delta t})$ and $\sqrt{(1-e^{-4\alpha\Delta t})}$ saturate to 1 as Δt grows, but the latter saturates more quickly. Since it is the denominator, the SNR will monotonically increase as the window width increases. Finally, it is seen that the SNR is inversely proportional to the $\sqrt{\kappa_2(0)} \approx \sqrt{\mathrm{PII}}$. The PII increases as the length of a pulse grows. Thus the SNR will decrease as the pulse length increases. It is important to note that the above SNR equation was derived for an arbitrary input function $g(t)$ with the only constraint on $g(t)$ being that $t_g \ll \tau_\sigma$. Thus this result is entirely independent of bandwidth. It has previously been seen that time-reversal requires a broadband signal; otherwise it becomes simple monochromatic phase conjugation. Good time-reversal focusing still requires broadband signals, but this is only a necessary and not a sufficient condition for good focusing. The result above shows that the PII (and thus pulse length) must also be small. The presence of $\kappa_2(0)$ comes from the incoherent variance term. It results from the fact that as the PII increases, $g(t)\otimes_t h(t)$ increases. Physically this can be associated with the long-range correlations created by the longer input function in the scattered signal. Previous work indicated that these long-range correlations come from the multiple-scattering medium; however, it is now evident that they may also arise due to the input signal.

Two time points are of particular interest, the minimum SNR ($t=T-t_{\mathrm{on}}$) and the SNR near the main lobe ($t \approx T$). For $t=T-t_{\mathrm{on}}$, the SNR becomes

$$\mathrm{SNR}_{\mathrm{min}} = \frac{1}{\sqrt{\alpha\kappa_2(0)}}\frac{1 - e^{-2\alpha\Delta t}}{\sqrt{(1 - e^{4\alpha\Delta t})}}\frac{(\max\{g(T-t)\} - \min\{g(T-t)\})}{e^{\alpha t_{\mathrm{on}}}}. \quad (28)$$

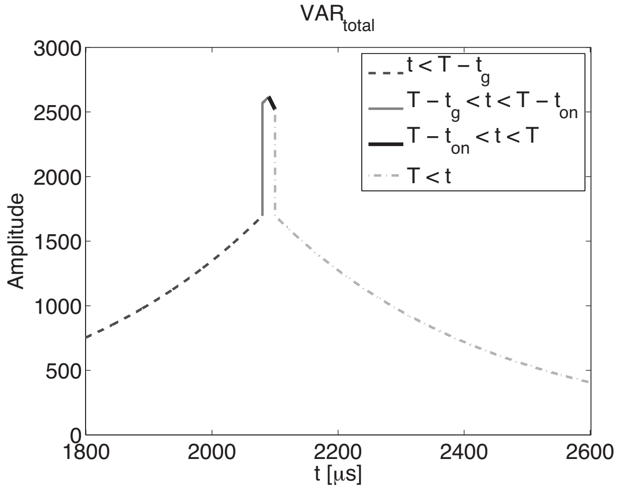


FIG. 4. The total variance plotted with each portion of the piecewise function shown with a different line style/color.

The directivity patterns can be defined by plotting the maximum signal over time at a particular location. SNR$_{\min}$ does this. Therefore Eq. (28) can also be used to describe how the SNR of the directivity pattern will change. Figure 6 shows the dependence of the minimum SNR on the decay constant, window width, window placement, and duration of $g(t)$ for a tone-burst input. Note that a negative SNR, on the decibel scale used, implies that the magnitude of the noise is greater than the signal itself. Previously it was mentioned that both the expectation value and the variance increase with τ_σ and Δt. Figure 6 shows, however, that the expectation value must increase faster since the SNR increases. Not surprisingly the SNR decreases with pulse length (independent of bandwidth) since the expectation value has no amplitude dependence on t_g but the variance does. Finally, it is seen that SNR$_{\min}$ decreases as t_{on} shifts. This is expected since the maximum of the incoherent variance does not change with t_{on}, but the expectation value decreases.

The SNR near the main lobe can be approximated from Eq. (27) evaluated at $t=T$ rather than $t=T\pm t_g/2$ because the incoherent variance is a slowly changing function on the time scale of t_g. In this case

$$\text{SNR}_{\text{near ML}} = \frac{1}{\sqrt{\alpha\kappa_2(0)}}\frac{1-e^{-2\alpha\Delta t}}{\sqrt{(1-e^{4\alpha\Delta t})}}(\max\{g(T-t)\} - \min\{g(T-t)\}). \tag{29}$$

The trends are also shown in Fig. 6. It is seen that the SNR near the main lobe is very similar to SNR$_{\min}$, except that it has no dependence on the window placement.

Minimizing the ratio of the standard deviation to the expectation value [the coefficient of variance (CV)] at the main lobe is desirable so the total amplitude at the focus is predictable. Recalling the assumptions [Eq. (13)],

$$\text{CV} = \sqrt{\alpha}e^{\alpha(T-t)} \times \frac{\sqrt{\kappa_2(0)(1+e^{-2\alpha\Delta t})+\kappa_1 e^{2\alpha(T-t)}}}{\sqrt{(1-e^{-2\alpha\Delta t})}\cdot g(T-t)}\Bigg|_{t\in[T-t_g,T]}. \tag{30}$$

Recalling that

$$\kappa_1 = \int_{-t_g/2}^{t_g/2} g(T-(t+\tau))g(T-(t-\tau))d\tau,$$

$$\kappa_2(0) \approx \int_0^{t_g} g^2(\tau)d\tau,$$

$$\text{CV} \approx \sqrt{\alpha}e^{\alpha(T-t)}\frac{\sqrt{(1+e^{-2\alpha\Delta t})\int_0^{t_g} g^2(\tau)d\tau + e^{2\alpha(T-t)}\int_{-t_g/2}^{t_g/2} g(T-(t+\tau))g(T-(t-\tau))d\tau}}{\sqrt{(1-e^{-2\alpha\Delta t})}\cdot g(T-t)}\Bigg|_{t\in[T-t_g,T]}. \tag{31}$$

Both terms in the numerator increase as the total "on-time" and amplitude of the input $g(t)$ grow, whereas the denominator only increases with the amplitude. As the window shifts to later times, the second term in the denominator grows, increasing the CV. This originates from the fact that as the window shifts to later times, the expectation value decreases. Finally, for small window widths, the numerator remains finite, while the denominator goes to zero. For large windows, all the terms with Δt go to zero, and thus no longer contribute. Therefore the CV gets very large for small window widths and decreases to a saturation value as the window width grows. The rate of this saturation depends on α. Also note that the CV decreases as the decay time constant $(1/\alpha)$ increases and that the window placement has no impact on the CV. These results are in Fig. 6 where the maximum CV (over time) is plotted.

IV. CONCLUSIONS

TRA has been a highly-successful method of focusing sound through high-order multiple-scattering media and has found application in many fields.[20–24,26–31] An initial shot

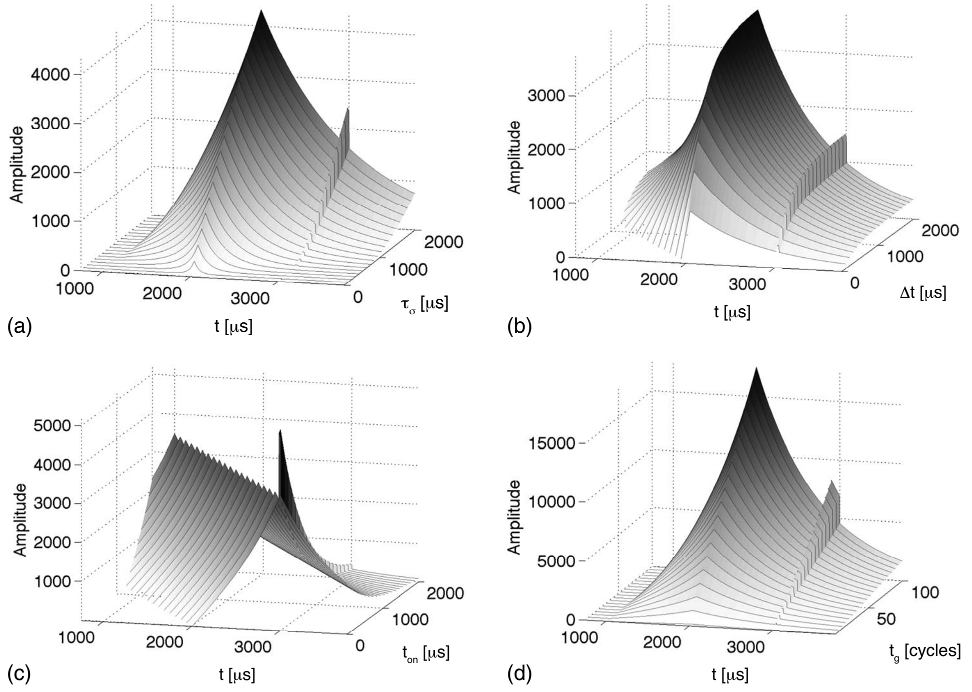


FIG. 5. The total variance is shown as a function of time and how this dependence changes with (a) the decay time constant τ_σ, (b) the window width Δt, (c) the window placement t_{on}, and (d) the duration of tone-burst input t_g (using τ_σ=1500 μs, t_{on}=1000 μs, Δt=1000 μs, and t_g=20 cycles as the respective constants).

noise model has been proposed by Derode *et al.*[4] to describe the expected signal and noise of this process. This model has been extended so that it applies to arbitrary input signals and windowing. The equations resulting from the extended model are novel and they confirmed previous predictions and also provided new predictions and explanations. This includes an explanation of the origin of the noise observed in multiple-scattering time-reversal (coherent versus incoherent contributions). Additionally it predicted that windowing can cause the peak of the noise to occur far from the main lobe and that the SNR depends on the PII of the original input signal.

The relatively complex results of Sec. II were then simplified for a set of parameters commonly found in experimental work. This in turn has allowed many trends to be identified. In particular, increasing the length of the input function degrades the SNR, independent of bandwidth. This is the result of the expectation value not depending on t_g, but the variance increasing with t_g due to long-range correlations in the scattered signal $g(t)\otimes_t h(t)$ due to $g(t)$. Additionally, it is seen that time-reversing later windows of a recorded signal does not affect the SNR near the main lobe, but it does reduce the main lobe to side lobe ratio for side lobes far from the main lobe [Fig. 2(d)]. Many of these results are also

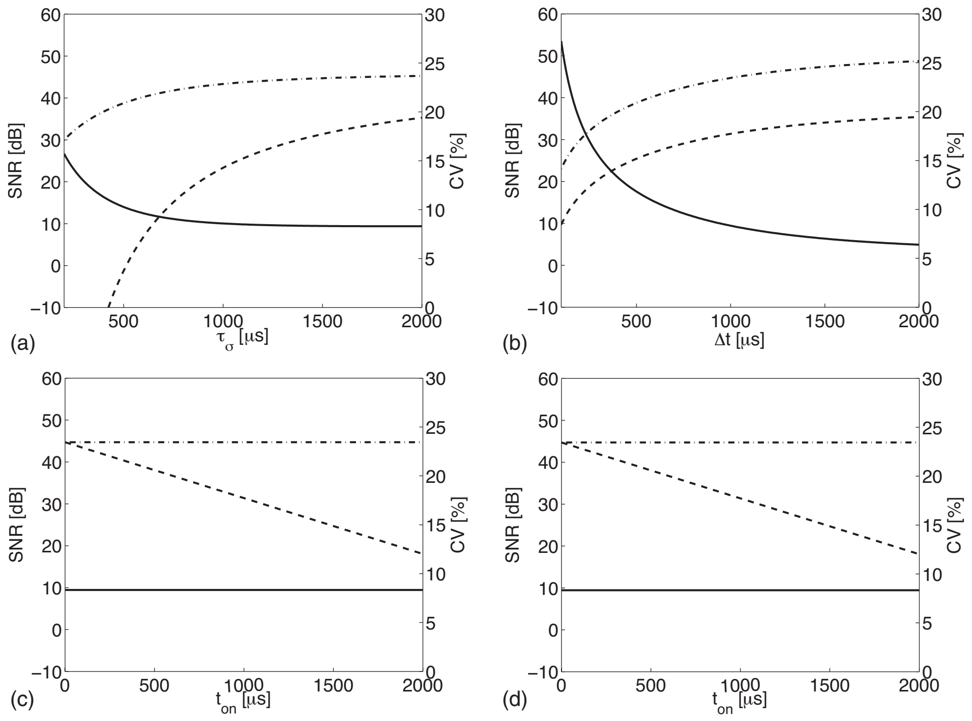


FIG. 6. The dependence of the SNR_{min} (dashed line), $\text{SNR}_{\text{near ML}}$ (dash-dot line), and CV_{max} (solid line) on (a) the decay time constant, (b) the window width, (c) the window placement, and (d) the duration of a tone-burst input (using t_σ=1500 μs, t_{on}=1000 μs, Δt=1000 μs, and t_g=20 cycles as the respective constants).

qualitatively similar to those seen for 1-bit time-reversal and an extension of this model to 1-bit time-reversal would be interesting but beyond the scope of this article.

Further analysis, specific to a particular application, can be done with these equations. This could include trying to maximize the SNR and minimize the CV for short pulses used in imaging. Alternatively, one might be interested in determining specific configurations that maximize the SNR and amplitude for long pulses that are used in thermal high-intensity focused ultrasound or acoustic radiation force experiments among other applications. Determining signals that maximize the amplitude for short pulses[19] may be useful for histotripsy.[38] One approach for imaging may be to use coded pulses, which would effectively doubly encode the signal, first with coded pulse and second with the reverberations in the multiple-scattering media.

In addition to looking at how the input function impacts the time-reversal focused signal, it is possible to do time-reversal using multiple channels (i.e., multiple transmitter-receiver pairs). It has been shown that if multiple pairs are used to transmit the same information, the SNR increases.[39] Recognizing that the main lobe will add coherently, its amplitude will increase proportional to the number of channels, the side lobes, however, will add incoherently (based on the physical interpretation provided) and thus will only go as the square root of the number of channels. This has been verified by Derode *et al.*[5]

In addition, it has been proposed that if multiple transmitters and receivers are used simultaneously, one can increase the bit-rate of sending information.[40] This is because the multiple-scattering medium makes each set of paths from receiver to transducer unique and independent (to first approximation, ignoring weak localization effects, recurrent scattering, correlated scatters, etc.). While it may be possible to increase the bit-rate, it is important to verify that the amount of noise (i.e., variance in the signal) does not dominate the signal (i.e., expectation value) under the conditions being used. The work of this paper provides a method for estimating these parameters based on its derivation of signal and noise of transmitting arbitrary pulses through a multiple-scattering medium. In particular it is found that the physical explanation of the incoherent (uncorrelated) noise term will be the same for each transmitter-receiver pair and will add as uncorrelated noise does, thus increasing the noise-floor for all channels as the square of the rms of the signal.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Matthew Schipper and Dr. Albert Levine for discussions on statistics. The authors would also like to thank the reviewers for their helpful and thought provoking comments. This work was supported in part by NIH Grant Nos. 5R01EB000281 and 1R21CA116043.

[1] A. Parvulescu and C. S. Clay, "Reproducibility of signal transmissions in the ocean," Radio Electron. Eng. **29**, 223–228 (1965).

[2] A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**, 4206–4210 (1995).

[3] C. Draeger and M. Fink, "One channel time-reversal in chaotic cavities: Theoretical limits," J. Acoust. Soc. Am. **105**, 611–617 (1999).

[4] A. Derode, A. Tourin, and M. Fink, "Ultrasonic pulse compression with one-bit time reversal through multiple scattering," J. Appl. Phys. **85**, 6343–6352 (1999).

[5] A. Derode, A. Tourin, and M. Fink, "Limits of time-reversal focusing through multiple scattering: Long-range correlation," J. Acoust. Soc. Am. **107**, 2987–2998 (2000).

[6] V. Leroy and A. Derode, "Temperature-dependent diffusing acoustic wave spectroscopy with resonant scatterers," Phys. Rev. E **77**, 036602 (2008).

[7] A. Derode, A. Tourin, and M. Fink, "Random multiple scattering of ultrasound. II. Is time reversal a self-averaging process?," Phys. Rev. E **64**, 036606 (2001).

[8] B. A. van Tiggelen, "Green function retrieval and time reversal in a disordered world," Phys. Rev. Lett. **91**, 243904 (2003).

[9] J. de Rosny, A. Tourin, A. Derode, B. A. van Tiggelen, and M. Fink, "Relation between time reversal focusing and coherent backscattering in multiple scattering media: A diagrammatic approach," Phys. Rev. E **70**, 046601 (2004).

[10] P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," J. Acoust. Soc. Am. **111**, 230 (2002).

[11] G. Ribay, J. de Rosny, and M. Fink, "Time reversal of noise sources in a reverberation room," J. Acoust. Soc. Am. **117**, 2866–2872 (2005).

[12] C. Draeger and M. Fink, "One-channel time reversal of elastic waves in a chaotic 2D-silicon cavity," Phys. Rev. Lett. **79**, 407–410 (1997).

[13] C. Draeger, J.-C. Aime, and M. Fink, "One-channel time-reversal in chaotic cavities: Experimental results," J. Acoust. Soc. Am. **105**, 618–625 (1999).

[14] N. Quieffin, S. Catheline, R. K. Ing, and M. Fink, "Real-time focusing using an ultrasonic one channel time-reversal mirror coupled to a solid cavity," J. Acoust. Soc. Am. **115**, 1955–1960 (2004).

[15] G. Montaldo, N. Perez, C. Negreira, and M. Fink, "The spatial focusing of a leaky time reversal chaotic cavity," Waves Random Complex Media **17**, 67–83 (2007).

[16] G. Montaldo, P. Roux, A. Derode, C. Negreira, and M. Fink, "Ultrasound shock wave generator with one-bit time reversal in a dispersive medium, application to lithotripsy," Appl. Phys. Lett. **80**, 897–899 (2002).

[17] A. Y. Sutin, E. Roides, and A. P. Sarvazyan, "Damage detection in composites using the time-reversal acoustics method (A)," J. Acoust. Soc. Am. **116**, 2567 (2004).

[18] Y. D. Sinelnikov, A. Y. Sutin, and A. P. Sarvazyan, "Time-reversal acoustic focusing with liquid resonator for medical applications," Sixth International Symposium on Therapeutic Ultrasound (2006), Vol. **911**, pp. 82–86.

[19] L. Fillinger, A. Y. Sutin, and A. P. Sarvazyan, "Time reversal focusing of short pulses," IEEE Ultrasonics Symposium (2007), pp. 220–223.

[20] M. Fink, G. Montaldo, and M. Tanter, "Time-reversal acoustics in biomedical engineering," Annu. Rev. Biomed. Eng. **5**, 465–497 (2003).

[21] A. S. Gliozzi, M. Griffa, and M. Scalerandi, "Efficiency of time-reversed acoustics for nonlinear damage detection in solids," J. Acoust. Soc. Am. **120**, 2506–2517 (2006).

[22] S. D. Kim, C. W. In, K. E. Cronin, H. Sohn, and K. Harries, "Reference-free ndt technique for debonding detection in cfrp-strengthened rc structures," J. Struct. Eng. **133**, 1080–1091 (2007).

[23] E. Larose, "Mesoscopics of ultrasound and seismic waves: Application to passive imaging," Ann. Phys. (Paris) **31**, 1–126 (2006).

[24] C. Larmat, J.-P. Montagner, M. Fink, Y. Capdeville, A. Tourin, and E. Clevede, "Time-reversal imaging of seismic sources and application to the great Sumatra earthquake," Geophys. Res. Lett. **33**, L19312 (2006).

[25] E. Larose, P. Roux, M. Campillo, and A. Derode, "Fluctuations of correlations and Green's function reconstruction: Role of scattering," J. Appl. Phys. **103**, 114907 (2008).

[26] G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Oceanic Eng. **27**, 602–609 (2002).

[27] P. Roux, A. Derode, A. Peyre, A. Tourin, and M. Fink, "Acoustical imaging through a multiple scattering medium using a time-reversal mirror," J. Acoust. Soc. Am. **107**, L7–L12 (2000).

[28] G. Montaldo, D. Palacio, M. Tanter, and M. Fink, "Building three-dimensional images using a time-reversal chaotic cavity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1489–1498 (2005).

[29] J. V. Candy, A. J. Poggio, D. H. Chambers, B. L. Guidry, C. L. Robbins, and C. A. Kent, "Multichannel time-reversal processing for acoustic communications in a highly reverberant environment," J. Acoust. Soc. Am. **118**, 2339–2354 (2005).

[30] R. C. Qiu, C. Zhou, N. Guo, and J. Q. Zhang, "Time reversal with miso for ultrawideband communications: Experimental results," IEEE Antennas

Wireless Propag. Lett. **5**, 269–273 (2006).

[31] G. Lerosey, J. de Rosny, A. Tourin, and M. Fink, “Focusing beyond the diffraction limit with far-field time reversal,” Science **315**, 1120–1122 (2007).

[32] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Series in Systems Science (McGraw-Hill, New York, 1984).

[33] K. J. Haworth, “Medical ultrasound aberration correction via acoustic droplet vaporization and time-reversal acoustics,” Ph.D. thesis, University of Michigan, Ann Harbor, MI (2009).

[34] A. Sarvazyan, “Ultrasonic transducers for imaging and therapy based on time-reversal principles,” J. Acoust. Soc. Am. **123**, 3429 (2008).

[35] G. Marsaglia and W. W. Tsang, “The Ziggurat method for generating random variables,” J. Stat. Software **5**, 1–7 (2000).

[36] J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).

[37] A. Tourin, A. Derode, A. Peyre, and M. Fink, “Transport parameters for an ultrasonic pulsed wave propagating in a multiple scattering medium,” J. Acoust. Soc. Am. **108**, 503–512 (2000).

[38] W. W. Roberts, T. L. Hall, K. Ives, J. S. Wolf, J. B. Fowlkes, and C. A. Cain, “Pulsed cavitational ultrasound: A noninvasive technology for controlled tissue ablation (histotripsy) in the rabbit kidney,” J. Urol. **175**, 734–738 (2006).

[39] M. Fink, “Acoustic time-reversal mirrors,” Top. Appl. Phys. **84**, 17–42 (2002).

[40] A. Derode, A. Tourin, J. de Rosny, M. Tanter, S. Yon, and M. Fink, “Taking advantage of multiple scattering to communicate with time-reversal antennas,” Phys. Rev. Lett. **90**, 014301 (2003).

Application of Renyi entropy for ultrasonic molecular imaging

M. S. Hughes, J. N. Marsh, J. M. Arbeit, R. G. Neumann, R. W. Fuhrhop, K. D. Wallace,
L. Thomas, J. Smith, K. Agyem, G. M. Lanza, and S. A. Wickline
Department of Medicine, Cardiovascular Division, Washington University School of Medicine,
Campus Box 8086, 660 South Euclide Avenue, St. Louis, Missouri 63110-1093

J. E. McCarthy
Department of Mathematics, Washington University in St. Louis, Cupples I Hall, One Brookings Drive,
St. Louis, Missouri 63130

(Received 18 November 2008; revised 12 February 2009; accepted 13 February 2009)

Previous work has demonstrated that a signal receiver based on a limiting form of the Shannon entropy is, in certain settings, more sensitive to subtle changes in scattering architecture than conventional energy-based signal receivers [M. S. Hughes *et al.*, J. Acoust. Soc. Am. **121**, 3542–3557 (2007)]. In this paper new results are presented demonstrating further improvements in sensitivity using a signal receiver based on the Renyi entropy.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3097489]

PACS number(s): 43.60.Bf, 43.60.Lq, 43.60.Cg [EJS] Pages: 3141–3145

I. INTRODUCTION

In an earlier paper we reported on the comparison between a Shannon entropy analog, H_f, and more conventional signal processing techniques, i.e., signal energy and its logarithm as applied to beam formed radiofrequency (rf) data. Both analysis techniques were applied to data obtained in backscatter measurements from nanoparticle targeted neovasculature.[1] The comparison study was undertaken after a preliminary conventional B-mode grayscale analysis of the data was unable to detect changes in backscattered rf arising from the accumulation of targeted nanoparticles in the neovasculature in the insonified region. This result implied that acoustic characterization of sparse collections of targeted perfluorocarbon nanoparticles presented challenges that might require the application of novel types of signal processing. We were able to show that signal processing based on a "moving window" H_f analysis could distinguish the difference in backscatter measured at 15 and 60 min and (although it was not stressed) able to detect accumulation of targeted nanoparticles 30 min post-injection. The signal energy, defined as the sum of squares of the signal amplitude over the same moving window, was unable to distinguish measurements made at any time during the 1 h experiment.

We stress that, although entropy-based techniques have a long history for image enhancement and postprocessing of reconstructed images, the approach we have taken in previous studies is different in that entropy is used directly as the quantity defining the pixel values in the image. Specifically, images were reconstructed by computing H_f for segments of the individual rf A-lines that comprise a typical medical image by applying a moving window, or "box-car," analysis facilitating local estimation of entropy values for regions within the image.

II. APPROACH

All rf data are obtained by sampling a continuous function, $y=f(t)$, and subsequently using the sampled values to compute its associated density function, $w_f(y)$.

A. The function $w_f(y)$

The density function $w_f(y)$ may be used to compute either the entropy H_f, as in previous studies, or the Renyi entropy as we do here. It corresponds to the density functions used in statistical signal processing. From it, other mathematical quantities are subsequently derived (e.g., mean values, variances, and covariances).[2–4] While the density function is usually assumed to be continuous, infinitely differentiable, and to approach zero at infinity in statistical signal processing of random signals, in our application $w_f(y)$ has (integrable) singularities.

As in previous studies, we employ the convention that the domain of $f(t)$ is $[0,1]$, so that, $w_f(y)$, the density function of $f(t)$, can be defined by the basic integral relation

$$\int_0^1 \phi(f(t))dt = \int_{f_{\min}}^{f_{\max}} \phi(y) w_f(y) dy. \tag{1}$$

Equation (1) implies

$$w_f(\xi) = \sum_{\{t_k | f(t_k)=\xi\}} \frac{1}{|f'(t_k)|}, \tag{2}$$

either by breaking the integral into a sum over intervals of monotonicity of $f(t)$ ("laps") and changing variables, or by choosing $\phi(y)$ to be a Dirac delta function and using the well-known expansion formula for a delta function of a function.[5] All of our digitized waveforms $f(t)$ are comprised of at least one monotonic section, or "lap." The lap boundaries are just the points t where $f'(t)=0$. Within a lap, $f(t)$ has a well-defined inverse function so that Eq. (2) may be rewritten as

$$w_f(y) = \sum_{k=1}^{N} |g_k'(y)|, \tag{3}$$

where N is the number of laps, $g_k(y)$ is the inverse of $f(t)$ in the kth-lap, and if y is not in the range of $f(t)$ in the kth-lap, $g_k'(y)$ is taken to be 0.

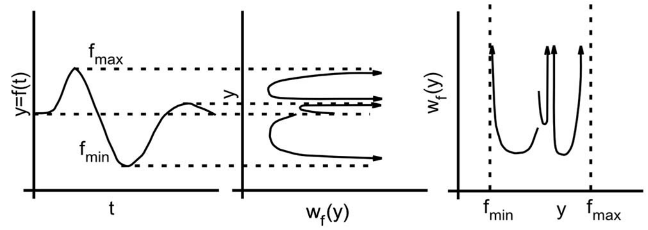


FIG. 1. A time-domain waveform, $f(t)$, with three critical points (left), and its associated density function $w_f(y)$ which has three corresponding (integrable) singularities.

We also assume that all experimental waveforms $f(t)$ have a Taylor series expansion valid in [0,1]. Then near a time t_k such that $f'(t_k)=0$

$$y = f(t) = f(t_k) + \tfrac{1}{2!}f''(t_k)(t-t_k)^2 + \cdots, \tag{4}$$

t_k is a lap boundary and on the left side of this point Eq. (4) may be truncated to second order and inverted to obtain

$$g_k(y) \sim t_k \pm \sqrt{2(y-f(t_k))/f''(t_k)}, \tag{5}$$

with

$$|g_k'(y)| \sim 1/\sqrt{2f''(t_k)(y-f(t_k))}. \tag{6}$$

The contribution to $w_f(y)$ from the right side of the lap boundary, from $g_{k+1}(y)$, is the same, so that the overall contribution to $w_f(y)$ coming from the time interval around t_k is

$$|g_k'(y)| \sim \sqrt{2/(f''(t_k)(y-f(t_k)))}, \tag{7}$$

for $0 < f(t_k) - y \ll 1$ for a maximum at $f(t_k)$ and $0 < y - f(t_k) \ll 1$ for a minimum. Thus, $w_f(y)$ has only a square root singularity (we have assumed that t_k is interior to the interval [0,1]; if not, then the contributions to w_f come from only the left or the right). If additionally, $f''(t_k)=0$, then the square root singularity in Eq. (6) will become a cube-root singularity, and so on, so that the density functions we consider will have only integrable algebraic singularities.

Figure 1 illustrates, schematically, one possible type of behavior possible in $w_f(y)$: both discontinuities and algebraic singularities [indicated by arrows on the plots of $w_f(y)$]. Progressing from left to right in the figure illustrates how to estimate qualitative features of $w_f(y)$ from $f(t)$. For instance, the maxima in $f(t)$ correspond to algebraic singularities in $w_f(y)$, plotted sideways in the middle panel to more clearly indicate the relationship between its features and those of $f(t)$. The rightmost panel shows $w_f(y)$ in a conventional layout (a rotated and flipped version of the plot in the middle panel). These plots show that the density functions possess significantly different attributes from those usually considered in statistical signal processing.

The mathematical characteristics of the singularities are important in order to guarantee the existence of the following integral on which we base our analysis of signals in this study:

$$I_f(r) = \frac{1}{1-r}\log\left[\int_{f_{\min}}^{f_{\max}} w_f(y)^r dy\right], \tag{8}$$

which is known as the Renyi entropy.[6] The physical significance of the parameter r appearing in Eq. (8) may be interpreted by analogy with statistical mechanics where the probabilities $w_f(y)$ are given in terms of system energy levels according to

$$E(y) = \frac{1}{\mu_0}\log[w_f(y)] \tag{9}$$

(with μ_0 being a physical constant), and thermodynamic quantities are derived from the partition function

$$Z = \int e^{-\mu E(y)}dy = \int w_f(y)^{-\mu/\mu_0}dy, \tag{10}$$

where $\mu = 1/(kT)$, with k being Boltzmann's constant and T being temperature.[7] From the equations we see that the Renyi entropy, $I_f(r)$, is very similar to the partition function in statistical mechanics and that the parameter r is analogous to an inverse "temperature." Moreover, $I_f(r) \to -H_f$, as $r \to 1$, using L'Hôpital's rule, so that I_f is a generalization of H_f as follows:

$$H_f = \int_{f_{\min}}^{f_{\max}} w_f(y)\log w_f(y)dy, \tag{11}$$

which previous studies have shown can be more sensitive to subtle changes in scattering architecture than are more commonly used energy-based measures.[1] The purpose of this study is to show that further sensitivity improvements may be obtained using I_f at the suitable value of r.

For the density functions $w_f(y)$ encountered in our study, $I_f(r)$ is undefined for $r \geq 2$. Moreover, as $r \to 2^-$, the integral appearing in Eq. (8) will grow without bound due to the singularities in the density function, $w_f(y)$ [i.e., Eq. (7)]. The behavior as $r \to 2$ is dominated by contributions from the singularities. If the kth critical point is a minimum (the argument for a maximum is similar) the contribution to the integral in Eq. (8) is asymptotic to

$$\lim_{\epsilon\to 0}\int_{f(t_k)}^{f_{\max}}\left(\frac{a_k}{\sqrt{y-f(t_k)}}\right)^{2-\epsilon}dy. \tag{12}$$

This is equal to

$$= \lim_{\epsilon\to 0} a_k^{2-\epsilon}\int_{f(t_k)}^{f_{\max}}(y-f(t_k))^{1-\epsilon/2}dy,$$

$$= \lim_{\epsilon\to 0} a_k^{2-\epsilon}\left.\frac{(y-f(t_k))^{\epsilon/2}}{\epsilon/2}\right|_{f(t_k)}^{f_{\max}},$$

$$= \lim_{\epsilon\to 0} a_k^{2-\epsilon}\frac{(f_{\max}-f(t_k))^{\epsilon/2}}{\epsilon/2},$$

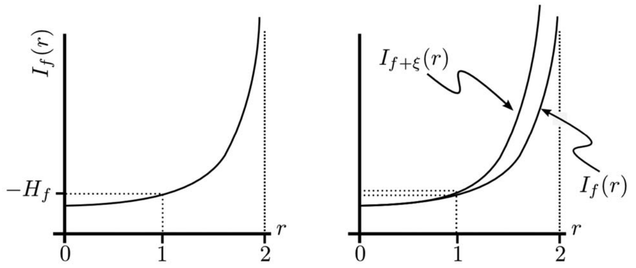


FIG. 2. Left panel: A plot of $I_f(r)$ (left) showing that $I_f(1)=-H_f$ and that $I_f(r)$ grows without bound as $r\rightarrow 2$. Right panel: Even though two similar waveforms $f(t)$ and $f(t)+\xi(t)$ may have nearly equal entropies, H_f, it is possible that as $r\rightarrow 2$ their Renyi entropies may diverge.

$$=\lim_{\epsilon\rightarrow 0}\frac{2a_k^2}{\epsilon}, \tag{13}$$

where $a_k=\sqrt{2/f''(t_k)}$ [for a maximum we have the asymptotic term $a_k=\sqrt{-2/f''(t_k)}$].

This behavior is shown in the left panel of Fig. 2. Moreover, as shown in the right panel, it is possible that two slightly different functions, $f(t)$ and $f(t)+\xi(t)$, where ξ is small, may have entropies, H_f and $H_{f+\xi}$ that are close, as shown in the figure, but whose Renyi entropies, $I_f(r)$ and $I_{f+\xi}(r)$, diverge as $r\rightarrow 2$. If this amplification effect is not dominated by noise, it may be exploited to distinguish subtly different functions, such as those obtained from measurements of backscattered ultrasound of targeted and nontargeted tissues. Our results show that this can happen in practice.

III. MATERIALS AND METHODS

A. Nanoparticles for molecular imaging

A cross-section of the spherical liquid nanoparticles used in our study is diagramed in Fig. 3. For *in vivo* imaging we formulated nanoparticles targeted to $\alpha_v\beta_3$-integrins of neovascularity in cancer by incorporating an "Arg-Gly-Asp" mimetic binding ligand into the lipid layer. Methods developed in our laboratories were used to prepare perfluorocarbon (perfluoro-octyl bromide, which remains in a liquid state at body temperature and at the acoustic pressures used in this study[8]) emulsions encapsulated by a lipid-surfactant monolayer.[9,10] The nominal sizes for each formulation were measured with a submicron particle analyzer (Malvern Zetasizer, Malvern Instruments). Particle diameter was measured at 200 ± 30 nm.

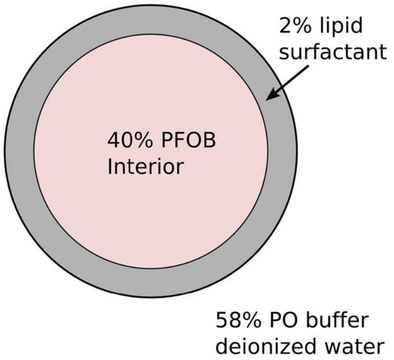


FIG. 3. (Color online) A cross-sectional diagram of the nanoparticles used in our study.

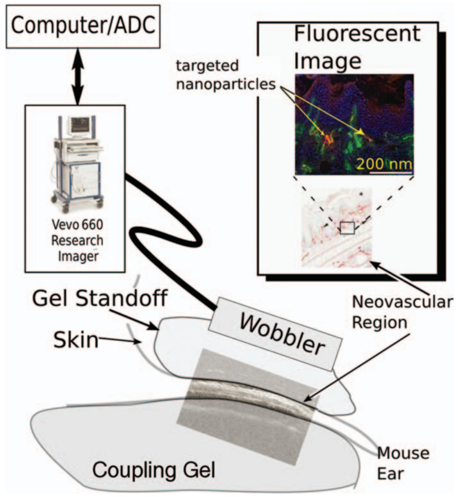


FIG. 4. A diagram of the apparatus used to acquire rf data backscattered from HPV mouse ears *in vivo* together with a histologically stained section of the ear indicating portions where $\alpha_v\beta_3$-targeted nanoparticles could adhere and a fluorescent image demonstrating presence of targeted nanoparticles.

B. Animal model

The study was performed according to an approved animal protocol and in compliance with the guidelines of the Washington University institutional animal care and use committee.

The model used is the transgenic K14-HPV16 mouse in which the ears typically exhibit squamous metaplasia, a precancerous condition, associated with abundant neovasculature that expresses the $\alpha_v\beta_3$-integrin. Eight of these transgenic mice[11,12] were treated with 1.0 mg/kg intravenous of either $\alpha_v\beta_3$-targeted nanoparticles ($n=4$) or untargeted nanoparticles ($n=4$) and imaged dynamically for 1 h using a research ultrasound imager modified to store digitized rf waveforms acquired at 0, 15, 30, and 60 min time points. In both targeted and untargeted cases, the mouse was placed on a heated platform maintained at 37 °C, and anesthesia was administered continuously with isoflurane gas (0.5%).

C. Ultrasonic data acquisition

A diagram of our apparatus is shown in Fig. 4. RF data were acquired with a research ultrasound system (Vevo 660, Visualsonics, Toronto, Canada), with an analog port and a sync port to permit digitization. The tumor was imaged with a 40 MHz single element "wobbler" probe and the rf data corresponding to single frames were stored on a hard disk for later off-line analysis. The frames (acquired at a rate of 30 Hz) consisted of 384 lines of 4096 eight-bit words acquired at a sampling rate of 500 MHz using a Gage CS82G digitizer card (connected to the analog-out and sync ports of the Vevo) in a controller PC. Each frame corresponds spatially to a region 0.8 cm wide and 0.3 cm deep.

The wobbler transducer used in this study is highly focused (3 mm in diameter) with a focal length of 6 mm and a theoretical spot size of 80×1100 μm (lateral beam width $\times$depth of field at -6 dB), so that the imager is most sensitive to changes occurring in the region swept out by the focal zone as the transducer is "wobbled." Accordingly, a gel standoff was used, as shown in Fig. 4, so that this region would contain the mouse ear.

A close-up view showing the placement of transducer, gel standoff, and mouse ear is shown in the bottom of Fig. 4. Superposed on the diagram is a B-mode gray scale image (i.e., logarithm of the analytic signal magnitude). Labels indicate the location of skin (top of image insert), the structural cartilage in the middle of the ear, and a short distance below this, the echo from the skin at the bottom of the ear. Directly above this is an image of a histological specimen extracted from a human papilloma virus (HPV) mouse model that has been magnified 20 times to permit better assessment of the thickness and architecture of the sites where $\alpha_v\beta_3$-targeted nanoparticle might attach (red by β_3 staining). Skin and tumor are both visible in the image. On either side of the cartilage (center band in image), extending to the dermal-epidermal junction, is the stroma. It is filled with neoangiogenic microvessels. These microvessels are also decorated with $\alpha_v\beta_3$ nanoparticles as indicated by the fluorescent image (labeled in the upper right of the figure) of a bisected ear from an $\alpha_v\beta_3$-injected K14-HPV16 transgenic mouse (Neumann *et al.*, unpublished). It is in this region that the $\alpha_v\beta_3$-targeted nanoparticles are expected to accumulate, as indicated by the presence of red β_3 stain in the magnified image of a histological specimen also shown in the image.

D. Ultrasonic data processing

Each of the 384 rf lines in the data was first upsampled from 4096 to 8192 points, using a cubic spline fit to the original data set in order to improve the stability of the thermodynamic receiver algorithms. Previous work has shown benefit from increased input waveform length.[13,14] Next, a moving window analysis was performed on the upsampled data set using a rectangular window that was advanced in 0.064 μs steps (64 points), resulting in 121 window positions within the original data set. This was done using both continuous entropy, H_f, and Renyi entropy, $I_f(1.99)$, analysis of the rf segments within each window in order to produce an image [either H_f or $I_f(1.99)$] for each time point in the experiment. As described previously, the density function, $w_f(y)$, used to compute H_f and $I_f(r)$ is computed using a Fourier series representation.[15] For this study, where the desire was to compute $I_f(r)$ as near to its singular value as possible, it was found that 16 384 terms were required for accurate estimation. In order to complete computations in a reasonable amount of time all calculations were performed on a Linux cluster using open message passing interface.

E. Image processing

All rf data were processed off-line to reconstruct images using information theoretic, either H_f or $I_f(1.99)$. Subsequently, a histogram of pixel values for the composite of the 0, 15, 30, and 60 min images was computed, either H_f or $I_f(1.99)$. Image segmentation of each type of image, either H_f or $I_f(1.99)$, at each time point in the experiment was then performed automatically using its corresponding histogram according to the following threshold criterion: The lowest 7% of pixel values were classified as "targeted" tissue, while the remaining were classified as "untargeted" (histogram analysis was also performed using 10% and 13% thresholds, with 7% having the best statistical separation between time points). The mean value of pixels classified as targeted was computed at each time post-injection.

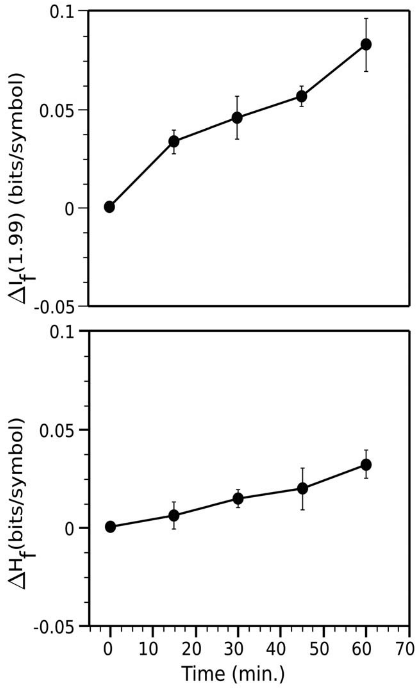


FIG. 5. A plot of average enhancement, i.e., change relative to value at time 0, obtained by analysis of $I_f(1.99)$ (top) and H_f (bottom) images from four HPV mice injected with $\alpha_v\beta_3$-targeted nanoparticles.

IV. RESULTS AND DISCUSSION

The results obtained after injection of targeted nanoparticles, by either the $I_f(1.99)$ or H_f receivers, are shown in the top and bottom panels of Fig. 5. Both panels compare the growth, with time, of the change (relative to 0 min) in mean value of receiver output in the enhanced regions of images obtained from all four of the animals in the targeted group. Standard error bars are shown with each point. At 15 min the change in mean value if $I_f(1.99)$ is more than two standard errors from zero, implying statistical significance at the 95% level. As the bottom panel shows it is 30 min before H_f is more than twice the standard error from zero.

The results obtained after injection of nontargeted nanoparticles, by either the $I_f(1.99)$ or H_f receivers, are shown in the top and bottom panels of Fig. 6. Neither receiver exhibits a statistically significant change in output over the course of the experiment.

The value of 1.99 was chosen after an initial round of numerical experimentation to assess numerical stability of receiver output [while also varying the number of terms required for the Fourier series reconstruction of $w_f(y)$] versus

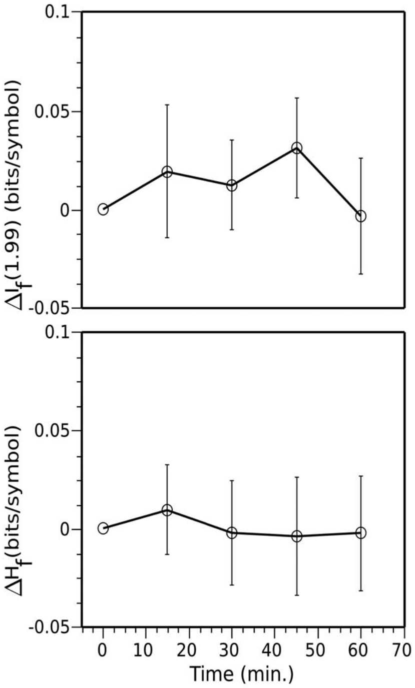


FIG. 6. A plot of average enhancement, i.e., change relative to value at time 0, obtained by analysis of $I_f(1.99)$ (top) and H_f (bottom) images from four HPV mice injected with nontargeted nanoparticles. Neither plot exhibits a statistically significant change.

computation time. As the goal is ultimately to develop an algorithm of clinical utility, the execution time of 1 week required to compute the $I_f(1.99)$ images for this study was taken as an upper acceptable bound. Comparison of the data in Figs. 5 and 6 show that $I_f(1.99)$ is able to detect accumulation of targeted nanoparticles in only half the time (post-injection) required by H_f. Pharmacokinetic dynamics would lead us to expect the steady increase in targeted nanoparticles in the region of insonification post-injection. Both plots of Fig. 5 are consistent with this model; however, we may conclude that $I_f(1.99)$ is more sensitive to their presence than H_f. Future studies will concentrate on increasing the computational efficiency of our algorithm so that the region closer to $r=2$ may be explored.

ACKNOWLEDGMENTS

This study was funded by NIH Grant Nos. EB002168, HL042950, and CO-27031 and NSF Grant No. DMS 0501079. The research was carried out at the Washington University Department of Mathematics and the School of Medicine.

[1] M. S. Hughes, J. E. McCarthy, J. N. Marsh, J. M. Arbeit, R. G. Neumann, R. W. Fuhrhop, K. D. Wallace, D. R. Znidersic, B. N. Maurizi, S. L. Baldwin, G. M. Lanza, and S. A. Wickline, “Properties of an entropy-based signal receiver with an application to ultrasonic molecular imaging,” J. Acoust. Soc. Am. **121**, 3542–3557 (2007).

[2] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes With Applications to Guidance* (Chelsea, New York, NY, 1987).

[3] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series* (Chelsea, New York, NY, 1984).

[4] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications* (MIT, Cambridge, MA, 1949).

[5] R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978).

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

[7] R. Tolman, *The Principles of Statistical Mechanics* (Dover, New York, NY, 1979).

[8] M. S. Hughes, J. N. Marsh, J. Arbeit, R. Neumann, R. W. Fuhrhop, G. M. Lanza, and S. A. Wickline, “Ultrasonic molecular imaging of primordial angiogenic vessels in rabbit and mouse models with $\alpha_v\beta_3$-integrin targeted nanoparticles using information-theoretic signal detection: Results at high frequency and in the clinical diagnostic frequency range,” Proceedings of the 2005 IEEE Ultrasonics Symposium (2005).

[9] S. Flacke, S. Fischer, M. J. Scott, R. J. Fuhrhop, J. S. Allen, M. McLean, P. Winter, G. A. Sicard, P. J. Gaffney, S. A. Wickline, and G. M. Lanza, “Novel MRI contrast agent for molecular imaging of fibrin implications for detecting vulnerable plaques,” Circulation **104**, 1280–1285 (2001).

[10] G. M. Lanza, K. D. Wallace, M. J. Scott, W. P. Cacheris, D. R. Abendschein, D. H. Christy, A. M. Sharkey, J. G. Miller, P. J. Gaffney, and S. A. Wickline, “A novel site-targeted ultrasonic contrast agent with broad biomedical application,” Circulation **94**, 3334–3340 (1996).

[11] J. M. Arbeit, R. R. Riley, B. Huey, C. Porter, G. Kelloff, R. Lubet, J. M. Ward, and D. Pinkel, “DFMO chemoprevention of epidermal carcinogenesis in k14-hpv16 transgenic mice,” Cancer Res. **59**, 3610–3620 (1999).

[12] J. M. Arbeit, K. Mnger, P. M. Howley, and D. Hanahan, “Progressive squamous epithelial neoplasia in k14-human papillomavirus type 16 transgenic mice,” J. Virol. **68**, 4358–4368 (1994).

[13] M. Hughes, “Analysis of digitized waveforms using Shannon entropy,” J. Acoust. Soc. Am. **93**, 892–906 (1993).

[14] M. Hughes, “Analysis of digitized waveforms using Shannon entropy. II. High-speed algorithms based on Green’s functions,” J. Acoust. Soc. Am. **95**, 2582–2588 (1994).

[15] M. Hughes, “Analysis of ultrasonic waveforms using Shannon entropy,” Proceedings of the 1992 IEEE Ultrasonics Symposium Vol. **2**, pp. 1205–1209 (1992).

Sound fields in generally shaped curved ear canals

H. Hudde and S. Schmidt
Institute of Communication Acoustics, Ruhr University Bochum, D-44780 Bochum, Germany

(Received 11 November 2008; revised 7 February 2009; accepted 9 February 2009)

The sound field in the external ear can be subdivided into a distinctly three-dimensional part in front of pinna and concha, a fairly regular part in the core region of ear canals, and a less regular part in the drum coupling region near the tympanic membrane. The different parts of the sound field and their interaction have been studied using finite elements. A "pinna box" enclosing the pinna provides both a realistic coupling of the external space to the ear canal and the generation of sound. The sound field in the core region turns out to be not that regular as mostly assumed: near pressure minima and maxima "one-sided" isosurfaces (surfaces of equal pressure magnitude) occur, which are inconsistent with the notion of a middle axis, in principle. Nevertheless such isosurfaces can be seen as part of a "fundamental sound field," which is governed by the principle of minimum energy. Actually, the sound transformation through narrow ducts is little affected by one-sided isosurfaces in between. As expected, the beginning of the core region depends on frequency. If the full audio range up to 20 kHz is to be covered, a location in the first bend of the ear canal is found.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097446]

PACS number(s): 43.64.Ha, 43.20.Bi, 43.40.Rj, 43.20.Mv [BLM] Pages: 3146–3157

I. INTRODUCTION

As ear canals are narrow acoustic ducts, the sound transmission therein is often mathematically formulated using low-frequency solutions, which are based on the evanescence of higher order modes. The sound transmission to the eardrum is significantly influenced by the variation in the cross-sectional area. A low-frequency solution taking the area functions into account is the well-known Webster equation (Webster, 1919), which has been frequently applied to ear canals (e.g., Hudde, 1983; Stinson, 1985b; Stinson and Khanna, 1989). In its original form the equation is based on the assumption of plane sound waves propagating in horns with straight axes. The actual three-dimensional effects due to the varying area function are considered using a generalized formulation of the continuity equation, which accounts for the area function. Webster's equation assumes purely fundamental modes propagating forth and back in spite of the variation in cross-section. Within the scope of this approximation, the shape of the cross-sections has no impact on propagation.

Real ear canals are considerably curved. This complicates the determination of the area function for a given ear canal. Obviously it is necessary to replace the straight axes of axially symmetric horns with curved axes adapted to the lengthwise course of the ear canal. Stinson and Khanna (1989) could show that Webster's equation can be maintained for curved ear canals if the axis s is properly chosen. In this case the area function $A(s)$ can still be defined as a function of a single parameter, namely, the arc length along the s-axis. The axis has to follow the center of the ear canal. Therefore it is often referred to as the "middle axis" of the ear canal. Models based on such middle axes are sometimes called "unidimensional" instead of "one-dimensional."

An implicit deficiency of this approach is the assumption of planar waves, which is maintained in spite of the curvature of the middle axis and the varying shape of the cross-sectional area. To cover the complete spatial structure of sound fields, it is indispensable to use formulations which exceed the level of fundamental modes. The most general way of computing sound fields is using numerical methods based on spatial grids such as finite elements (FEs) or boundary elements. Actually all the simulations presented in this paper are obtained using FE.

A successful analytic approach aiming at mathematical coupling of the ear canal sound field to the vibrations of the eardrum has been developed by Rabbitt and Holmes (1988). To take advantage of the ear canal slenderness, a series representation of the wave in terms of fundamental and higher order modes is well suited. In the paper just mentioned, the geometry of the ear canal is restricted to the case of axisymmetric straight ear canals. In a later paper (Rabbitt and Friedrich, 1991), the modes are adapted to noncircular shapes. The usage of modes is particularly successful if the coupling of the central sound field to both ends of the ear canal (sound source and eardrum) is to be investigated. If exclusively the central region is considered, simpler approaches assuming fundamental modes are usually preferred. A decisive step toward a more accurate representation of sound fields in real ear canals using fundamental modes has been derived by Agullo *et al.* (1998). The authors introduced a theory taking into account curved equipotential surfaces. In this way they found a modification of Webster's equation, which allows sound field computations for new types of curvilinear horns.

Later on, the curvilinear horn theory has been extended to curved ducts with general area functions (Farmer-Fedor and Rabbitt, 2002). The usage of an "effective Webster area function" indicates a close relationship to Webster's solution. Due to the curved equipotential surfaces, this area function can considerably differ from the area function determined assuming plane areas. It is difficult to verify the validity of such theories by experiments because probe tubes disturb the

0001-4966/2009/125(5)/3146/12/$25.00 © 2009 Acoustical Society of America

sound field. Farmer-Fedor and Rabbitt (2002) used an elaborate setup to measure the sound pressure along straight paths in the ear canal and found data which delivered plausible results when evaluated according to their theory. This confirms the general applicability of the extended curvilinear horn theory. However, local deviations from the theory could not be detected by only using one straight path.

Scanning the spatial structure of sound fields by means of microphones can be very inaccurate. Therefore very thin probes are mandatory (e.g., 0.2 mm outer diameter, as used in Stinson and Daigle, 2007). The only way to get rid of sound field disturbances is by computing rather than measuring the pressure distribution. Models of rigidly walled ear canals are very simple acoustic systems, which can be easily treated using FE or other techniques. It would be more ambitious to model also the soft tissue lining the ear canal. However, this paper focuses on effects caused by the interaction with the external sound field, the curvature of the ear canal, and the eardrum vibrations. Effects of skin and cartilage in the ear canal are excluded. Wave effects occurring in hard-walled loss-free ducts are perfectly reproduced using FE. The only precondition is that the discretization of meshes is made fine enough.

The validity of an analytic horn equation approach has already been numerically checked by using boundary elements (Stinson and Daigle, 2005). It turned out that this approach is able to predict the pressure distribution along ear canal middle axes fairly well. However, some deviations in transverse directions are also reported. A certain degree of variation is to be expected when the transverse pressure distribution is examined on planar surfaces. The actual curvature of equipotential surfaces mostly produces moderate pressure variations in the order of 1 dB if plane cross-sectional slices are considered. In addition, huge variations of more than 25 dB were found near a pressure minimum. This reveals that pure fundamental mode approaches cannot correctly reproduce sound fields everywhere in the ear canal, even within the central “core region,” which will be specified below. The occurrence of such irregular parts in the sound field will be extensively discussed in this paper (Sec. IV).

For practical applications, the question of how far sound sources influence the form of the field in the ear canal has high significance. Sound sources feeding the ear canal can be very different: they can be positioned far or close to the pinna or—in the case of different types of hearing aids—they can act at the entrance or even within the ear canal. Another interesting case is a measuring device coupled to the ear canal. Near the source proximity effects take place, which can considerably disturb measurements when fundamental mode propagation is erroneously assumed at microphone positions too close to the source. Such effects have been studied in an arrangement simulating a hearing aid as sound source (Stinson and Daigle, 2007). In the paper at hand, the influence of the source on the sound field in the ear canal is only investigated for sound sources outside the ear canal. The aim of these investigations was to locate the position of the “entrance” of the ear canal.

At the end of the ear canal, the eardrum causes disturbances (Rabbitt and Holmes, 1988). Due to its inclined position and retracted conical shape, the eardrum would evoke higher order modes even if it was rigid. The main source of modes, however, is given by the vibrations. Therefore it is necessary to include a model which is able to simulate the vibrations of the eardrum realistically. A complete model of the middle ear is necessary because the ossicular chain has essential impact on the eardrum vibrations. The middle ear is formulated as a FE model as well.

In this paper no new analytical method is presented, which could further improve the computation of ear canal sound fields. If there is actually a reason to calculate the sound field in individual ear canals, there is no way other than using general numerical methods. The main problem is reproducing the geometry of the ear canal geometry correctly rather than performing sound field calculations. The goal of this paper is to elucidate the characteristics of ear canal sound fields and to identify the limits of unidimensional approaches using fundamental modes. A deeper understanding of the characteristics of ear canal sound fields can be obtained by considering the fields under the aspect of minimizing the sound energy (Sec. V). This leads to the notion of “fundamental sound fields,” which include not only approximate fundamental modes but also irregular structures, which are referred to as “one-sided isosurfaces.”

II. FINITE ELEMENT MODEL

An adequate representation of widespread external sound fields around the head cannot be obtained using FE because the number of nodes rapidly increases when the air-filled space around the head is enlarged. In this paper, however, only sound fields in the ear canal and fairly close to the pinna are studied. Therefore a comparably small volume encompassing the pinna, referred to as “pinna box” in the following, is sufficient for the investigations. The anatomical and physiological part of the FE model comprises the pinna including a small part of the surrounding head surface, the ear canal, and a middle ear model. These elements are shown in Fig. 1, which also provides the scales to allow for estimating the dimensions. As already mentioned, non-rigid tissue was not included in the present investigation. The ear canal walls and the pinna were assumed perfectly rigid and clamped. Thus an external source produces sound waves in the ear canal but no vibrations of the pinna or the ear canal walls. The maximum lateral dimension of the ear canal in the central part is about 7.5 mm. Therefore in the audio frequency range, no higher order modes can propagate.

As the vibrations of the eardrum considerably depend on the ossicular chain loading the tympanic membrane, a complete middle ear model needs to be used. The model is kept comparably simple. Nevertheless it is in agreement with many measured results. Actually this agreement is the basic idea to determine the parameters. The parameters are chosen to reproduce a pool of different transfer function data measured in many cadaver ears. The data pool comprises acoustic eardrum impedances and several transfer functions relating sound pressure and volume velocity at the eardrum to forces and velocities measured at the *processus lenticularis* and the stapes footplate under different boundary conditions

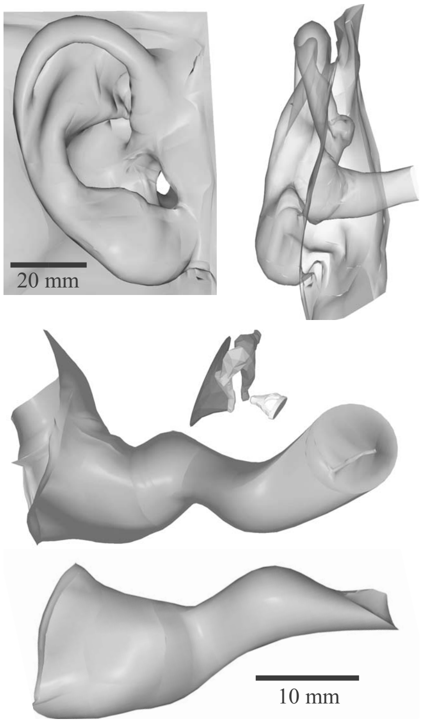


FIG. 1. Details of the finite-element model used. In the upper part, the pinna is shown from lateral and frontal views. In the lower part, the ear canal is depicted from two views in an enlarged scale. The ear canal walls are assumed as rigid. The middle ear in the center of the figure is depicted in the same scale as used for the ear canal. It consists of the eardrum, the three ossicles, and several ligaments and muscles which are not shown in the figure.

(Hudde and Engel, 1998a, 1998b, 1998c). The data pool is supplemented by further measurements using laser vibrometry (Weistenhöfer, 2002). In particular, spatial velocity components at different points on the three ossicles are measured to obtain data on translational and rotational motions. Such measurements are performed using acoustic and mechanic stimulation as well. The acoustic signals are applied at the eardrum by means of a measuring head, whereas mechanic excitation is realized by shaking the whole temporal bone in three orthogonal directions.

Compared to FE models which provide very refined representations of the eardrum (e.g., Tuck-Lee *et al.*, 2008; Funnell and Decraemer, 1996), the tympanic membrane of the model is fairly simple. It is modeled by two homogeneous shells representing pars flaccida and pars tensa (mass density of $1.2\ \mathrm{g/cm^3}$, thickness of 0.2 mm, Young's moduli of 1.0 and 2.1 MPa, respectively, Poisson's ratio of 0.4, and β-damping of $10\ \mu\mathrm{s}$). The retracted shape of the tympanic membrane corresponds to average data, whereas the geometry of the middle ear ossicles is obtained from an individual middle ear. The shape of the ossicles is measured using an optical method (Weistenhöfer and Hudde, 1999). The ossicles have the following masses: malleus, 23.5 mg; incus, 29.9 mg; stapes, 3.36 mg. The model further comprises the two joints, the anterior malleal and the posterior incudal ligament, the annular ligament, and the two muscles. In the work of Weistenhöfer (2002), all the elastic elements are described by matrices containing translational and rotational compliances and resistances. In the FE model the elastic elements are modeled as solids approximating the matrices. The most important parameters are the translational and rotational compliances in the direction of the main axis of operation. For the malleal and the incudal ligaments, the same values [translational compliance of 0.12 mm/N, rotational compliance of $5/(\mathrm{N\ mm})$] are approximated. Both muscles are predominantly specified by translational compliances of 0.7 mm/N. The incudomalleal joint is rather stiff, having compliances of $5\times10^{-3}\ \mathrm{mm/N}$ (translational) and $0.75\times10^{-3}/(\mathrm{N\ mm})$ (rotational). The incudostapedial joint is translationally fairly stiff (0.07 mm/N) as well but rotationally compliant $[5/(\mathrm{N\ mm})]$. The annular ligament is mostly characterized by its longitudinal compliance being 0.8 mm/N. The hydroacoustic load of the stapes has only weak influence on the sound field in the ear canal. It is approximated as a mechanical second order vibrator (mass of 10 mg, compliance of 11 N/mm, and mechanical resistance of $70\ \mathrm{N\ s/mm}$).

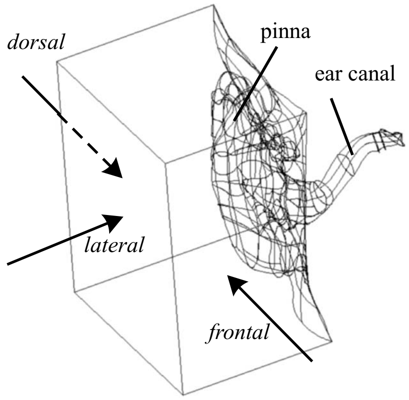


FIG. 2. The pinna box, a small air volume encompassing the pinna, is used to approximate free-field conditions and to implement external sound fields. It originates from a cuboid with the dimensions of 80 mm (height), 70 mm (width), and 50 mm (depth), which have been adapted to the shape of the head around the pinna. All the walls of the box are lined absorbingly. The three walls indicated by arrows can also be used as sound sources approximating laterally impinging sound incidence and grazing sound waves from frontal and dorsal directions.

The pinna box (Fig. 2) has almost rectangular shape. Basically, the box provides an absorbing volume, which approximates the free space. In addition, external sound sources can be implemented by using the walls of the pinna box as vibrating pistons. Distinctly different sources approximating sound incidence from different directions can be generated if only one surface is "active." The approximation of the free space is obtained by "coating" the walls with the specific field impedance, ρc, which is the product of the static mass density and the speed of sound. Although the assignment of the specific impedance generates perfect absorbers only for perpendicular sound incidence, the pinna box turned out to provide a fairly good approximation of the free space. This has been checked by considering the influence of box dimensions on the radiation impedance. Enlarging the volume of the utilized pinna box leaves the radiation impedance almost unchanged.

Meshing of the model is done by automatic algorithms provided by ANSYS. The node density is chosen according to the requirement that the distance between two nodes should be not greater than a tenth of the wavelength at the highest frequency considered. This leads to maximum distances of about 1.5 mm in the fluid of the ear canal and 0.7 mm on the eardrum and also on the ossicles. In addition, the geometry has to be captured with sufficient accuracy. This yields a higher node density in strongly structured regions. Correct results are ensured by checking the results after refining the meshes: if the results are identical within reasonable tolerances, the spatial resolution is chosen appropriately. In Secs. III and IV sound fields calculated by the FE model for different excitations and in different regions are presented and interpreted.

III. SOUND FIELDS NEAR THE EARDRUM

First, the sound field near the eardrum, in the drum coupling region, is to be considered. All the FE calculations have been performed as harmonic analysis. Thus the results represent the steady state in terms of complex phasors of the sound pressure. The structure of resulting sound fields is visualized using surfaces of equal pressure, which are referred to as isosurfaces for brevity. In general, surfaces of equal pressure magnitude and phase are not identical. If not indicated otherwise, "isosurface" means a surface of equal pressure magnitude. If surfaces of equal pressure phases are addressed, they are denoted as phase isosurfaces.

Figure 3 shows isosurfaces for different frequencies.[1] As expected the structure of the sound field near the eardrum turned out to be independent of the source used. This was checked for the three sources approximating frontal, lateral, and dorsal sound incidence. Within the resolution of the graphic representations, no differences can be observed when switching between the sources. For clarity, it should be noted that—of course—the sound source determines the sound field everywhere. If the term "source independence" is used, it exclusively refers to the structure of the sound field, which is established by the invariant shape of the isosurfaces produced by different sources.

At low and high frequencies (results for 640 Hz, 8 kHz, and 12 kHz), the waves at the eardrum look very simple: the eardrum guides the waves in a similar manner as the ear canal walls do. The isosurfaces are almost perfectly perpendicular not only to the ear canal walls but to the eardrum surface as well. This means that for the given frequencies the eardrum vibrations are so small that they cannot significantly alter the sound field in the ear canal. In other words, the eardrum behaves fairly stiff with respect to the adjacent air in the ear canal. The missing effect of eardrum vibrations at 12 kHz is immediately proven by comparison with the sound field arising for an eardrum, which is made perfectly rigid (Fig. 4). The incident wave is reflected in the innermost part of the tympanomeatal corner resulting in typical standing wave patterns. At higher frequencies, the maximum pressure in the drum coupling region arises near point T in the tympanomeatal corner. Here the wave appears to be guided by the eardrum just like by the ear canal walls. The main reflection of the sound wave takes place in the "termination point" T. The shape of the isosurfaces is almost constant for frequencies above 8 kHz.

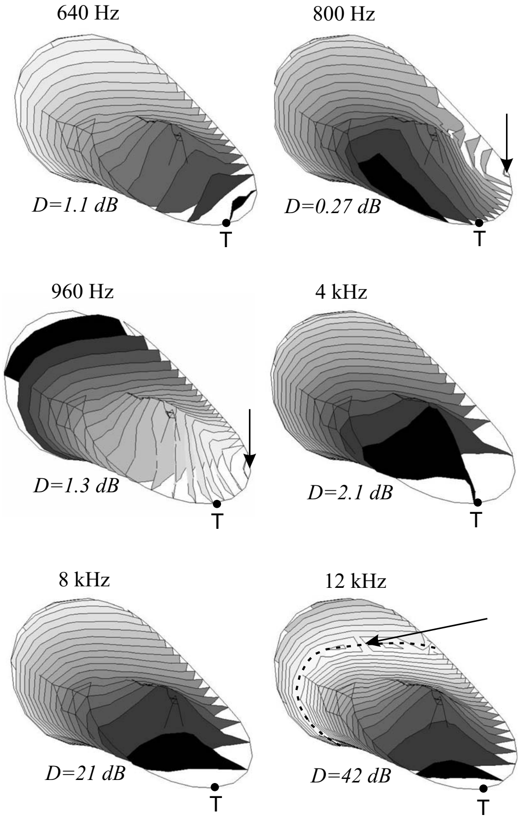


FIG. 3. Surfaces of equal pressure magnitude (isosurfaces) in the drum coupling region at different frequencies. The sound field is depicted viewing from inside through the (transparent) tympanic membrane and ear canal walls. Gray tones indicate magnitudes of complex pressure. The darkest surface belongs to maximum pressure. The lighter the gray tone, the lower the pressure magnitude. The gradation is adapted to the range of pressure magnitudes, separately for each frequency. Thus the gray tones are not comparable between different frequencies. To indicate the scale, the dynamic ranges D of pressure magnitudes occurring in the drum coupling region are explicitly given. The point T in the tympanomeatal corner denotes the innermost location of the ear canal. The arrows in the panels for 800 and 960 Hz indicate local pressure minima arising in the tympanomeatal corner. The arrow at 12 kHz shows the quarter wavelength minimum, which appears on an isosurface.

It should be mentioned that the strong reflections at high frequencies calculated using the FE model correspond to the parameter choice based on the authors' own measurements (Hudde and Engel, 1998c). These measurements agree with others regarding the reflectance minimum due to the main middle ear resonances, but at higher frequencies they give higher reflectance magnitudes compared to several other authors. A nice collection of data is given in Farmer-Fedor and Rabbitt (2002). Our data come very close to the data of these

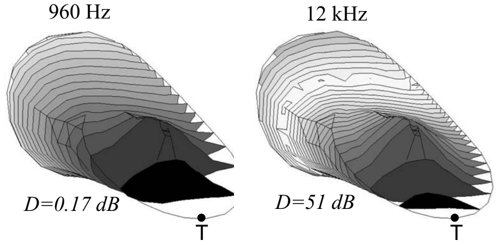


FIG. 4. Isosurfaces near the eardrum at 960 Hz and 12 kHz for the case of a rigid eardrum. Scale is given as dynamic range of pressure magnitudes.

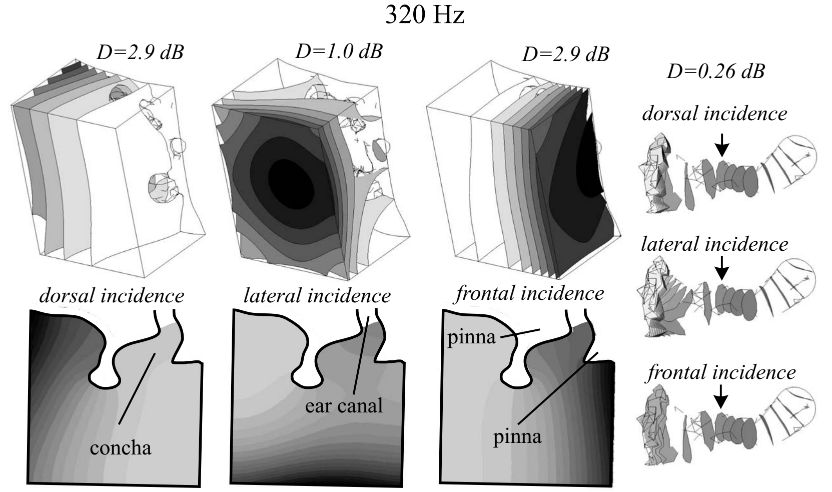


FIG. 5. Sound fields inside and outside the ear canal for dorsal, lateral, and frontal sound incidence at 320 Hz. Higher pressure magnitudes are indicated by darker gray tones, and lower pressures by lighter ones. Scale is given as dynamic range of pressure magnitudes. Upper left row: isosurfaces in the pinna box. Lower left row: contours of pressure isosurfaces in a horizontal plane section through the entrance of the ear canal. Right column: pressure isosurfaces in the ear canal. The foremost surfaces that can be considered identical in shape for all the three excitations are marked by arrows.

authors, except that the reflectance remains high beyond 10 kHz as well. However, also the reflectances measured by Farmer-Fedor and Rabbitt (2002) are higher than all the data published by other authors.

The most striking difference between the sound fields near normal and rigid eardrums is observed at 960 Hz (compare Figs. 3 and 4). In the case of a natural eardrum, the locations of maximum and minimum are almost interchanged. A closer inspection reveals that the tympanic membrane has noticeable effect in the frequency range from about 600 Hz to 4 kHz, where the middle ear as a whole and the tympanic membrane resonate. Passing through this range, the maximum near T is continuously shifted away from the eardrum, takes its remotest position at 960 Hz, and is shifted back to its normal position. Figure 3 can only show a small selection of these sound fields. The eardrum radiates a wave superimposing on the undisturbed sound field, which would occur for a rigid eardrum. The resulting sound fields cannot be described as unidimensional functions along a constant middle axis. According to Fig. 3, for 640 Hz, 960 Hz, and possibly for 4 kHz, an axis of propagation could rather be determined, but it would change with frequency. At 800 Hz a maximum and a minimum arise in opposite positions close to point T, a condition which obviously does not cope with a middle axis.

However, one has to keep in mind that a quarter wavelength at 1 kHz is as large as about 85 mm. Although the isosurfaces always suggest sound waves, the drum coupling region behaves as a lumped element rather than as a transmission line at this frequency. Even at 4 kHz the effect of wave propagation near the eardrum is comparably small yet. At higher frequencies, when significant wave propagation occurs, the sound field disturbance due to tympanic membrane vibrations widely drops. The maximum deviation $p_{\max}/p_T$ between the pressure magnitude at T and the absolute maximum occurring somewhere in the drum coupling region is only about 1.05 dB. This value appears at 960 Hz where the maximum arises at the remotest position within the drum coupling region. Above this frequency, the deviations continuously decrease because the pressure maximum actually arises very close to T.

Comparing directly the sound fields for a natural and a stiffened eardrum, level differences up to 10 dB occur, but only in a narrow frequency range near 3 kHz. Apart from this frequency range, the deviations are mostly below 1 dB. In summary, the eardrum vibrations only have noticeable impact on the sound field in the frequency range up to 3–4 kHz, where the drum coupling region approximately behaves as a lumped element. Hence the problems of specifying a middle axis in this frequency range are not very significant. The applicability of an ear canal middle axis will be further discussed in the light of general sound field structures presented in the Sec. IV.

IV. SOUND FIELDS WITHIN AND OUTSIDE THE EAR CANAL

The sound fields in front of the pinna and within the ear canal considerably differ in structure. The external field is a superposition of the source field and the response caused by scattering and reflection at the pinna. Therefore the resulting field has distinct three-dimensional characteristics. In contrast, the sound field in the core region of the ear canal is fairly regular because it is widely determined by fundamental modes propagating along the middle axis. It is reasonable to study both types of fields simultaneously to get an insight into the interaction which takes place in the transition region in between. Calculations have been performed at all multiples of 160 Hz up to 16 kHz. In the following, only a small selection can be represented graphically.

To get an impression of the sound fields in front of and behind the ear canal entrance, multiple views are necessary. Figure 5 depicts the sound fields for frontal, lateral, and dorsal sound incidence as produced by the pinna box (Fig. 2) at a low frequency (320 Hz). In the upper row on the left, some isosurfaces in the pinna box are visualized. The row below shows the traces of isosurfaces as boundary contours between different gray tones ("contour plots"). The pressure magnitudes are calculated in horizontal sections running through the pinna box and the center of the ear canal entrance. The upper boundary of this slice is given by the traces of the pinna and the ear canal walls in that plane. Here the transition of the external sound field into the ear canal can be

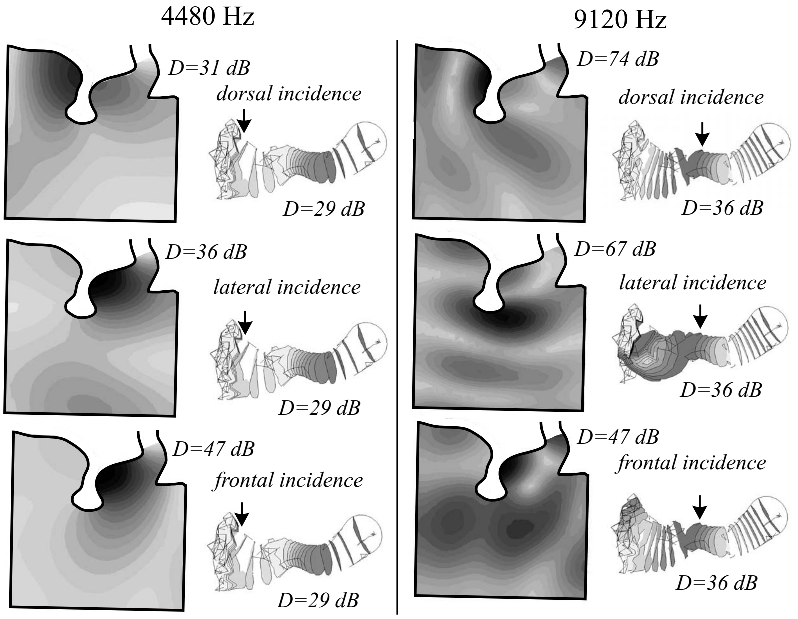


FIG. 6. Sound field inside and outside the ear canal for dorsal, lateral, and frontal sound incidence at 4480 Hz (left) and 9120 Hz (right). Scale is given as dynamic range of pressure magnitudes. For both frequencies contour plots in horizontal slices running through the center of the ear canal entrance as well as isosurfaces in the ear canal are provided. The foremost surfaces that can be considered identical in shape for all the three excitations are marked by arrows.

pursued. Its continuation is given in the third view, the column on the right of which shows the isosurfaces inside the ear canal.

The upper row shows that vibrating pinna box walls can actually generate approximately grazing wave fronts for dorsal and frontal incidence and impinging wave fronts for lateral incidence. The same is observed in the lower row. The ear canal isosurfaces to the right visualize variations due to the different sound sources. As expected the sound field close to the eardrum turns out to be independent of the source. The isosurfaces for sound coming from the side are noticeably different at the position where the entrance could be assumed, whereas the ear canal sound fields for dorsal and frontal incidence look identical up to the beginning of the concha. The arrows point to the first surface, which seems identical for all the three sources. This position is surprisingly far inside the ear canal. However, one should have in mind that at 320 Hz the absolute pressure differences are extremely small.

At 4480 Hz the sound field structure in the complete ear canal and even in the concha becomes almost independent of sources (Fig. 6). The similarity of the sound fields in the concha and the ear canal can be seen in the contour plots and the isosurfaces as well. Note that similar fields in the concha occur although the pressure distribution on the boundaries of the pinna box differs considerably. At 9120 Hz the conditions are completely changed. The sound fields for frontal and dorsal incidence remain rather similar, but for lateral incidence the sound field takes a very different structure. For lateral incidence the contours at the entrance exhibit more irregular structures which are not in line with the central axis of the ear canal. In the ear canal the field structure is considerably altered as well. The arrows indicating the foremost isosurfaces which are not affected by different sources are shifted toward the eardrum. Obviously, the extension of the core region strongly depends on frequency.

The striking differences between the sound fields at 4480 and 9120 Hz are caused by the conditions at the entrance. At 4480 Hz a pressure minimum arises, which corresponds to a velocity maximum (long vectors in the top panel in Fig. 7). The large velocity in the transition region constitutes a strong field interaction between the concha and the ear canal. The sound field in the ear canal continues outward. As a consequence, even the field in the posterior part of the concha is almost independent of the direction of sound incidence. Converse conditions are found at 9120 Hz where a velocity minimum occurs at the entrance. The minimum forms a surface on which the acoustic input impedance is high, approximating a rigid surface. Here the pressure is not subject to any boundary condition. Hence the sound fields in front of and behind this surface are only weakly coupled and can look very different. Thus, a bit surprisingly, the weak coupling allows the sound sources to influence the fields up to more posterior positions in the ear canal. Vice versa, the strong coupling established by large velocities in the transition region leads to a regular continuation of the ear canal sound field outward to the concha and therefore to the weak influence of sound sources.

The weak coupling at 9120 Hz is also expressed by the time-variant orientation of velocity vectors. The velocities in the concha and in front of the minimum continuously change their orientation during a cycle of oscillation, whereas behind the minimum the direction of the velocity vectors is almost constant. Arbitrary points in time have been chosen to illustrate these variations in the four lower panels of Fig. 7. The simultaneous constancy inside the ear canal and variability in the concha illustrates the decoupling of fields very intuitively. In contrast, the invariability of the velocities at 4480 Hz is expressed by the fact that a single panel can represent all the sound sources and times simultaneously.

With increasing frequency, the distances between pressure maxima and minima become shorter. Therefore both a maximum and a minimum arise near the entrance. Discrimination of strong and weak coupling at the ear canal entrance becomes increasingly ambiguous. As a result such strong coupling as observed at 4480 Hz is not found again at fre-

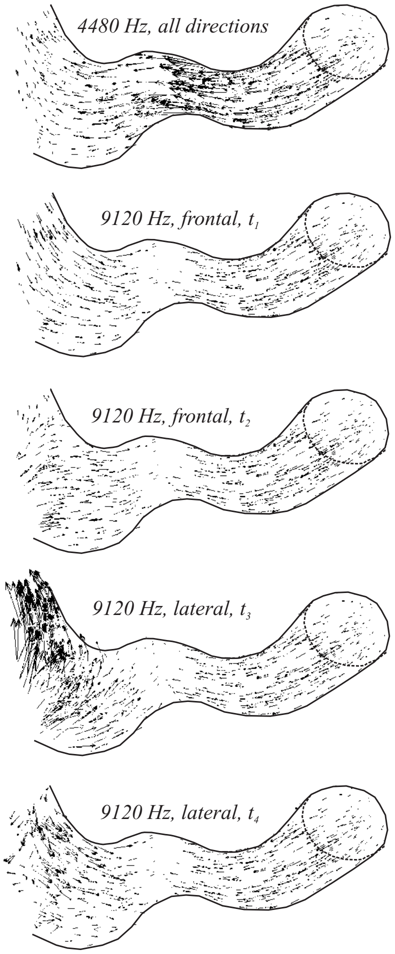


FIG. 7. Velocities in the transition region between concha and ear canal at 4480 and 9120 Hz. The underlying geometry is identical to that in Fig. 6. The concha is on the left, and the tympanic membrane on the right. Top: The orientation of the velocities at 4480 Hz is almost identical for the three directions of incidence and constant with time. Other four panels below: At 9120 Hz the orientation of the velocity vectors in the concha changes when the direction of sound incidence is altered. In addition, it varies during a cycle of the oscillation (time instants t_1-t_4).

quencies beyond 10 kHz. The complexity of sound fields generally grows with increasing frequency. An example is given at 14 240 Hz (Fig. 8). Here the beginning of the core region is similar to that for 9120 Hz. Generally this position does not change too much at higher frequencies.

In the sound field at 14 240 Hz (Fig. 8, right), a very peculiar type of isosurfaces is observed, which is oriented to one side of the ear canal wall instead of to the middle axis. These surfaces form a set of domes arching over a certain point on one side of the ear canal wall. Such structures are denoted as one-sided isosurfaces in the following. They occur at lower and higher frequencies as well. One-sided isosurfaces do not conform to sound waves propagating along a middle axis. They go beyond the theory of fundamental modes. All the approaches addressed in Sec. I are unable to predict such isosurfaces because isosurfaces being normal to the middle axis are presupposed.

V. FUNDAMENTAL SOUND FIELDS

In this section the sound fields in the core region are considered in more depth. These fields are per definition independent of the external sound source and widely characterized by "regular isosurfaces," which completely cover the duct aperture. The surfaces allow a middle axis to be constructed running through the surface centroids. Figure 9 outlines such regular isosurfaces and a corresponding middle axis for the special ear canal used throughout this paper. However, as we have seen, the sound fields are not exclusively composed of regular isosurfaces but comprise dome-shaped one-sided surfaces. In the following a generalized concept of fundamental sound fields is given, which includes these isosurfaces, although they cannot be represented by pure fundamental modes.

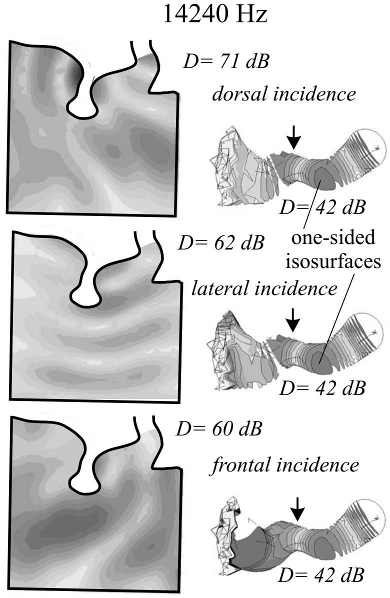


FIG. 8. Sound fields inside and outside the ear canal for dorsal, lateral, and frontal sound incidence at 14240 Hz.

The concept of fundamental sound fields is based on low-frequency approximations. Hence no accurate definition can be given. However, several properties and features can be specified, which constitute fundamental sound fields. Fundamental sound fields can only occur if the duct has the following properties: (a) the walls are almost rigid, (b) the duct has a recognizable direction of wave propagation which can be expressed by a middle axis, (c) the lateral dimensions are small compared to wavelength in the frequency range

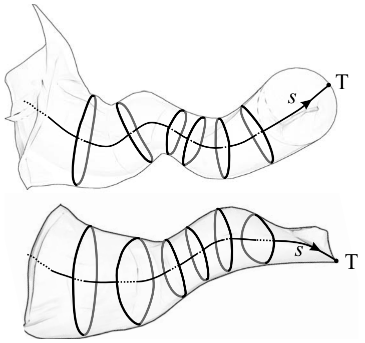


FIG. 9. Schematic representation of regular isosurfaces and a corresponding middle axis. The middle axis ends in the termination point T in the tympanomeatal corner. The ear canal is shown from top and lateral views.

considered, (d) the size and the shape of cross-sectional areas may vary along the middle axis, but only gradually, and (e) the duct may be curved to a moderate degree. Ear canals meet these requirements in the range of audible frequencies.

A fundamental sound field can only arise in regions which are not affected by near-field effects of sound sources, discontinuities, or other reflecting objects in or at the ends of the duct. At a sufficient distance from such disturbances, the structure of the sound field, i.e., the shape of the isosurfaces, is characterized by certain features which are discussed below. The sound field close to the drum is not fundamental but is nevertheless independent of the sound source outside the ear canal because the influence of the source ends at the beginning of the core region. For this reason, the ending of the core region is of lower practical interest.

One-sided isosurfaces can appear at every location in the core region. Seemingly regular isosurfaces on both sides of one-sided isosurfaces are also warped somehow. Thus the shape of these isosurfaces changes with frequency. However, regular isosurfaces sufficiently far away from one-sided isosurfaces turn out to be fairly constant. Therefore a middle axis is certainly meaningful in regions of regular isosurfaces. The general usefulness of a middle axis and the best way to construct it will be further discussed in Sec. VI after having studied the conditions in regions of one-sided surfaces.

A reasonable generalization of fundamental modes to fundamental sound fields is obtained by regarding the field structure from the viewpoint of energy. According to a general principle, many physical systems tend to approach a state of minimum energy. In a duct which is so narrow that higher order modes are evanescent, the transition to a stationary state of minimal energy is enforced by the slenderness of the duct. This will be elucidated in the following.

Let us first consider a simple cylindrical duct. If a certain sound power is to be transmitted via the duct, this can be done using any wave mode which can propagate in the duct. It can be shown that the mean energy density in the duct becomes minimum if the fundamental mode, which means plane waves in the cylindrical duct, is chosen. Thus in the case of a cylindrical duct, the fundamental mode is definitely the mode of minimum energy. It seems likely that this principle of minimum energy will be valid also for noncylindrical ducts. In the following, the sound fields actually arising in ear canal are investigated under this aspect. This analysis confirms that the principle of minimum energy applies for general curved ducts as well. Not only the regular isosurfaces, which are close to plane waves, but also the one-sided isosurfaces in combination with the corresponding phase isosurfaces turn out to take a structure minimizing the energy density in the duct. Therefore one-sided isosurfaces are considered as part of the fundamental field in spite of the striking differences to plane waves.

The energy density at a point given by its position vector **r** can be written as

$$W'(\mathbf{r}) = \frac{p_{\mathrm{RMS}}^2(\mathbf{r})}{\rho c^2} + \rho v_{\mathrm{RMS}}^2(\mathbf{r}). \qquad (1)$$

Herein root-mean-square (RMS) values are used instead of amplitudes because no simple relation between the amplitude and the RMS value exists for velocities having time-variant orientations in space. According to Eq. (1), minimum energy density is achieved if the RMS values of sound pressure p_{RMS} and velocity v_{RMS} are as small as possible under the given conditions.

From the general expression of harmonic pressure $p(\mathbf{r}) = |p(\mathbf{r})|e^{j\varphi_p(\mathbf{r})}$ at an angular frequency ω, the velocity is obtained using Euler's equation

$$v(\mathbf{r}) = -\frac{\mathrm{grad}\{p(\mathbf{r})\}}{j\omega\rho} = \frac{e^{j\varphi_p(\mathbf{r})}}{\omega\rho}(j \cdot \mathrm{grad}\{|p(\mathbf{r})|\} - |p(\mathbf{r})| \cdot \mathrm{grad}\{\varphi_p(\mathbf{r})\}). \qquad (2)$$

The complex acoustic power penetrating a surface A is obtained by integrating the product of pressure and conjugate complex velocity,

$$S(\omega) = \frac{1}{2}\int_A p(\mathbf{r}) \cdot \mathbf{v}^*(\mathbf{r}) d\mathbf{A} = -\int_A \frac{|p(\mathbf{r})|^2 \mathrm{grad}\{\varphi_p(\mathbf{r})\}}{2\omega\rho} d\mathbf{A} - j\int_A \frac{|p(\mathbf{r})|\mathrm{grad}\{|p(\mathbf{r})|\}}{2\omega\rho} d\mathbf{A}. \qquad (3)$$

If the power transmitted through a duct is to be computed, the integral has to be taken over a surface A, which completely covers the duct aperture. In this case Eq. (3) provides the separation of the apparent power into its active and reactive parts. The expression for the active power, the real part in Eq. (3), is a somewhat generalized form of an equation given earlier (Stinson, 1985a). The active power is constant along the middle axis. This establishes the close relationship between the magnitude and the phase gradient of the pressure.

If the integral in Eq. (3) is taken over a regular isosurface, one can conclude that velocity components tangential to the isosurface do not contribute to the apparent power. On the other hand, such components lengthen the velocity vector and therefore increase its RMS value and thus the energy density due to Eq. (1). Hence velocity vectors in minimum energy sound fields must be normal to the isosurfaces, at least approximately. Near the (rigid) walls this condition is met in any case. The requirement of vanishing velocity components normal to the walls enforces vanishing pressure gradients in the direction perpendicular to the walls.

From Eq. (2) the temporal dependence of the velocity can be written as

$$v(\mathbf{r},t) = \mathrm{Re}\{v(\mathbf{r})e^{j\omega t}\} = -\frac{1}{\omega\rho}\mathrm{grad}\{|p(\mathbf{r})|\}\sin(\omega t + \varphi_p(\mathbf{r})) - \frac{1}{\omega\rho}|p(\mathbf{r})| \cdot \mathrm{grad}\{\varphi_p(\mathbf{r})\}\cos(\omega t + \varphi_p(\mathbf{r})). \qquad (4)$$

This means that in general the direction of velocities is time-variant. The instantaneous direction of the velocity oscillates between the directions of the gradients of magnitude and phase. This contradicts the postulation of velocity vectors

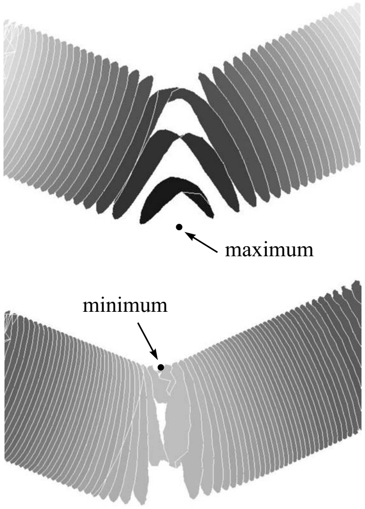


FIG. 10. One-sided magnitude isosurfaces for a pressure maximum and minimum occurring in a bend (tube diameter of 7.5 mm and frequency of 5.8 kHz). Increasing darkness of gray tones indicates increasing pressure magnitude. The pressure differences between the isosurfaces are kept constant. Thus the higher density of surfaces near the pressure minimum indicates larger gradients compared to the maximum. (Disregard the wrong lines arising left of the bend in both panels: these lines belong to the duct geometry and could not be suppressed in the graphic.)

being normal to isosurfaces except for coinciding isosurfaces of magnitude and phase. Only in that case the direction of the velocities is time-invariant and normal to the isosurface. Hence these conditions characterize regular isosurfaces: the pressures on regular isosurfaces oscillate in phase, and the velocities at all points on the isosurfaces have constant orientations being normal to the surface. The velocities on regular isosurfaces define local directions of propagation. In particular, the velocity in the surface centroid provides the local direction of the middle axis.

Next, one-sided magnitude isosurfaces have to be considered. The origin of such surfaces is easily found. Consider a duct with a single bend. Due to reflections at the wall, the sound pressure of a fundamental wave is a little higher on the concave side of the bend than on the opposite convex side. Thus, if a local pressure maximum occurs in the bend, it will be situated on the concave side. The isosurface which belongs to this pressure magnitude degenerates to an isolated point. The isosurfaces of slightly smaller pressures cannot reach the other side of the ear canal as well. As a result the isosurfaces must take the shape of domes arching over the location of the maximum. Hence one-sided isosurfaces systematically arise in general curved ear canals. They can occur at arbitrary positions but are particularly pronounced in bends. The case of maxima arising on a closed circumference, instead of at a point on one side of the wall, is an exception that can only occur in ideal straight cylindrical ducts and axisymmetric horns.

To study one-sided isosurfaces in pure form, the sound transmission through a single bend in an otherwise straight duct of constant cross-section has been considered. The upper panel in Fig. 10 shows the isosurfaces for the case of a pressure maximum in the bend. Using an analogous argument, it becomes clear that a local minimum in a bend must evoke one-sided magnitude isosurfaces as well. As expected the minimum appears on the convex side (lower panel of Fig. 10). Shifting of the extrema to the center of the bend has been achieved by adjusting the termination of the duct accordingly. The terminating impedance was chosen to be double and half the wave impedance to obtain symmetric cases for the extrema. Nevertheless, the conditions at pressure maxima and minima are not uniform. Pressure maxima are always broader than minima, particularly if the incident waves are strongly reflected at the end of the duct like in ear canals.

It is interesting to examine how the sound field can maintain the property of minimum energy in the region of one-sided magnitude isosurfaces. At the locations of pressure minima and maxima, the gradient of the pressure magnitude vanishes. Therefore, due to Eq. (2) the velocity in the extrema is only determined by the phase gradient. In contrast to the magnitude, the phase monotonically decreases from the source to the termination. No local phase extrema exist, which could disturb regular phase isosurfaces. Therefore, in order to maintain the orientation of velocity vectors pointing mostly in the direction of propagation, the phase isosurfaces should be regular also in regions of one-sided magnitude isosurfaces. This behavior is fully confirmed by the FE calculations. Obviously phase isosurfaces indicate the direction of propagation much better than magnitude isosurfaces since they are not warped near extrema.

The sound field near the extrema is highly symmetric. This can be analytically seen by reconsidering the active power expressed in Eq. (3). The meaning completely changes if the integral is taken over a one-sided magnitude isosurface. The real part of the integral taken over a regular isosurface means the active power transmitted through the duct, whereas the integral must be zero if it is taken over one-sided isosurfaces. It follows that

$$\int_A \mathrm{grad}\{\varphi_p(\mathbf{r})\}d\mathbf{A} = 0. \tag{5}$$

This results in symmetrically domed magnitude isosurfaces and phase gradients, which are equal at opposite points on the surfaces.

The conditions in the vicinity of a pressure maximum are schematically shown in Fig. 11. At sufficient distance from the maximum, the isosurfaces of magnitude and phase coincide. Near the pressure maximum, the gradients differ in direction. Therefore, the velocity changes its direction during a cycle of the harmonic sound wave. Since the gradient of the magnitude is small near a pressure maximum, the direction of the velocity is mostly given by the phase gradient which points in the direction of propagation. Just during a short time interval around zero-crossing of $\cos(\omega t+\varphi_p(\mathbf{r}))$, the small magnitude gradient near the maximum determines the direction of velocity according to Eq. (4). As a result, the velocities in pressure maxima mostly point in the direction of propagation. Only in the short time interval around zero-crossing does the direction rapidly turn to the opposite direction. The temporal changes in velocity fields in pressure extrema are depicted in Fig. 12, where the instant of direction

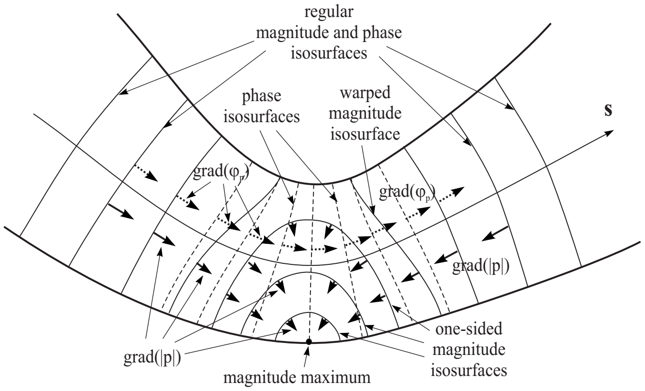


FIG. 11. Schematic representation of the isosurfaces of pressure magnitude and phase and the corresponding gradients in the vicinity of a pressure maximum. All the surfaces are perpendicular to the walls.

reversal is denoted as t_0. This event is repeated after each half cycle under reversed conditions. At t_0 the velocity in the pressure maximum is directed normal to the walls, except immediately at the walls where the normal velocity has to vanish.

In a pressure minimum, the conditions are a little more intricate. Exactly in the minimum, again the phase gradient determines the direction of velocity because the gradient of the pressure magnitude vanishes. However, on both sides of the minimum, the gradient of the pressure magnitude rapidly increases, particularly in narrow minima. Thus, contrary to the conditions in the broad maximum, the gradient of the pressure magnitude significantly contributes to the total velocity vector. The direction of this component is determined by the one-sided isosurfaces of the pressure magnitude. Close to the minimum location, the gradient of the pressure magnitude points toward the minimum. Thus it is oriented almost perpendicularly to the direction of wave propagation, a condition which violates the "rules" governing energy minimal sound fields. To meet the rules at least as well as possible, such unfavorable conditions must be restricted to narrow sections around the minima. This can actually be observed in Fig. 10. The one-sided isosurfaces form very steep domes; i.e., the surfaces approach regular surfaces already very close to the minimum. Due to their limited spatial extension, one-sided isosurfaces near minima can be easily overlooked when the isosurfaces are regarded. On the other hand, the pressure differences between the one-sided isosurfaces are much greater than in maxima. Therefore the large level differences in cross-sections through the pressure minimum are easily detected. As already mentioned in Sec. I, such lateral level differences have already been reported by others (Stinson and Daigle, 2005).

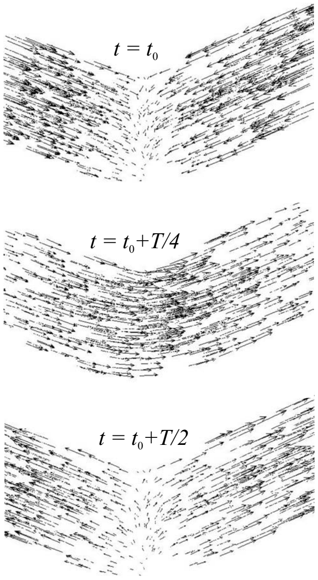


FIG. 12. Velocity vectors in a region of one-sided isosurfaces during a cycle of period T. At $t = t_0 + kT/2$, k being integer, the velocity field changes its direction in the duct. Apart from these instants, the velocity vectors almost permanently point in or contrary to the direction of propagation.

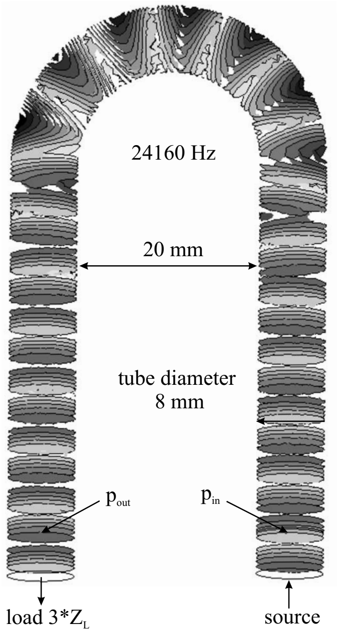


FIG. 13. Sound field in a toroidal duct at a frequency slightly below the cutoff frequency of the first higher order mode. The transmission via the curved part of the duct is hardly affected by the one-sided isosurfaces if it is considered from some distance using straight tube adaptors of sufficient length.

The effect of one-sided isosurfaces on sound transmission is low. This is seen when the transmission via a curved section in an otherwise straight duct is examined using FEs. The tube shown in Fig. 13 consists of a toroidal part and two straight adaptors at both ends. The diameter of the curved section and the adaptors is 8 mm. At the right end a nonreflecting source generates a sound field, whereas at the left end a load impedance is applied. The adaptors are chosen long enough to ensure almost perfect plane waves at the ends of the adaptors in the frequency range of interest. The pressure values p_{in} generated near the source (5 mm from the entrance) and p_{out} generated near the acoustic load (5 mm from the outlet) are taken from the numerical solution to compute the transfer function $p_{\text{out}}/p_{\text{in}}$. This transfer function is compared to the transfer function of a homogeneous straight tube of the same length.

The differences between the transfer functions are surprisingly small in spite of strong one-sided isosurfaces appearing in the bended part of the duct and a little beyond. At high frequencies, the sound field structure approaches that of the first higher order mode of the toroid. Figure 13 represents the sound field at 24 160 Hz, which is just below the cutoff frequency of the mode. Up to this frequency, the deviations of the transfer functions are not greater than 1 dB in magnitude and 12° in phase. Thus, with respect to sound transmission, the one-sided isosurfaces are rather marginal disturbances if they are evanescent.

VI. SUMMARY AND CONCLUSIONS

The sound field in ear canals and its continuation to the concha and a volume enclosing the pinna have been investigated for sound incidence from different directions. As also found by others, the sound field can be subdivided into three parts, the drum coupling region near the tympanic membrane, the external part in front of an entrance, and the core region in between.

At the end of the ear canal, the vibrations of the eardrum disturb the simplicity of the fundamental sound field. At frequencies sufficiently above the resonances of the middle ear, i.e., beyond about 4 kHz, the eardrum appears to guide the sound waves like the ear canal walls because it behaves fairly stiff compared to the adjacent air. This is best seen comparing the sound fields obtained for real and rigid tympanic membranes. In the frequency range between 600 Hz and 4 kHz, the eardrum noticeably radiates sound, which disturbs the simple field structure to some extent. However, the corresponding pressure variation in the drum coupling region remains fairly small because the wavelengths are large compared to the length of the section. These findings support circuit models replacing the true middle ear load by a lumped-element acoustic load impedance (Hudde and Engel, 1998a).

The structure of external sound fields outside the ear canal is mainly formed by the sound source and the pinna. Therefore, it has distinctly three-dimensional character. In contrast, the sound field structure in the core region, expressed by isosurfaces, is independent of the sound source and behaves essentially unidimensionally. The first surface that is independent of sources, within a certain tolerance, specifies the entrance, the beginning of the core region. Its location is mainly influenced by the location of pressure extrema near the entrance. In a pressure minimum the corresponding large velocity strongly couples the sound fields in the ear canal and the concha. In this case, the influence of the sound source becomes comparably weak, which means that the core region extends outward up to the concha. Vice versa, a pressure maximum at the entrance decouples the outer and inner sound fields, which shifts the entrance to more posterior locations into the ear canal. Using broadband excitation, the beginning of the core region is determined by the most unfavorable frequency. This yields a point immediately behind the first bend of a typical ear canal. The estimation of the entrance position according to isosurfaces is very strict. Using weaker criteria, e.g., the source independence of pressure transfer functions between a point at the entrance and a point in the rear section of the ear canal, an entrance location in front of the first bend is found.

The sound field in most sections of the core region follows a middle axis and is regular; i.e., the isosurfaces are not very different from cross-sectional areas, albeit slightly curved. However, also domelike-shaped one-sided isosurfaces, which stick to one side of an ear canal wall, systematically arise in real, curved, and non-axisymmetric ear canals. One-sided isosurfaces originate from pressure minima and maxima and can therefore arise at any position, but they are most pronounced in strongly curved sections.

In spite of such irregular isosurfaces, the sound field in the core region is referred to as fundamental sound field because, like in the case of ideal plane waves, the sound field transmits acoustic power producing minimum energy density in the duct. Plane waves and regular waves having identical isosurfaces of pressure magnitude and phase transmit acoustic power with minimum energy because the velocity vectors are always perpendicular to the pressure isosurfaces. An examination of one-sided isosurfaces and the associated phase surfaces reveals that the principle of minimum energy further applies in a weakened form. The disturbance of the regular sound field is restricted to a section of minimal extension. Moreover time intervals of velocities being not perpendicular to pressure isosurfaces are kept as short a possible. The physical reason for the energy minimization is the slenderness of the ear canal, which enforces conversion of locally appearing higher order modes to fundamental modes within short distances. Thus one-sided isosurfaces describe the sound field structure of short evanescent regions, which have been found by others as well (Farmer-Fedor and Rabbitt, 2002).

A middle axis is usually determined as the line interconnecting the centroids of all isosurfaces. However, in general the isosurfaces of pressure magnitude and phase disagree and vary with frequency. Thus, in principle, the middle axis of ear canals also changes with frequency if it can be specified at all. Fortunately, the real conditions are not that unfavorable. Apart from regions of one-sided isosurfaces, the isosurfaces of magnitude and phase almost coincide. Moreover, the shape of isosurfaces turns out to be only weakly dependent on frequency except near one-sided isosurfaces. As the one-sided isosurfaces are bound to pressure extrema, they are shifted through the ear canal if the frequency is altered. Hence isosurfaces at all locations change with frequency when a one-sided isosurface passes through. Regular surfaces in the vicinity of one-sided isosurfaces are also warped if one-sided isosurfaces approach or move away due to variation in frequency. However, the phase isosurfaces remain regular even in regions of one-sided magnitude isosurfaces. This is a consequence of the minimum energy principle. The phase isosurfaces are approximately independent of frequency in the full audio range.

The latter finding implies that phase isosurfaces are almost identical for incident and reflected waves. Otherwise the superposition of both waves would alter the resulting isosurface as a function of frequency. Phase isosurfaces are closely related to wave fronts, which represent points of equal time delay after impulsive excitation of a sound wave. Obviously, wave fronts are completely independent of the ear canal behind the surface just reached by the wave. The same argument holds for the wave in opposite direction. Wave fronts for both directions can significantly disagree, for instance, at a discontinuity in the area function. However, in continuously varying ducts such as ear canals, wave fronts and, therefore, phase isosurfaces are almost independent of the direction of propagation.

The sound velocities in regions of regular isosurfaces always point in the direction of propagation. Here the isosurfaces of pressure magnitude and phase and the wave fronts coincide. The directions of magnitude and phase gradients on the surface are the same and determine the time-invariant local direction of the velocity. In regions of one-sided magnitude isosurfaces, the directions of magnitude and phase gradients disagree. Nonetheless, also in this case the velocities are mostly oriented in the direction of propagation normal to the phase isosurfaces. Only during short time intervals near zero-crossing does the velocity momentarily change its direction. Thus even near pressure extrema, the deviation from the ideally time-invariant and frequency independent orientation of velocity vectors is minimal.

In summary, the phase isosurfaces almost exclusively depend on the shape of the ear canal and not on frequency. Hence the middle axis in the core region is independent of frequency as well. It is best derived from the regular phase isosurfaces. The middle axis remains meaningful, also in regions of one-sided magnitude isosurfaces, because it represents the orientation of velocity vectors most of the time except near the zero-crossing of the velocity. Thus one-sided isosurfaces play the role of local disturbances in an otherwise fairly regular field. Actually, the given investigation of a toroidal duct with two straight adaptors demonstrates that unidimensional transformations of acoustic field variables over irregular regions are hardly altered by one-sided isosurfaces. However, this does not mean that one-sided isosurfaces can be ignored. Actually the appearance of one-sided isosurfaces marks locations where the sound field is particularly sensitive to interference with measuring equipment such as probe tube microphones. This item will be discussed elsewhere.

ACKNOWLEDGMENTS

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) in the project Hu 352/8. The authors would like to acknowledge the valuable comments and suggestions made by two anonymous reviewers. This has considerably improved the manuscript.

[1] This representation differs from the one automatically generated by the FE program ANSYS, which provides the real part of the complex pressure instead of the magnitude. Therefore, the diagram produced by ANSYS shows the pressure distribution at a certain moment. In contrast, the magnitude of the pressure phasor used throughout this paper characterizes the time-invariant spatial distribution as standing wave pattern.

Agullo, J., Barjau, A., and Keefe, D. H. (**1999**). "Acoustic propagation in flaring, axisymmetric horns: I. A new family of unidimensional solutions," Acust. Acta Acust. **85**, 278–284.

Farmer-Fedor, B. L., and Rabbitt, R. D. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.

Funnell, W. R. J., and Decraemer, W. F. (**1996**). "On the incorporation of moiré shape measurements in finite-element models of the cat eardrum," J. Acoust. Soc. Am. **100**, 925–932.

Hudde, H. (**1983**). "Estimation of the area function of human ear canals by sound pressure measurements," J. Acoust. Soc. Am. **73**, 24–31.

Hudde, H., and Engel, A. (**1998a**). "Measuring and modeling basic properties of the human middle ear and ear canal. Part I: Model structure and measuring techniques," Acust. Acta Acust. **84**, 720–738.

Hudde, H., and Engel, A. (**1998b**). "Measuring and modeling basic properties of the human middle ear and ear canal. Part II: Ear canal, middle ear cavities, eardrum, and ossicles," Acust. Acta Acust. **84**, 894–913.

Hudde, H., and Engel, A. (**1998c**). "Measuring and modeling basic properties of the human middle ear and ear canal. Part III: Eardrum impedances, transfer functions, and complete model," Acust. Acta Acust. **84**, 1091–1108.

Rabbitt, R. D., and Friedrich, M. T. (**1991**). "Ear canal cross-sectional pressure distributions: Mathematical analysis and computation," J. Acoust. Soc. Am. **89**, 2379–2390.

Rabbitt, R. D., and Holmes, M. H. (**1988**). "Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane," J. Acoust. Soc. Am. **83**, 1064–1080.

Stinson, M. R. (**1985a**). "Spatial variation of phase in ducts and the measurement of acoustic energy reflection coefficients," J. Acoust. Soc. Am. **77**, 386–393.

Stinson, M. R. (**1985b**). "The spatial distribution of sound pressure within scaled replicas of the human ear canal," J. Acoust. Soc. Am. **78**, 1596–1602.

Stinson, M. R., and Daigle, G. A. (**2005**). "Comparison of an analytic horn equation approach and a boundary element method for the calculation of sound fields in the human ear canal," J. Acoust. Soc. Am. **118**, 2405–2411.

Stinson, M. R., and Daigle, G. A. (**2007**). "Transverse pressure distributions in a simple model ear canal occluded by a hearing aid test fixture," J. Acoust. Soc. Am. **121**, 3689–3702.

Stinson, M. R., and Khanna, S. M. (**1989**). "Sound propagation in the ear canal and coupling to the eardrum, with measurements on model systems," J. Acoust. Soc. Am. **85**, 2481–2491.

Tuck-Lee, J. P., Pinsky, P. N., Steele, C. R., and Puria, S. (**2008**). "Finite element modeling of acousto-mechanical coupling in the cat middle ear," J. Acoust. Soc. Am. **124**, 348–362.

Webster, A. G. (**1919**). "Acoustical impedance, and the theory of horns and of the phonograph," Proc. Natl. Acad. Sci. U.S.A. **5**, 275–282.

Weistenhöfer, Ch. (**2002**). "Funktionale Analyse des menschlichen Mittelohres durch dreidimensionale Messung und Modellierung (Functional analysis of the human middle ear based on three-dimensional measurements and models)," Ph.D. thesis, Bochum, Germany.

Weistenhöfer, Ch., and Hudde, H. (**1999**). "Determination of the shape and inertia properties of the human auditory ossicles," Audiol. Neuro-Otol. **4**, 192–196.

Otoacoustic emissions evoked by 0.5 kHz tone bursts

W. Wiktor Jedrzejczak, Artur Lorens, Anna Piotrowska,
Krzysztof Kochanek, and Henryk Skarzynski
Institute of Physiology and Pathology of Hearing, ul. Zgrupowania AK “Kampinos” 1, 01-943 Warszawa, Poland

(Received 11 December 2008; revised 2 February 2009; accepted 10 February 2009)

The aim of this research is to extend previous studies of the time-frequency features of otoacoustic emissions (OAEs) using information about the properties of the signals at low frequencies. Responses to 0.5 kHz tone bursts were compared to OAEs that were evoked by click stimuli and by 1, 2, and 4 kHz tone burst stimuli. The OAEs were measured using 20 and 30 ms intervals between stimuli. The analysis revealed no differences in the time-frequency properties of 1, 2, and 4 kHz bursts measured using these two different acquisition windows. However, at 0.5 kHz the latency of the response was affected significantly if a shorter time window was used. This was caused by the fact that the response reached a maximum after an average time of 15.4 ms, and lasted a few milliseconds longer. Therefore, for this particular stimulus, the use of a 30 ms time window seems more appropriate. In addition, as an example of the possible application of low-frequency OAEs, signals were measured in patients suffering from partial deafness, characterized by steep audiograms with normal thresholds up to 0.5 kHz and almost total deafness above this frequency. Although no response to clicks was observed in these subjects, the use of 0.5 kHz tone bursts did produce OAEs. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097464]

PACS number(s): 43.64.Jb, 43.60.Hj [BLM] Pages: 3158–3165

I. INTRODUCTION

Otoacoustic emissions (OAEs) are well established in clinical practice and are a valuable tool in basic research (as reviewed by Probst *et al.*, 1991). The most commonly used OAEs are click-evoked OAEs (CEOAEs) and distortion product OAEs (DPOAEs), which are evoked by two tonal stimuli. OAEs are believed to be good predictors of hearing status, particularly in the 1–4 kHz range. However, while CEOAEs are better indicators of cochlear function at 1 kHz, DPOAEs perform more satisfactorily at 4 kHz. Both CEOAEs and DPOAEs perform poorly at 0.5 kHz (e.g., Gorga *et al.*, 1993; Suckfüll *et al.*, 1996; Harrison and Norton, 1999). This is mostly due to the problems associated with various types of noise and middle ear effects. Another difficulty is related to the use of high pass filters in acquisition systems, which eliminates some of the response at frequencies as low as 0.5 kHz (Prieve *et al.*, 1993; Hurley and Musiek, 1994). For these reasons, the characteristics of low-frequency emissions are described much less commonly in the literature than those in the 1–4 kHz range.

OAEs are known to be frequency-specific. In most cases, OAEs do not occur when hearing loss is greater than 35 dB HL. However, if hearing is preserved at a certain frequency, there is then the possibility of recording OAEs at this specific frequency (e.g., Kemp *et al.*, 1986; Probst *et al.*, 1987). It is the basal region of the cochlea that is most prone to damage. This is reflected by the fact that reduction in OAE amplitudes or even disappearance of OAEs starts at higher frequencies (e.g., Lonsbury-Martin *et al.*, 1991). Loss of hearing at high frequencies can also influence emissions at frequencies lower than those corresponding to the actual hearing loss. For example, Murnane and Kelly (2003) showed that the levels of CEOAEs for subjects with hearing loss at frequencies greater than 2 kHz were lower in the low-frequency region than the levels of CEOAEs obtained from ears with normal hearing, even though there were no significant differences at low-frequency, pure-tone thresholds.

Transiently evoked OAEs can also be elicited by so-called “tone bursts,” which are short stimuli centered at a certain frequency. Several studies (e.g., Elberling *et al.*, 1985; Probst *et al.*, 1986) have shown that the spectral content of a sum of tone burst-evoked OAEs (TBOAEs) centered in the 1–4 kHz corresponds to CEOAE. It is nevertheless the case that for tone bursts, the same level of stimulus as the equivalent click gives a higher level of OAE (e.g., Prieve *et al.*, 1996).

The idea of making use of tone burst stimuli in clinical practice resurfaces in the literature from time to time (e.g., Hauser *et al.*, 1991; Lichtenstein and Stapells, 1996; McPherson *et al.*, 2006). For example, Lichtenstein and Stapells (1996) compared hearing thresholds with OAEs evoked using tone bursts as well as clicks. They concluded that 0.5 kHz tone bursts could act as much better stimuli than clicks when searching for responses at this frequency. They also showed that OAEs evoked by 0.5 kHz tone bursts provide a better estimate of hearing status at 0.5 kHz than those evoked by clicks. Recently Zhang *et al.* (2008a) proposed the use of 1 kHz bursts as a useful method for investigating cochlear function at specific frequency ranges in neonates.

The time-frequency (TF) characteristics of CEOAEs have been studied extensively using a range of different methods (e.g., Wit *et al.*, 1994; Ozdamar *et al.*, 1997; Hatzopoulos *et al.*, 2000; Jedrzejczak *et al.*, 2004; Marozas *et al.*, 2006). These methods have been used to identify the general properties of CEOAEs in healthy subjects (e.g., Tognola *et al.*, 1997; Jedrzejczak *et al.*, 2006; Zhang *et al.*, 2008b), as well as the influence of disturbing factors (e.g., Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2005; Paglialonga *et al.*, 2007). The TF characteristics of emissions evoked by tone bursts have been analyzed in this way less often (e.g., Konrad-Martin and Keefe, 2003; Jedrzejczak *et al.*, 2004; Jedrzejczak *et al.*, 2008b). In Jedrzejczak *et al.* (2004), CEOAEs were compared with the responses to half-octave tone bursts at 1–4 kHz. The results showed similar TF patterns between CEOAEs and TBOAEs and similar frequency-latency relationships. However, in that work as well as in the other mentioned above TF studies the low frequencies of OAEs were not investigated.

The aim of the study reported herein was to extend the results of previous TF studies of OAEs (and especially those from Jedrzejczak *et al.*, 2004) using information about the properties of OAE responses in the 0.5 kHz region. The 0.5 kHz TBOAEs were compared with emissions evoked by clicks and higher-frequency tone bursts. In addition, an investigation was carried out into the dependency of the response on the duration of the acquisition window. It was suggested by Prieve *et al.* (1993) that poor results from CEOAEs at a frequency of 0.5 kHz could be at least partially due to the fact that the duration of the acquisition window was too short. Other studies (e.g., Norton and Neely, 1987) have shown that 0.5 kHz TBOAEs can last longer than 20 ms. For these reasons, all measurements were performed using a standard 20 ms time window as well as a 30 ms time window. The potential clinical application of 0.5 kHz TBOAEs was tested on a group of partially deaf patients with severe high-frequency hearing loss starting at 0.5 kHz.

II. EXPERIMENTAL PROCEDURE

OAEs from both ears of ten subjects (five females and five males, aged 27–35 years) were measured in low-noise ambient conditions using an ILO-96 apparatus (Otodynamics Ltd., London) running software Version 5.

These subjects were laryngologically healthy and did not have any otoscopic ear abnormalities. Impedance audiometry tests revealed normal type A tympanograms and normal acoustic reflexes. In pure-tone audiometry, hearing thresholds were better than 20 dB HL for all tested frequencies (0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz).

Standard click stimuli and a set of 0.5, 1, 2, and 4 kHz tone bursts (average amplitude of −80 ± 3 peak dB SPL, nonlinear averaging protocol) were used to evoke a total of 520 OAE responses. Previous studies (e.g., Norton and Neely, 1987) have shown that lower level stimuli could not be sufficient to evoke OAEs in 0.5 kHz. Therefore relatively high levels for all stimuli were used, similar to Lichtenstein and Stapells (1996). The tone bursts were four cycles long with equal rise/fall times and no plateau. The initial part of the response was windowed automatically by the system to minimize the artifact of stimuli. The onset of the window was 2.5 ms for click, and 10, 5, 2.5, and 1.25 ms for consecutive tone bursts of different frequencies. All recordings were performed in two acquisition windows: the standard—ending at 20 ms and a longer—ending at 30 ms.

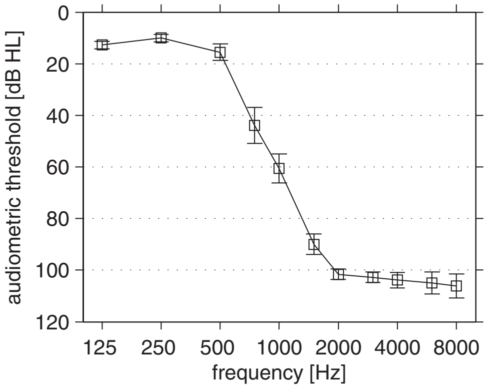


FIG. 1. Averaged audiogram for a group of hearing-impaired subjects.

The synchronized spontaneous otoacoustic emission (SSOAE) spectra were also measured using the default ILO protocol. In accordance with this protocol, the OAEs evoked by click stimuli of 80 dB SPL were recorded in an 80 ms window. The first 20 ms of each averaged response were discarded and the responses from the following 60 ms were considered to be synchronized spontaneous emissions.

In addition, OAEs were measured in nine ears of six subjects (three females and three males, aged 26–66) with a hearing impairment called partial deafness (Skarzynski *et al.*, 2003). This type of hearing loss is characterized by normal thresholds at low frequencies and almost total deafness at high frequencies. The average audiogram for all the tested ears is shown in Fig. 1. For these subjects, measurements of OAEs evoked by clicks were performed using a standard 20 ms window and those evoked by 0.5 kHz tone bursts using a 30 ms window (average amplitude of −80 ± 3 peak dB SPL). SSOAE tests showed an absence of spontaneous components.

III. METHOD

The OAE signals were analyzed using the matching pursuit (MP) method (introduced by Mallat and Zhang (1993). The MP algorithm relies on the adaptive decomposition of the signal into waveforms (also called atoms) from a large and redundant set of functions (called dictionary). Finding the optimal approximation of a signal by selecting functions from such a set of functions is computationally intractable problem, and for this reason sub-optimal solutions were applied. The waveforms were fitted using an iterative procedure, starting with the waveform that accounted for the largest part of the signal energy. Then, the next waveforms were fitted to the residues. This procedure was continued until 99.5% of the original signal energy was accounted for. A stochastic dictionary proposed by Durka *et al.* (2001) was used, which was composed of 10^6 Gabor functions (sine-modulated Gaussians), described by parameters such as frequency, latency, time span, amplitude, and phase. In Jedrzejczak *et al.* (2006) it has been shown that most of the energy of CEOAE responses from healthy adults can be described efficiently by seven to ten waveforms (resonant modes). These components are characteristic not only for each subject but also for each ear. A detailed description of the MP procedure may be found in Jedrzejczak *et al.* (2004). The

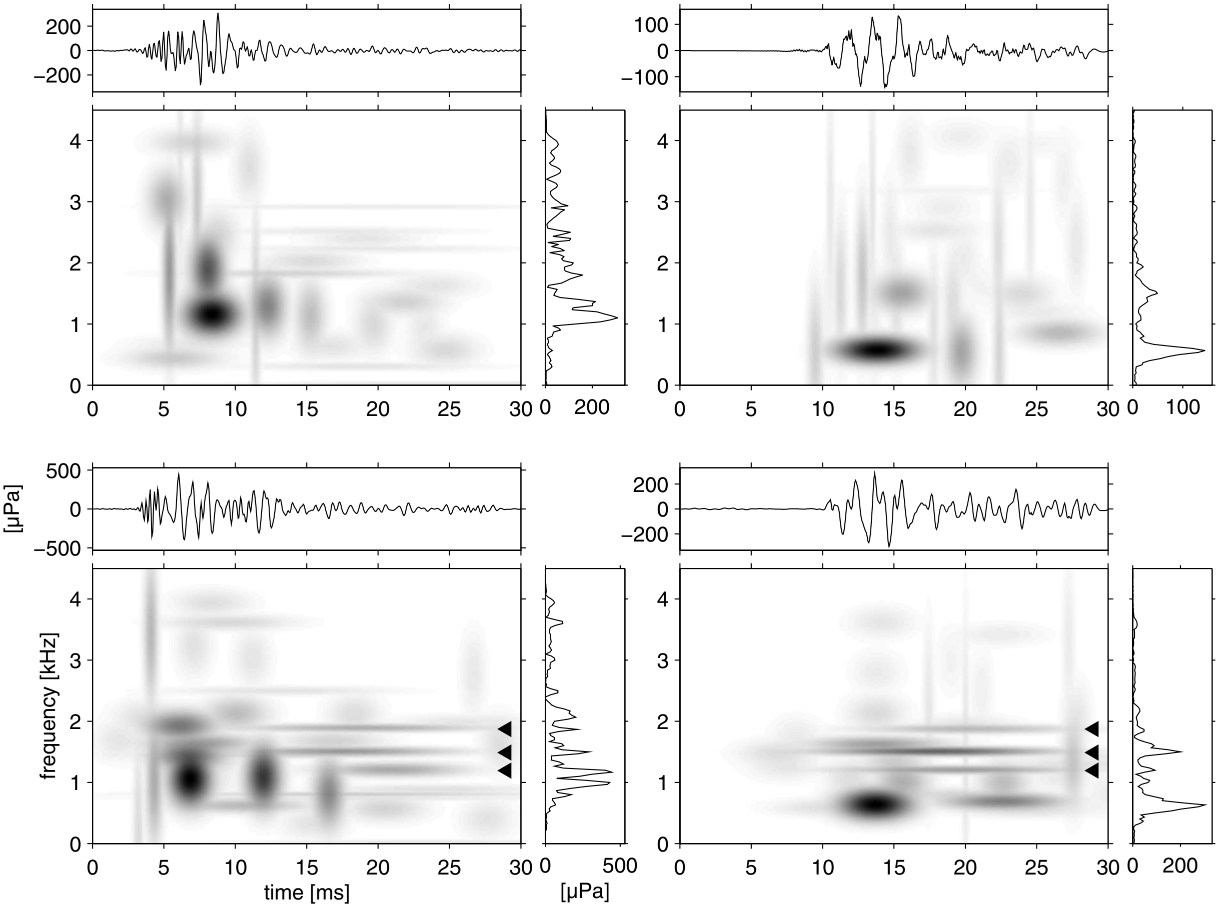


FIG. 2. Examples of TF distributions of energy of OAEs for subjects without SOAEs (top panels) and with SOAEs (bottom panels). From left to right: response to click and response to 0.5 kHz tone burst. Triangles mark the frequencies of SSOAEs. In addition, the time evolution of signal is plotted above, and corresponding spectrum is shown to the right of each TF map.

relationship between the waveforms determined by the MP procedure and the resonant modes of OAEs was determined in Blinowska *et al.* (2007).

Each single-frequency waveform found by the MP algorithm is described using physically meaningful parameters in the following ways: (i) the amplitude is determined as the maximum of the modulus of the waveform at a certain frequency, (ii) the latency is measured as the time taken from onset of the stimulus to the maximum point in the waveform envelope, and (iii) the time span parameter is defined as half the width of the waveform envelope and can be interpreted as the duration of the waveform.

The TF distributions of the signal energy were constructed by superimposing the Wigner transforms of individual waveforms. In this way, the distribution is free of cross-terms, which are a cause of blurring in the TF maps obtained using the Wigner–Ville, Choi–Williams, or other similar approaches (e.g., Jedrzejczak *et al.*, 2004).

The Wilcoxon rank sum test was used for the statistical analysis of the data with the criterion of significance set at $p<5\times10^{-2}$. This test is an equivalent of the student's *t*-test and is generally used when the populations that are analyzed do not have normal distributions.

IV. RESULTS

Examples of the TF analysis of OAEs evoked by click and 0.5 kHz tone burst for two subjects with normal hearing, the first without SOAEs and the second with SOAEs, are shown in Fig. 2. The TF distributions of the signal energy are shown for OAEs measured using 30 ms time windows. Both click and TBOAEs are usually composed of a few dominant components. The evoked part of the signal is represented by components of various durations and frequency spans. SOAEs are visible in the form of long, narrow components that appear on TF maps for both click and TBOAEs. In the case of responses to clicks in both subjects, there is not much energy below 1 kHz and the evoked part of the signal ends before 20 ms. However, in the ear with SOAEs, spontaneous components continue until the end of the acquisition window. Most of the energy of the 0.5 kHz TBOAE is close to the stimulus frequency. It may be seen that in this case, some components exceed 20 ms, which is the standard window in most measurement setups. TF maps for higher-frequency TBOAEs are not shown because these were studied extensively in Jedrzejczak *et al.* (2004). Furthermore, it was shown that in general these are similar to parts of OAE maps for click stimuli in the corresponding frequency bands.

Figure 3 shows the average values of the parameters of OAEs determined by means of the MP method for clicks and tone bursts in 20 and 30 ms acquisition windows. The values were determined for the highest energy component in each octave frequency band. Here, linear filters were not used to obtain frequency band information from clicks. Instead, components from given bands were selected because the MP

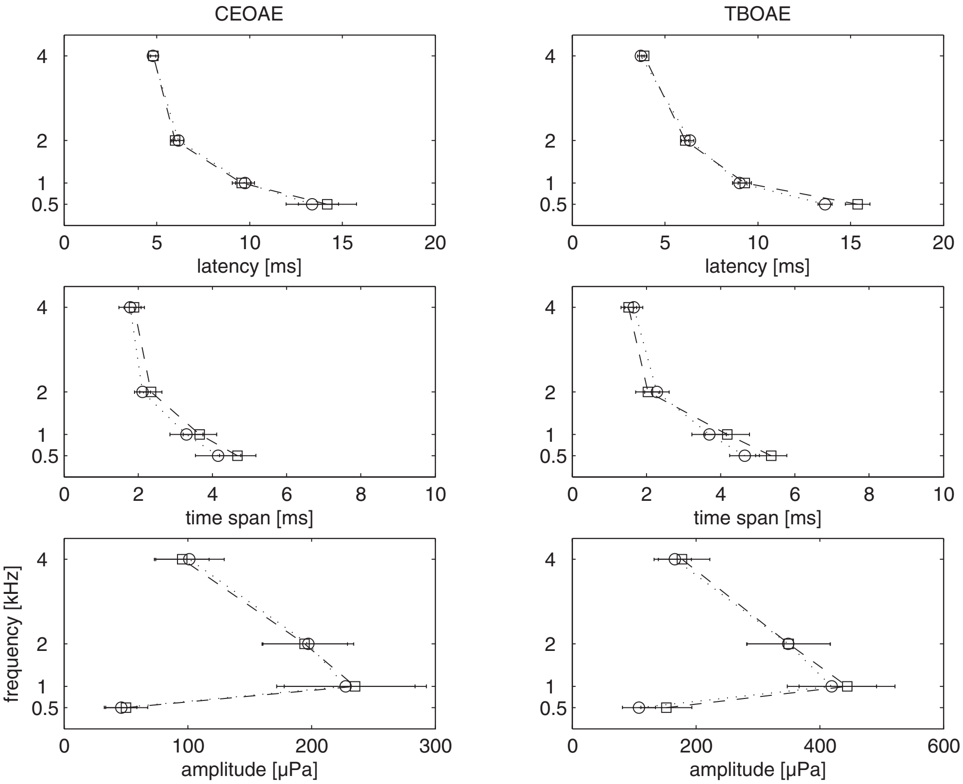


FIG. 3. Average latency, time span, and amplitude of click (left-hand panels) and TBOAEs (right-hand panels) for octave frequencies of 0.5, 1, 2, and 4 kHz. The parameters were calculated for signals measured in 20 ms (circle) and 30 ms windows (rectangle). The bars represent standard errors.

method determines their exact frequency. As mentioned previously, spontaneous activity was identified for each ear by SSOAE measurement. Components that reflect SOAEs (i.e., long and narrow waveforms as seen in Fig. 2, bottom panels) can significantly alter the estimates of the parameters of the evoked signal (Jedrzejczak *et al.*, 2007). For this reason, these components were removed from the process of estimating the average values of MP parameters.

The properties of the clicks recorded in the 20 and 30 ms windows are very similar (left column of Fig. 3). There is a good repeatability of the responses for the 1, 2, and 4 kHz octave bands. The latencies range from 5 ms (at 4 kHz) to 10 ms (at 1 kHz), in agreement with the results of previous studies that were performed using the same method (Jedrzejczak *et al.*, 2004; Jedrzejczak *et al.*, 2005). However, the results from the 0.5 kHz band are not very reliable. In this band the standard error for latency is higher than in other frequency bands, which means that there is a greater spread of values at this frequency. The average amplitude in the 0.5 kHz band is significantly lower (at $p<0.02$) than even the lowest amplitude in the 4 kHz band. The use of the 30 ms window seems to have little effect on the CEOAEs properties. It did not improve OAE performance at 0.5 kHz.

The main difference between responses to clicks and tone bursts is that the amplitude is higher for TBOAEs (bottom plots of Fig. 3). This may be expected because tone bursts have a smaller bandwidth and thus a stimulus of the same intensity has a greater spectral density than the equivalent click stimulus. This results in TBOAEs having a higher amplitude than the CEOAE in that frequency band. This is important, particularly in the case of the 0.5 kHz frequency band. Despite the fact that the amplitude for TBOAE in this band is still the lowest, it is now very close to the TBOAE amplitude in the 4 kHz band.

The latencies and time spans in the 1–4 kHz range showed similar results for tone bursts and clicks (top and middle plots of Fig. 3). There are slight differences, which are caused mainly by different durations of click and tone burst stimuli. As may be expected, using a 20 ms window only affected the response for the 0.5 kHz burst. The latency of the response was significantly shorter (at $p<0.05$) than when a 30 ms window was used. For longer acquisition windows, an average latency of 15.4 ± 0.9 ms was obtained. The differences in time span and amplitude between the two acquisition windows for the 0.5 kHz octave band were not significant. However, they were higher than for the 1–4 kHz band.

The most common method of detecting the presence of OAEs and assessing the quality of their measurement is by means of a reproducibility parameter. This is defined as the correlation between the two buffers of the sub-averages of the single responses (Kemp *et al.*, 1986). Reproducibility values were calculated for all tone burst and click stimuli measured using 20 and 30 ms acquisition windows. For all click and tone burst stimuli (even 0.5 kHz) the values for the same types of responses measured using different acquisition windows differed by only a few percent, and were not significant. However, while reproducibility values for OAEs evoked by click and 1–4 kHz tone bursts were on average at the level of 90%, the equivalent values for 0.5 kHz TBOAEs were lower.

In the context of the present study, the most interesting feature seems to be the comparison between results from the more frequently used click stimulus and the 0.5 kHz tone burst. For this reason, the growth in the reproducibility as a function of the number of averages used was investigated for signals evoked by these stimuli. All single responses were

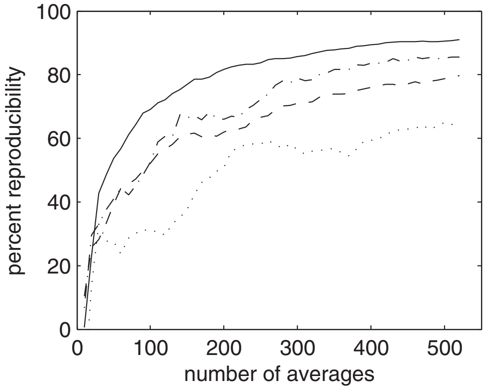


FIG. 4. Average reproducibility for normal hearing subjects as a function of the number of averaged responses for OAEs evoked by click (solid line), click filtered using a 0.5 kHz octave band (dotted line), 0.5 kHz tone burst (dashed line), and 0.5 kHz tone burst using octave band filtering (dash-dotted line).

recorded for off-line analysis, with 520 responses in each of the two buffers. This is the double of the default number of averages in the ILO system, and it was used to determine the improvement that could be achieved using a longer measurement regime. The results averaged over the group of ten subjects (20 ears) are shown in Fig. 4. In this case, the response to three clicks of the same polarity, and one with reversed polarity and three times higher amplitude, is treated as one response. This is standard for most acquisition systems and is known as the non-linear mode (as described by Kemp *et al.*, 1986). It can be seen that the reproducibility values for the 0.5 kHz bursts are smaller by around 10% than when using clicks. The extension of the averaging process from 260 responses to 520 (using clicks) improved the reproducibility by 6% on average, while for the 0.5 kHz burst, it is improved by 12%. This would suggest that better results could in general be obtained by extending the averaging process in the case of the 0.5 kHz TBOAE measurement. The results for clicks in 0.5 kHz octave band were poor, while filtering of the responses to 0.5 kHz tone bursts improved reproducibility by a few percent. The reproducibility of the filtered CEOAEs was 20% lower than the equivalent results using 0.5 kHz TBOAE.

The clicks and 0.5 kHz bursts were also used to verify the hearing status of patients with partial deafness. The averaged audiogram for the tested ears of these patients is shown in Fig. 1. An example of the TF maps of the signals measured after a click and a 0.5 kHz tone burst for one partially deaf subject is presented in Fig. 5. The response to the click contains mostly noise, and the characteristic pattern of the components of the responses with decreasing frequency and increasing latency (as seen in Fig. 2) is not observed. There are no components in the 20–30 ms range because in this case, only the responses in the 20 ms window were measured using click stimuli. The map is shown using a duration of 30 ms in order to preserve the same scale as for the 0.5 kHz burst and the plots shown in Fig. 2. The responses to the 0.5 kHz bursts are similar to those for subjects with normal hearing. The reproducibility values for clicks and 0.5 kHz bursts for normal and partially deaf subjects are summarized in Table I. The reproducibility of the responses to clicks for partial deafness is very low, which indicates that these stimuli did not produce OAEs in these subjects. In contrast, the reproducibility of the unfiltered responses to the 0.5 kHz bursts is only slightly lower than that obtained for normal subjects. This difference is probably due mainly to the fact that the high-frequency responses in partially deaf subjects consist mostly of noise. After octave band filtering, the reproducibility of the 0.5 kHz TBOAEs for normal and partially deaf subjects was similar. In addition, the TF parameters of 0.5 kHz TBOAEs for partially deaf subjects have values similar to those of a normal group (latency of 14.5 ± 0.9 and time span of 6.2 ± 0.9). The average amplitude was slightly lower than for the normally-hearing group, but this difference was not significant.

V. DISCUSSION

In the study reported herein, the TF analysis of TBOAEs as conducted by Jedrzejczak *et al.* (2004) was extended using information from the 0.5 kHz frequency region. As was the case for previous studies of 0.5 kHz TBOAEs (e.g., Lichtenstein and Stapells, 1996) more reliable signals were recorded when the acquisition window was extended from 20 to 30 ms. The 20 ms window discards a few milliseconds

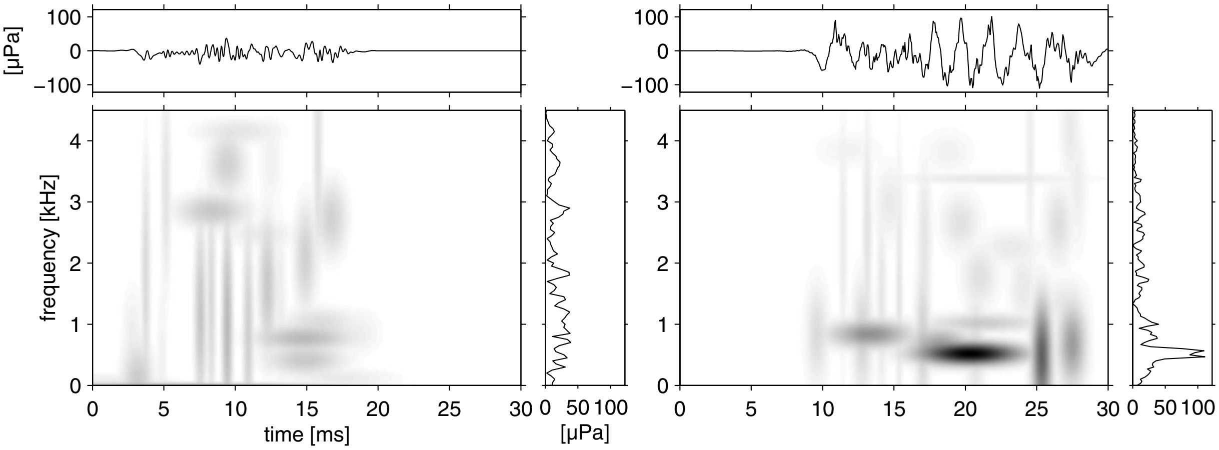


FIG. 5. Example of the TF distributions of the energy of OAEs for a partially deaf subject. From left to right: response to click and response to 0.5 kHz tone burst. In addition, the time evolution of signal is plotted above, and corresponding spectrum is shown to the right of each TF map.

TABLE I. Average reproducibility values (±standard error) of OAEs evoked by clicks and 0.5 kHz tone bursts for normal hearing and partial deafness subjects. In fourth and fifth columns values are shown for OAEs filtered ±half-octave around 0.5 kHz.

	No. of ears	Click	0.5 kHz burst	Click filtered	0.5 kHz burst filtered
Normal hearing	20	91 ± 3	80 ± 5	65 ± 7	86 ± 4
Partial deafness	9	26 ± 4	66 ± 5	42 ± 14	91 ± 2

of the response at 0.5 kHz, which influences the estimation of latency (Fig. 3). For the case of clicks and 1, 2, and 4 kHz tone bursts, OAEs had similar amplitude and TF properties when measured in acquisition windows of different durations. The differences were very small, mainly due to minor variations in amplitude between measurements. This is not unexpected, given that the purely evoked part of these responses is at most 20 ms long. Nevertheless, when synchronized spontaneous OAEs or other long-lasting components are present, both CEOAEs and TBOAEs can exceed 20 ms (e.g., Probst *et al.*, 1986). This is where the application of the MP method can be an advantage because it can be useful for the separation of the evoked and spontaneous components (Jedrzejczak *et al.*, 2008a).

The OAE amplitude characteristic presented here (bottom plots of Fig. 3) was similar to that seen in previous studies (e.g., Probst *et al.*, 1991, Lonsbury-Martin *et al.*, 1991), i.e., the greatest response was in the 1–2 kHz range. The amplitude is much smaller at 0.5 kHz than at other frequencies, particularly in the case of the response to clicks. This feature, together with the low reproducibility values obtained, indicates that clicks are not a reliable method of evaluating OAE status at 0.5 kHz. This is probably due to the lower middle ear transfer at about 0.5 kHz in relation to higher frequencies (e.g., Puria, 2003). The other contributing factors could be measurement-noise bias (Backus, 2007) and the low-frequency filters used in the ILO equipment (Prieve *et al.*, 1993; Hurley and Musiek, 1994). Use of the 20 ms window seems to have a negative effect only in the case of emissions evoked by 0.5 kHz tone bursts. The results in the low-frequency region using clicks are equally poor for both the window durations used. For this reason, the latency estimates in this region are also not very reliable.

The TBOAE latency in the 0.5 kHz region estimated here was, on average, 15.4 ms. This is slightly lower than the values reported by Prieve *et al.* (1996). However, the highest stimulus level used in their study was 10 dB SPL lower than that used here. It is known that the latency of OAEs lengthens as the intensity of the stimulus decreases (e.g., Norton and Neely 1987; Tognola *et al.*, 1997; Sisto and Moleti, 2007). The other feature that could influence the results are SOAEs. These are present in a majority of ears (e.g., Penner and Zhang, 1997) and it is known that synchronized SOAEs can dominate evoked responses (e.g., Probst *et al.*, 1986). Recently, it has been shown that synchronized SOAE components can also be responsible for variations in latency (Jedrzejczak *et al.*, 2007). In our study, therefore, the components that reflected spontaneous activity were removed from the latency calculation, thus restricting the differences between individual latency values. It was not possible to do this in previous studies, which used much simpler methods of analysis.

Tognola *et al.* (1997) proposed a power-law fit for the latency-frequency relationship in CEOAEs. This was later confirmed in several studies (e.g., Sisto and Moleti, 2002). The power-law fit also yielded good results for the latency-frequency relationship for 1–4 kHz tone bursts (Jedrzejczak *et al.*, 2004). The proposed function predicted that at 0.5 kHz, the latency should be on average 15 ms, which is consistent with the experimental results presented here.

It is already known that CEOAEs show good specificity in relation to hearing loss (e.g., Lonsbury-Martin *et al.*, 1991). Even if, at most frequencies, a subject's thresholds are below norm when hearing is preserved at least at one frequency, it is still possible to measure OAEs. The presence of emissions has been found when hearing loss started at 1 kHz (e.g., Robinette, 2003). Here results for subjects with even greater hearing loss (normal hearing only up to 0.5 kHz) were shown. The reproducibility values were considerably lower than 50%, which proved that CEOAEs were not present. However, it was possible to measure 0.5 kHz TBOAEs. The responses for these subjects had TF properties similar to the 0.5 kHz TBOAEs of normal subjects.

We showed in our study that 0.5 kHz TBOAEs provide additional information to standard CEOAEs. However, the measurements require a very quiet environment and yield lower reproducibility values on average than those of CEOAEs, although by doubling the standard averaging time the reproducibility can be improved by up to 10%. This technique, combined with the use of a 30 ms window, causes a slight lengthening of the measurement time in comparison with the standard CEOAE test. However, the technique could be restricted in its use to those cases where the click test failed. Recently, McPherson *et al.* (2006) suggested a similar procedure for 1 kHz burst and showed that the lowest referral rates in newborn screening were achieved when a 1 kHz TBOAE was measured following a lack of response to click stimuli.

The CEOAE is absent in a majority of patients who undergo standard cochlear implantation. The presence of a 0.5 kHz OAE could be used as one of the tests in the qualification procedure for partial deafness cochlear implantation (PDCI) introduced by Skarzynski *et al.* (2003). The PDCI procedure differs from standard cochlear implantation in that its aim is to preserve low-frequency hearing for combined electric acoustic stimulation (EAS). The PDCI technique is a more efficient method of treatment than cochlear implantation without hearing preservation for a selected group of subjects with residual hearing (e.g., Lorens *et al.*, 2008). The 0.5 kHz TBOAE test could be particularly helpful when pure-tone audiometry is not possible.

It has been shown that for tone bursts (Lichtenstein and Stapells, 1996) and continuous stimulation (Ellison and Keefe, 2005), the best properties of OAEs are achieved at the frequency that corresponds to the stimulus. This points to the importance of filtering out the lower and higher parts of the signal. The problems could be noise, SOAEs, and other evoked components, arising possibly due to non-linear intermodulation distortion (Yates and Withnell, 1999). Similar conclusions were drawn by Chan and McPherson (2000). They found that half-octave band analysis at the frequency corresponds to the stimulus reflected TBOAE performance more reliably than broadband analysis. Here, filtering also improved reproducibility values for 0.5 kHz TBOAEs. This was particularly the case for partial deafness subjects in which the higher frequencies of response were occupied mostly by noise.

VI. CONCLUSIONS

TF analysis was used to characterize CEOAEs and TBOAEs in different acquisition windows. The main area of interest was the low-frequency range of responses around 0.5 kHz. The general TF properties of the responses in this region are similar to those of higher frequency TBOAEs. These usually only consist of a few components, and their main distinctive feature is a very long latency. The 0.5 kHz TBOAE is more reliable than CEOAE in the case of activity at low frequencies. Even the use of a 30 ms window in the measurement of CEOAEs did not improve the detection of 0.5 kHz activity. On the other hand, this window seems better suited to 0.5 kHz TBOAEs, given that its duration exceeds 20 ms in most cases.

The reproducibility values for low-frequency responses are significantly lower than those obtained using standard wideband click stimuli. Therefore the prolongation of measurement and/or the lowering detection criteria should be considered. Nevertheless, 0.5 kHz TBOAE is a promising tool for the detection of emissions in patients with deep, high-frequency hearing loss when click stimuli do not produce OAEs.

ACKNOWLEDGMENTS

We thank David T. Kemp for providing information about the structure of the data files created by the ILO equipment. We would also like to thank two anonymous reviewers who provided many helpful suggestions that improved the manuscript. This research was partially funded by the Polish Ministry of Science and Higher Education within the framework of the 2007–2010 Scientific Support Program (Grant No. N N518 0924 33).

Backus, B. C. (**2007**). “Bias due to noise in otoacoustic emission measurements,” J. Acoust. Soc. Am. **121**, 1588–1603.

Blinowska, K. J., Jedrzejczak, W. W., and Konopka, W. (**2007**). “Resonant modes of otoacoustic emissions,” Physiol. Meas. **28**, 1293–1302.

Chan, R. H., and McPherson, B. (**2000**). “Test-retest reliability of tone-burst-evoked otoacoustic emissions,” Acta Oto-Laryngol. **120**, 825–834.

Durka, P. J., Ircha, D., and Blinowska, K. J. (**2001**). “Stochastic time-frequency dictionaries for matching pursuit,” IEEE Trans. Signal Process. **49**, 507–510.

Elberling, C., Parbo, N. J., Johnsen, N. J., and Bagi, P. (**1985**). “Evoked acoustic emissions: Clinical application,” Acta Oto-Laryngol., Suppl. **99**, 77–85.

Ellison, J. C., and Keefe, D. H. (**2005**). “Audiometric predictions using stimulus-frequency otoacoustic emissions and middle ear measurements,” Ear Hear. **26**, 487–503.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). “A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects,” J. Acoust. Soc. Am. **94**, 2639–2648.

Harrison, W. A., and Norton, S. J. (**1999**). “Characteristics of transient evoked otoacoustic emissions in normal-hearing and hearing-impaired children,” Ear Hear. **20**, 75–86.

Hatzopoulos, S., Cheng, J., Grzanka, A., and Martini, A. (**2000**). “Time-frequency analyses of TEOAE recordings from normals and SNHL patients,” Audiology **39**, 1–12.

Hauser, R., Probst, R., and Löhle, E. (**1991**). “Click- and tone-burst-evoked otoacoustic emissions in normally hearing ears and in ears with high-frequency sensorineural hearing loss,” Eur. Arch. Otorhinolaryngol. **248**, 345–352.

Hurley, R. M., and Musiek, F. E. (**1994**). “Effectiveness of transient-evoked otoacoustic emissions (TEOAEs) in predicting hearing level,” J. Am. Acad. Audiol. **5**, 195–203.

Jedrzejczak, W. W., Blinowska, K. J., Kochanek, K., and Skarzynski, H. (**2008b**). “Synchronized spontaneous otoacoustic emissions analyzed in a time-frequency domain,” J. Acoust. Soc. Am. **124**, 3720–3729.

Jedrzejczak, W. W., Blinowska, K. J., and Konopka, W. (**2005**). “Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise,” Hear. Res. **205**, 249–255.

Jedrzejczak, W. W., Blinowska, K. J., and Konopka, W. (**2006**). “Resonant modes in transiently evoked otoacoustic emissions and asymmetries between left and right ear,” J. Acoust. Soc. Am. **119**, 2226–2231.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). “Identification of otoacoustic emission components by means of adaptive approximations,” J. Acoust. Soc. Am. **115**, 2148–2158.

Jedrzejczak, W. W., Hatzopoulos, S., Martini, A., and Blinowska, K. J. (**2007**). “Otoacoustic emissions latency difference between full-term and preterm neonates,” Hear. Res. **231**, 54–62.

Jedrzejczak, W. W., Smurzynski, J., and Blinowska, K. J. (**2008a**). “Origin of suppression of otoacoustic emissions evoked by two-tone bursts,” Hear. Res. **235**, 80–89.

Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). “Acoustic emission cochleography–practical aspects,” Scand. Audiol. Suppl. **25**, 71–95.

Konrad-Martin, D., and Keefe, D. H. (**2003**). “Time-frequency analyses of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: Testing cochlear model predictions,” J. Acoust. Soc. Am. **114**, 2021–2043.

Lichtenstein, V., and Stapells, D. R. (**1996**). “Frequency-specific identification of hearing loss using transient-evoked otoacoustic emissions to clicks and tones,” Hear. Res. **98**, 125–136.

Lonsbury-Martin, B. L., Whitehead, M. L., and Martin, G. K. (**1991**). “Clinical applications of otoacoustic emissions,” J. Speech Hear. Res. **34**, 964–981.

Lorens, A., Polak, M., Piotrowska, A., and Skarzynski, H. (**2008**). “Outcomes of treatment of partial deafness with cochlear implantation: A DUET study,” Laryngoscope **118**, 288–94.

Mallat, S. G., and Zhang, Z. (**1993**). “Matching pursuit with time-frequency dictionaries,” IEEE Trans. Signal Process. **41**, 3397–3415.

Marozas, V., Janusauskas, A., Lukosevicius, A., and Sörnmo, L. (**2006**). “Multiscale detection of transient evoked otoacoustic emissions,” IEEE Trans. Biomed. Eng. **53**, 1586–1593.

McPherson, B., Li, S. F., Shi, B. X., Tang, J. L., and Wong, B. Y. (**2006**). “Neonatal hearing screening: Evaluation of tone-burst and click-evoked otoacoustic emission test criteria,” Ear Hear. **27**, 256–262.

Murnane, O. D., and Kelly, J. K. (**2003**). “The effects of high-frequency hearing loss on low-frequency components of the click-evoked otoacoustic emission,” J. Am. Acad. Audiol. **14**, 525–533.

Norton, S. J., and Neely, S. T. (**1987**). “Tone-burst-evoked otoacoustic emissions from normal-hearing subjects,” J. Acoust. Soc. Am. **81**, 1860–1872.

Ozdamar, O., Zhang, J., Kalayci, T., and Ulgen, Y. (**1997**). “Time-frequency distribution of evoked otoacoustic emissions,” Br. J. Audiol. **31**, 461–471.

Paglialonga, A., Tognola, G., Parazzini, M., Lutman, M. E., Bell, S. L., Thuroczy, G., and Ravazzani, P. (**2007**). “Effects of mobile phone exposure on time frequency fine structure of transiently evoked otoacoustic emissions,” J. Acoust. Soc. Am. **122**, 2174–2182.

Penner, M. J., and Zhang, T. (**1997**). “Prevalence of spontaneous otoacoustic emissions in adults revisited,” Hear. Res. **103**, 28–34.

Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). “Click- and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears,” J. Acoust. Soc. Am. **99**, 3077–86.

Prieve, B. A., Gorga, M. P., Schmidt, A., Neely, S., Peters, J., Schultes, L., and Jesteadt, W. (**1993**). “Analysis of transient-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears,” J. Acoust. Soc. Am. **93**, 3308–3319.

Probst, R., Coats, A. C., Martin, G. K., and Lonsbury-Martin, B. L. (**1986**). “Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears,” Hear. Res. **21**, 261–275.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). “A review of otoacoustic emissions,” J. Acoust. Soc. Am. **89**, 2027–2067.

Probst, R., Lonsbury-Martin, B. L., Martin, G. K., and Coats, A. C. (**1987**). “Otoacoustic emissions in ears with hearing loss,” Am. J. Otolaryngol. **8**, 73–81.

Puria, S. (**2003**). “Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions,” J. Acoust. Soc. Am. **113**, 2773–2789.

Robinette, M. S. (**2003**). “Clinical observations with evoked otoacoustic emissions at Mayo Clinic,” J. Am. Acad. Audiol. **14**, 213–224.

Sisto, R., and Moleti, A. (**2002**). “On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears,” J. Acoust. Soc. Am. **111**, 297–308.

Sisto, R., and Moleti, A. (**2007**). “Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels,” J. Acoust. Soc. Am. **122**, 2183–2190.

Skarzynski, H., Lorens, A., and Piotrowska, A. (**2003**). “A new method of partial deafness treatment,” Med. Sci. Monit. **9**, 20–24.

Suckfüll, M., Schneeweiss, S., Dreher, A., and Schorn, K. (**1996**). “Evaluation of TEOAE and DPOAE measurements for the assessment of auditory thresholds in sensorineural hearing loss,” Acta Oto-Laryngol. **116**, 528–533.

Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). “Time-frequency distributions of click-evoked otoacoustic emissions,” Hear. Res. **106**, 112–122.

Wit, H. P., van Dijk, P., and Avan, P. (**1994**). “Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions,” Hear. Res. **73**, 141–147.

Yates, G. K., and Withnell, R. H. (**1999**). “The role of intermodulation distortion in transient-evoked otoacoustic emissions,” Hear. Res. **136**, 49–64.

Zhang, V. W., McPherson, B., and Zhang, Z. G. (**2008a**). “Tone burst-evoked otoacoustic emissions in neonates: Normative data,” BMC Ear Nose Throat Disord. **8**, art. no. 3.

Zhang, Z. G., Zhang, V. W., Chan, S. C., McPherson, B., and Hu, Y. (**2008b**). “Time-frequency analysis of click-evoked otoacoustic emissions by means of a minimum variance spectral estimation-based method,” Hear. Res. **243**, 18–27.

Long-term stability of spontaneous otoacoustic emissions

Edward M. Burns
Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, Washington 98105

(Received 5 January 2009; revised 18 February 2009; accepted 19 February 2009)

Spontaneous otoacoustic emissions (SOAEs) were measured longitudinally for durations up to 19.5 years. Initial ages of the subjects ranged from 6 to 41 years. The most compelling finding was a decrease in frequency of all emissions in all subjects, which was approximately linear in %/year and averaged 0.25%/year. SOAE levels also tended to decrease with age, a trend that was significant, but not consistent across emissions, either within or across subjects. Levels of individual SOAEs might decrease, increase, or remain relatively constant with age. Several types of frequency/level instabilities were noted in which some SOAEs within an ear interacted such that their levels were negatively correlated. These instabilities often persisted for many years. SOAEs were also measured in two females over the course of their pregnancies. No changes in SOAE levels or frequencies were seen, that were larger than have been reported in females over a menstrual cycle, suggesting that levels of female gonadal hormones do not have a significant direct effect on SOAE frequencies or levels. *© 2009 Acoustical Society of America.* [DOI: 10.1121/1.3097768]

PACS number(s): 43.64.Jb, 43.66.Hg [BLM] Pages: 3166–3176

I. INTRODUCTION

Spontaneous otoacoustic emissions (SOAEs), along with the various stimulus-related OAEs, have been studied extensively since their discovery by Kemp (1979), in part because they provide a noninvasive technique for probing cochlear mechanics. Current models of SOAE generation (Talmadge *et al.*, 1998; Shera, 2003) propose that SOAEs are actively maintained cochlear standing waves, which arise from the coherent reflection of forward traveling waves in the region of maximum basilar-membrane (BM) displacement and the reflection of backward traveling waves at the stapes. In these models, potential SOAE frequencies are related to the total phase change involved in round-trip travel of a wave moving from the stapes to the region of reflection and back. These models explain the quasiperiodicity and characteristic minimal spacing of SOAE frequencies (reviewed in Talmadge *et al.*, 1998) and the influence of middle-ear properties on SOAE frequencies (Shera, 2003).

In the ultra-short term, on the order of 1 min, human SOAEs are apparently frequency-locked based on their extremely low variability in frequency and amplitude. For frequency, half-power bandwidths can be less than 0.1% of SOAE frequency; for amplitude, rms amplitude fluctuations can be less than 6.3% of mean SOAE amplitude (Van Dijk and Wit, 1990). In the short term, on the order of hours to months, SOAE frequencies and amplitudes are more variable. A major source of variability in measurements made over a single recording session are slow drifts of SOAE frequency (up or down) and amplitude (typically increases), which occur over a period of about 30 min after insertion of the microphone in either the measured ear or, in the interim between measurements, in the contralateral ear (Zurek, 1981; Rabinowitz and Widen, 1984; Whitehead, 1988, 1991), apparently as the result of tactually induced efferent effects. The average drifts in frequency are on the order of 0.1%, with maximum shifts on the order of 0.5%; average drifts in amplitude are on the order of 1.5 dB (Whitehead, 1988, 1991).

Repeated measurements over several weeks or months show frequency variability of the same order of magnitude as within-session variability, but much larger amplitude variability. Whitehead (1988) found average maximum frequency shifts of about 0.4% and average maximum amplitude shifts of about 8 dB for 19 SOAEs in one subject measured weekly for 6–8 weeks. Calculations based on the data of Fritze (1983) show a standard deviation of 0.26% for 41 measurements of a single SOAE over 4 months. Some of this weekly or monthly variability is presumably related to small shifts, less than 1%, which follow a diurnal cycle and, in the case of menstruating females, a monthly cycle (Wit, 1985; Wilson, 1986; Bell, 1992; Haggerty *et al.*, 1993; Penner, 1995).

SOAE frequencies and amplitudes are influenced by factors that change middle-ear impedance, including the transmission between the stapes and the cochlea. Elicitation of the middle-ear reflex results in frequency and amplitude shifts, which are mostly increases in frequency of less than 1% and decreases in amplitude of 25 dB or more, with the maximum shifts occurring in SOAEs with frequencies below 3 kHz (Schloth and Zwicker, 1983; Rabinowitz and Widen, 1984; Mott *et al.*, 1989; Harrison and Burns, 1993; Burns *et al.*, 1993b). Increasing or decreasing the static pressure in the ear canal (Kemp, 1979; Wilson and Sutton, 1981; Schloth and Zwicker, 1983; Whitehead, 1988), or increasing or decreasing ambient pressure (Hauser *et al.*, 1993) results in a similar pattern of frequency and amplitude shifts; however, frequency shifts from static pressure changes may be as large as 5% (Kemp, 1979; Wilson and Sutton, 1981). Finally, shifts in posture, which presumably increase middle-ear stiffness via changes in intracochlear pressure, likewise result in a pattern of frequency and amplitude shifts consistent with those elic-

0001-4966/2009/125(5)/3166/11/$25.00 © 2009 Acoustical Society of America

ited by the more direct methods of increasing middle-ear stiffness (Bell, 1992; Whitehead, 1988; De Kleine *et al.*, 2000).

Activation of the cochlear efferent system, for example, with contralateral acoustic stimulation, also can produce small positive shifts in SOAE frequencies, less than 0.3%, and small decreases in amplitude (Mott *et al.*, 1989; Long, 1989). These effects, however, are very difficult to separate from middle-ear reflex effects (Harrison and Burns, 1993).

Finally, there is a small frequency modulation of SOAEs, about 0.1% that correlates with heart rate (Long and Talmadge, 1997). Because most SOAE measurements are analyzed with time-frequency distributions (TFDs) having relatively high-frequency-resolution and low-time-resolution, e.g., spectrally-averaged Fourier transforms, this modulation is usually manifest as low-level sidebands around the SOAE.

The above-cited data are representative of the majority of SOAEs. However, there are many SOAEs that are much less stable both in the short term and in the ultra-short term. In the course of our studies, the author and colleagues have measured a large number of SOAEs whose ultra-short-term frequency variability was more than an order of magnitude greater than that measured by Van Dijk and Wit (1990). Most of these were low-level SOAEs that presumably were perturbed by noise (Talmadge *et al.*, 1993); however, we also have measured a number of high-level frequency-unstable SOAEs. In addition, we have measured a number of SOAEs whose frequencies were stable, but whose amplitude varied in an "on-off" manner over a time frame of tenths-of-seconds to tens-of-seconds (Burns and Keefe, 1992; Burns, 1996).

While the instabilities in individual SOAEs described above are apparently independent of instabilities in other SOAEs in the same ear, another type of SOAE instability that has been documented is the "energy-sharing" of SOAEs where pairs, or groups, of SOAEs co-vary in a highly correlated manner. There are at least three variations of this type of instability. Noncontiguous-linked SOAEs are groups of SOAEs, usually separated by stable SOAEs, whose levels co-vary in an on-off manner; i.e., when one group is present the other is absent (Burns *et al.*, 1984; Whitehead, 1988). The spacing between the frequencies of these SOAEs is consistent with the characteristic minimum spacing of SOAE frequencies–roughly 6%–observed in literature and predicted by the standing-wave models. In some cases the on-off switching from one group of SOAEs to another (i.e., between "states") can be elicited by changing middle-ear pressure (Whitehead, 1988), or by the presentation of an external tone (Burns *et al.*, 1984). The change from state to state of these unstable SOAEs often results in small, less than 1%, shifts in the frequencies of adjacent stable SOAEs. The time course of spontaneous state shifting can vary from seconds to months.

Another variation of the energy-sharing instability are contiguous-linked SOAEs. In this case the amplitudes of pairs of contiguous SOAEs, whose frequency spacing is again consistent with the characteristic minimum spacing of SOAEs, co-vary inversely. Analyses of this type of instability by short-time Fourier transform (STFT) and other TFDs having high-time-resolution show that this co-variation is typically not an on-off co-variation; rather, both SOAEs are usually present simultaneously, but with one or the other of them at a much lower level (Burns and Pitton, 1993).

The third variation of this energy-sharing, the bimodal energy-sharing SOAE, is a SOAE that shifts between two frequencies whose spacing is usually smaller than the characteristic minimum spacing between adjacent SOAEs (Whitehead, 1988; Burns and Keefe, 1992). Analyses of these SOAEs with TFDs having high-time-resolution show that the switching between frequencies is in an on-off manner and is very rapid, often occurring over times that are shorter than the time-resolution of the TFDs (Burns and Keefe, 1992; Burns and Pitton, 1993). Presumably this instability reflects a single SOAE that has a bimodal frequency distribution. For this instability the time course of the shifting varies from tenths-of-seconds to months.

There are relatively few data on the long-term (over years) frequency and amplitude stability of SOAEs. Kohler and Fritze (1992) reported on measurements of 13 SOAEs in 7 ears over a single measurement interval of from 3.5 to 7.2 years. All SOAEs showed a decrease in frequency that averaged 0.14%/year across all emissions.[1] Burns *et al.* (1993a) measured 54 emissions from 12 subjects for periods up to over 9 years and also noted a decline in frequency of all emissions, which averaged about 0.2%/year. Although there is an obvious trend in these data for a long-term decrease in SOAE frequency with age, the small magnitude of this apparent shift, relative to the magnitude of the short-term shifts in SOAE frequency from other factors noted above, requires that longitudinal measurements of SOAE frequencies be carried out over a very long period.

The author and three of his children are "super emitters" who had participated in the study of Burns *et al.* (1993a) on SOAE stability.[2] Having easy additional access to these ears provided the author with a core group of subjects well suited for a continuation and expansion of the earlier study. In the end, repeated measurements were available for some subjects for over 19 years.

II. METHODS

A. Long-term longitudinal measurements

The subjects for the long-term longitudinal study were 6 children (initial age range 6–12 years) and 12 adults (initial age range 21–41 years). Details concerning length of measurement period, number of measurements, initial age, etc., are given in Table I. Subjects 3, 4, 5, and 6 are the author's children and subject 18 is the author. The other adult subjects had been measured initially while graduate students in the Speech and Hearing Sciences Department at the University of Washington and had either remained in the Seattle area or had made the mistake of revisiting the department. At the time of the initial measurements, all subjects had normal tympanograms and thresholds within normal limits at all audiometric frequencies, with the exception of the 41-year-old author who had a permanent threshold shift of greater than 40 dB at 8000 Hz in both ears. The minimum measurement period for inclusion in the study was 4 years.

TABLE I. Individual subject information and average SOAE changes. The numbers in parentheses are standard deviations.

Subject	Initial age	Sex	No. of SOAEs	Years measured	Measurements	Average frequency shift (%/year)	Average level shift (dB/year)
1	6	F	3	4.0	4	−0.27 (0.07)	−0.6 (0.5)
2	6	F	19	4.1	5	−0.26 (0.09)	0.4 (2.3)
3	7	F	6	19.3	16	−0.19 (0.04)	−0.7 (0.3)
4	9	F	11	16.3	15	−0.23 (0.03)	−0.5 (5.8)
5	9	M	2	13.5	11	−0.19	−0.5
6	12	F	9	19.5	15	−0.35 (0.06)	−0.6 (5.0)
7	23	F	3	14.1	3	−0.19 (0.07)	0.0 (0.4)
8	27	F	4	7.1	3	−0.29 (0.07)	−0.6 (1.4)
9	28	F	2	13.8	3	−0.16	−0.9
10	30	F	6	6.4	8	−0.13 (0.07)	−1.1 (9.3)
11	30	F	2	5.1	2	−0.27	−0.1
12	31	F	10	5.0	2	−0.21 (0.06)	−1.5 (1.2)
13	32	F	2	5.0	2	−0.13	−0.3
14	36	F	3	5.1	2	−0.41 (0.01)	−0.2 (0.6)
15	36	M	1	4.1	2	−0.17	−2.1
16	37	M	1	6.8	2	−0.41	−0.8
17	38	M	1	15.0	5	−0.21	−0.5
18	41	M	11	19.3	47	−0.28 (0.07)	−0.3 (0.5)

SOAE measurements for the subjects studied the longest, those measured over periods of greater than 16 years, commenced in mid-1983 to early 1984. Measurements from 1983 to April 1986 were made with a custom-built insert microphone that utilized a Knowles model 1842 microphone and a Grason–Stadler oto-admittance meter earpiece fitted with an appropriately sized rubber eartip. From April 1986 to April 1988, measurements were obtained with an Etymotic model ER-10 insert microphone using foam eartips. From April 1988 until the end of the study in 2003 measurements were made with an Etymotic model ER-10B microphone, which also used the eartips of an oto-admittance meter. All microphones were calibrated in a Zwislocki coupler mounted in a KEMAR mannequin. Measurements were obtained with the subjects sitting quietly in a sound-treated chamber.[3] Tympanograms were measured at the initial session and at any session where SOAE measurements indicated that there might be a middle-ear problem (e.g., no SOAEs, or very low levels, in an ear which had previously had SOAEs), but were not routinely measured at every session.

Microphone signals were high-pass filtered (c/o 300 Hz, roll-off 6 dB/octave). From 1983 until 1991, the signals were analyzed online with a Wavetek-Rockland 5820A spectrum analyzer. Each measurement was typically based on 64 spectral averages. The frequency region from 0 to 10 kHz was examined. Frequency measurements were made with an analysis binwidth of 1.25 Hz; however, an "improved accuracy" option of the analyzer, utilizing weighted averaging of the analysis bins adjacent to that containing the maximum energy, gave a nominal frequency-measurement resolution of 0.125 Hz. Level measurements were made with an analysis binwidth of 12.5 Hz to account, to some degree, for the wider apparent bandwidth of the less-frequency-stable SOAEs.

From 1991 until the end of the study, microphone signals were digitally recorded (44.1 kHz sampling rate) and analyzed offline. Typically, recordings of 20-s duration were analyzed via discrete Fourier transform (DFT) with spectral averaging with a 2:1 overlap factor. Frequency measurements were based on spectra obtained by applying a 262144-sample Hanning window, which gave an analysis binwidth of 0.17 Hz. Level measurements were obtained using a 4096-sample Hanning window, which gave an analysis binwidth of 10.8 Hz..

For the pre-1991 (online) measurements the existence criteria for SOAEs were based on measurement repeatability and comparison with calibration data. For the offline measurements SOAEs are clearly discriminable from electrical pure-tone artifacts based on their bandwidths in the high-frequency-resolution mode; that is, even the most frequency-stable SOAEs have a wider half-power bandwidth than a pure-tone electrical signal, and SOAEs also have a pronounced broadening near the noise floor (Long and Talmadge, 1997).

The SOAEs followed were selected to some extent. For example, super-emitter subjects can have many borderline, low-level, highly frequency-unstable SOAEs, which often disappear from session to session, and whose frequency is difficult to characterize. These weak SOAEs were not followed longitudinally; however, some initially frequency-stable SOAEs that later became unstable were followed as possible. Additionally, in some cases, SOAEs that later became stable were not present in the initial measurement sessions, but appeared in later years. Some of the latter SOAEs, in particular, those which comprised one member of an energy-sharing pair, were followed but are not included in the averaged data.

B. Threshold fine structure measurements

The correlation between SOAE frequencies and minima in the microstructure of the auditory threshold is well documented and is addressed in the current models of SOAE generation (e.g., Schloth, 1983; Long and Tubis, 1988; Talmadge *et al.*, 1998). Specifically, minima in threshold fine structure correspond to maxima in stimulus-frequency otoacoustic emission (SFOAE) fine structure, which in turn correspond to frequencies at which SOAEs may be present. That is, SOAEs always correspond to a threshold minimum, but not all threshold minima have a SOAE associated with them. For subject 18, who was followed over a period of 19.5 years, threshold microstructure was measured in both ears at the time of initial measurement (1983), 11 (1994), and 19 years (2003). Thresholds were obtained for pure tones at frequencies separated by 10 Hz intervals using a computerized Bekesey-tracking procedure. 500-ms duration tones were presented at the rate of 1/s. Consecutive tones increased or decreased in level by a fixed decibel increment. The direction of level change was controlled by the subject. Thresholds were based on the last six of ten turn-arounds.

C. Longitudinal measurements during pregnancy

One explanation posited for the monthly variation of SOAE frequency in females is based on a direct effect of the hormones estrogen and/or progesterone on the generation mechanism of SOAEs (Haggerty *et al.*, 1993; Penner, 1995). Because the levels of these hormones vary much more dramatically over the course of a pregnancy than during a normal menstrual cycle, variation in SOAE frequencies would presumably also vary more dramatically during pregnancy. SOAEs in two females were measured every few weeks over the course of their pregnancies. Subject P1 was followed from 28 week prepartum to 48 week postpartum. Subject P2 was followed from 38 week prepartum to 58 week postpartum. For subject P1, tympanograms were also measured at each session. These two subjects were not part of the main study.

III. RESULTS

A. Longitudinal frequency shifts

96 SOAEs from the 18 subjects were measured. All 96 SOAEs showed decreases in frequency over their respective measurement periods. Figure 1 shows the relative changes in SOAE frequencies over a period of 19.5 years for subject 6, who was 12 years old when first measured. The initial frequency and the ear of each SOAE are given in the figure legend. These results are representative in that all SOAEs showed a gradual decrease in frequency over time; these decreases were approximately linear in percent-frequency-shift-per-year, and there was a tendency for lower-frequency SOAEs to show a greater shift.

For the 77 SOAEs that were measured at least three times, the plots of frequency shift by time were fitted by linear regression. Of these, the 45 SOAEs that were measured for periods of at least 12 years all had coefficients of determination greater than 0.85. The slopes of the linear fits for all SOAEs are shown in Table I as an average slope for each subject. For subjects whose SOAEs were measured only twice, the slopes were calculated from the two measurements. The average slope across all SOAEs was −0.25%/year. The range of slopes for all SOAEs was from −0.033%/year to −0.539%/year; the range for SOAEs measured for at least 12 years was from −0.132%/year to −0.440%/year. Lower initial frequencies tended to show larger negative slopes.

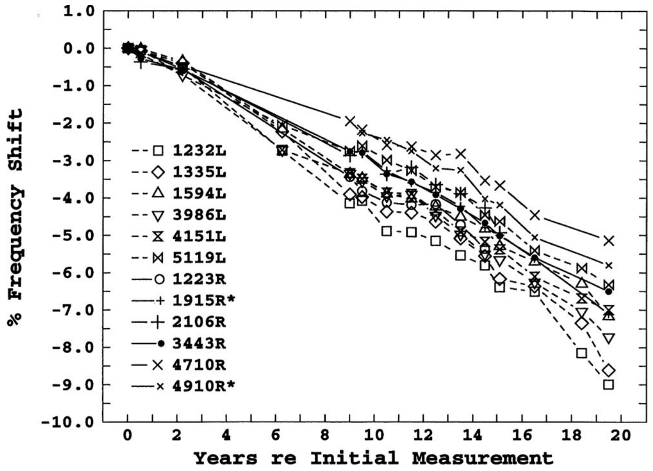


FIG. 1. Longitudinal frequency shifts for 11 SOAEs in subject 6. The * indicates the member of a contiguous-linked SOAE pair not present in the initial years of the study.

A linear regression of frequency-shift-slope on initial frequency, initial level, initial age, duration of measurement, and ear showed a small but significant effect of initial frequency ($\beta = 0.015\%/\text{year kHz}$; $r = 0.242$, $p = 0.017$). No other predictors were significant.

Some other aspects of the results shown in Fig. 1 are noteworthy. This subject was one of the subjects with noncontiguous-linked unstable SOAEs studied by Burns *et al.* (1984), and the initial longitudinal measurements were taken at about the same time as the measurements analyzed in that paper. Those data were later reanalyzed by Keefe *et al.* (1990) who concluded that the SOAEs that comprised the noncontiguous-linked SOAEs could be characterized as high harmonics of a common low fundamental. The SOAEs with frequencies of 1232, 1335, and 1594 Hz shown in Fig. 1 are three of those SOAEs. If they remained harmonics of a common fundamental they would, of course, be expected to show exactly the same percentage frequency shifts over time. This was the case for the first 2 years and for the first 6 years for two of the SOAEs, but after that they all showed significantly different frequency shifts. The instability, a spontaneous "state switching" between two groups of SOAEs, was not seen in any measurements following the 2-year measurements, and state switching could not be induced with the presentation of external tones, as it had been in the original study. It also should be noted that this ear had both the largest average SOAE shift (−0.385%/year) of any ear studied, as well as the largest shift for an individual SOAE measured more than 12 years (−0.440%/year).

The frequency shifts for the right ear of subject 6 show the same form as the left-ear data, but the average frequency shift across SOAEs (−0.2%/year) is smaller. Although frequency shifts in the two ears were not significantly different across subjects, in this subject the difference is highly significant ($t = 7.12$, $p < 0.001$), despite a similar range of SOAE

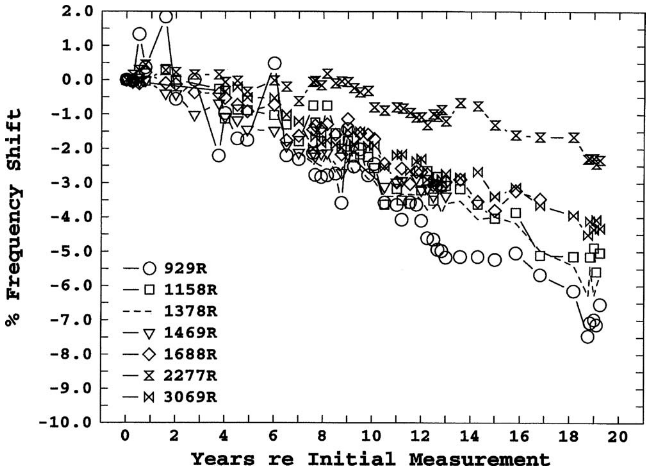


FIG. 2. Longitudinal frequency shifts for six SOAEs in the right ear of subject 18. The dashed line indicates the member of a bimodal-SOAE pair not present in the initial years of the study.

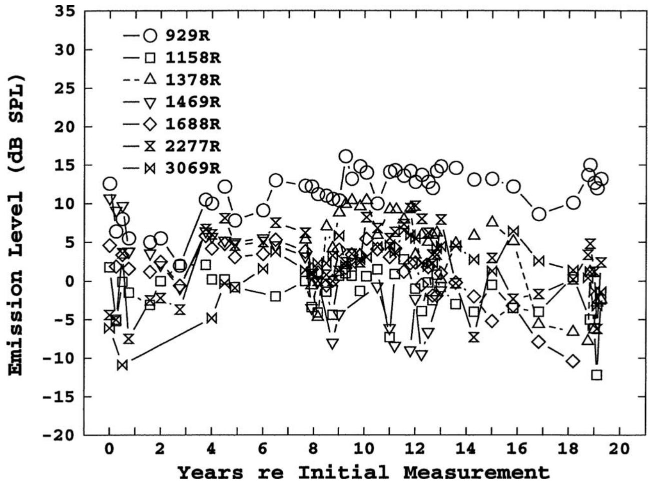


FIG. 3. Longitudinal level shifts for 11 SOAEs in subject 18.

frequencies. The two SOAE frequencies, which are followed by *s, 1915, and 4910, are examples of the contiguous-linked instability described in the Introduction. These SOAEs both appeared at about the eighth year of measurement (age 19): Their frequency shifts at that point were arbitrarily plotted as the same as their energy-sharing neighbors, 2106 and 4710, respectively. Both the 4710 and 4910 SOAEs were simultaneously present in measurements until the end of the study, whereas the 2106 SOAE was eventually replaced by the 1915 SOAE at about 15 years (age 27). The two emissions comprising each pair show similar frequency-shift slopes, and, in fact, the 4710 and 4910 frequency shifts show a similar fine structure in these plots.

Figure 2 shows the frequency shifts for the right ear of subject 18. Several aspects of these data are of interest. First, this ear exhibits two of the instabilities we have characterized as bimodal: SOAEs that are separated in frequency by less than the characteristic minimum SOAE spacing and whose amplitudes co-vary in an on-off manner. The irregular fine structure of the shifts in the 929-Hz SOAE illustrates this phenomenon. The initial SOAE frequencies for this subject (at age 41) are based on an average of 57 measurements per SOAE taken over a period of 1 month. Most of the frequency measurements showed an approximately normal distribution, with an average standard deviation across SOAEs of 0.18%. The 929-Hz SOAE, however, showed a bimodal distribution with modes at 929 and 941 Hz. The 929-Hz mode was chosen as the reference because that was the predominate mode, but the obvious irregularity of the fine structure of the frequency-shift plot during the first 10 years reflects the fact that for some of the longitudinal measurements, the SOAE was in the higher-frequency mode.

Another example of this instability is the 1378 and 1469 Hz SOAE pair. For the first 6 years only the (nominal) 1469 Hz SOAE was present.[4] For the period from 7 to 13 years, sometimes only one or the other was present, and sometimes both were present during a single measurement, from 14 to 19 years only the (nominal) 1378 Hz SOAE was present. Further analyses of the recordings from measurement sessions where both SOAEs were present, using TFDs having higher-time-resolution than our standard DFT analysis, showed that the two SOAEs were not present simultaneously. Rather, there was an on-off switching between the two frequencies, usually with an on-time on the order of 1 s. The subject heard this switching as a "trill-like" tinnitus, a rapid alteration between two pitches a flat semitone apart, that has been described in another paper (Burns and Keefe, 1992). This tinnitus allowed an estimate of the effective level of the SOAEs by loudness matching (Burns, 1996).[5] Although it is virtually impossible to see in Fig. 2 because of the scales used, the slopes and fine structure of the frequency shifts of these two SOAEs are nearly identical.

Finally, there is the anomalous behavior of the 2277 Hz SOAE. This SOAE did not show any significant frequency shift for the first 9 years, the longest shift-free period of any SOAE measured. It also showed the shallowest slope (−0.132%/year) of any of the emissions measured for at least 12 years.

B. Longitudinal level shifts

Longitudinal shifts in level were much more variable than shifts in frequency. A representative example is shown in Fig. 3, the level shifts for the right ear of subject 18. Although, overall, there was a general decrease in SOAE levels, individual SOAEs might show a decrease, an increase, or no change. However, among the 45 SOAEs measured for longer than 12 years, 39 SOAEs showed a decrease in level and only 6 showed either an increase in level or no change. As noted, it was also common for individual SOAEs to disappear, or for new ones to appear. This was particularly true in subjects with numerous SOAEs in one ear.

Simple curves could not be fitted to most of the plots of shifts in level. Therefore the estimates of level-shifts-per-year shown in Table I were simply calculated from the differences in levels between the initial and final levels. These average changes are somewhat biased by the fact that they include the large decreases in levels for SOAEs that eventually disappeared, but do not include SOAEs that were not initially present but appeared later in the longitudinal measurements. Generally, SOAEs having higher initial levels showed greater level shifts.

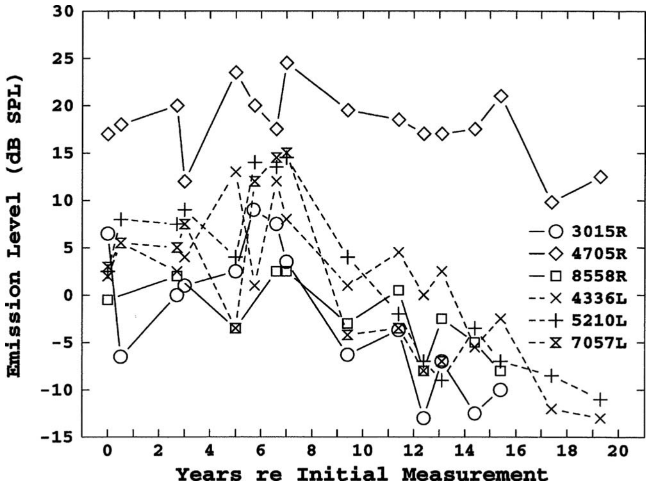


FIG. 4. Longitudinal level shifts for six SOAEs in subject 3.

A linear regression of level-shift-slope on initial frequency, initial level, initial age, and duration of measurement showed a highly significant effect of initial level ($\beta = -0.063$ dB/year dB SPL; $r = -0.401$, $p = 0.00$). No other predictors were significant.

The results shown in Fig. 3 were representative of the vast majority of subjects with multiple SOAEs. Only one subject with more than four SOAEs showed consistent decreases in the levels of all SOAEs. The results for this subject are shown in Fig. 4. After the seventh measurement year, which corresponds to age 14, all her SOAEs, with the possible exception of the strongest, showed consistent decreases in level, and three eventually disappeared completely.

C. Threshold fine structure

Threshold fine structure measurements for the left ear of subject 18, taken at the time of initial measurements (1983, age 41), at 11 years (1994), and at final measurements (2003), are shown in Fig. 5. Minima in the fine structure shifted down in frequency by an amount consistent with the shifts in SOAE frequencies. For example, the minima at 1640/50 Hz (1983), 1590 Hz (1994), and 1560 Hz (2003), correspond to a SOAE whose frequencies were 1642 Hz in 1983, 1595 Hz in 1994, and 1565 Hz in 2003. As noted in the Introduction, a SOAE is not always associated with a minimum. The minima at 1470 Hz in 1983 and 1430 Hz in 1994 are associated with a SOAE that was at 1465 Hz in 1983 and 1431 Hz in 1994. This SOAE was no longer present after 1995; however, the minimum at 1380/1390 Hz in 2003 presumably corresponds to the same SFOAE maximum, which no longer had a measurable SOAE. Threshold fine structure measurements obtained in the right ear of this subject also were consistent with his SOAE shifts.

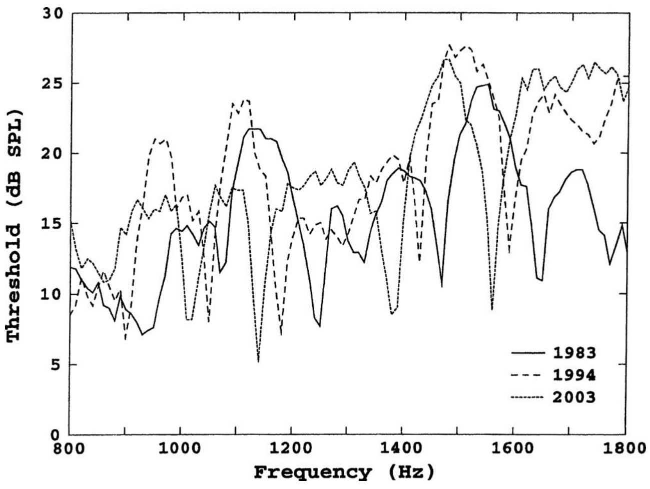


FIG. 5. Behavior threshold fine structure for subject 18, over the range from 800 to 1800 Hz, measured in the initial (1983), 11th (1994), and final (2003) years of study.

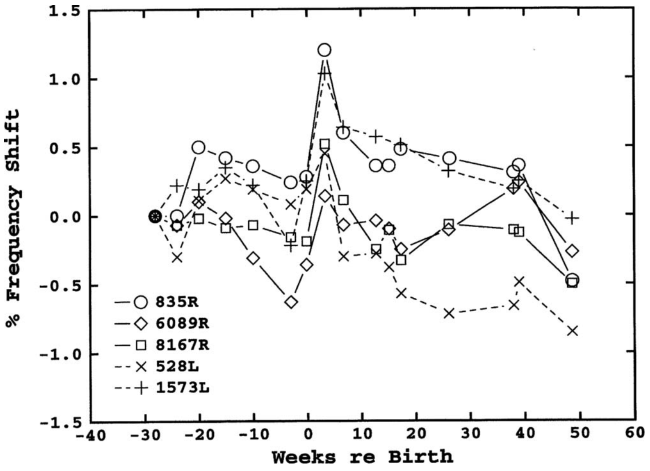


FIG. 6. Longitudinal frequency shifts for five SOAEs in subject P1 during 28 weeks of her pregnancy and for 48 weeks after giving birth.

D. SOAE changes during pregnancy

Figure 6 shows the frequency shifts in the SOAEs of subject P1 for the period from 28 week prepartum to 48 week postpartum. The frequency shifts in subject P2 were roughly similar. The only consistent change in SOAE frequencies across the two subjects was an increase in frequency for all SOAEs from the last prepartum measurement to the first postpartum measurement, which averaged about 0.6% in both subjects. There was no consistent pattern in SOAE level changes over the course of pregnancy in either subject.

IV. DISCUSSION

A. Long-term frequency shifts

Perhaps the most compelling result reported here was the ongoing decrease in SOAE frequencies of about 0.25%/year. This rate of decrease was independent of subject age, i.e., it was essentially the same for subjects with initial ages of 6 years as for subjects with final ages of 60 years. Evidence from a separate longitudinal study of SOAEs in children from ages 1 month to 8 years suggests that this decrease starts shortly after birth (Burns, 1999). Figure 7 shows the frequency shifts of 41 SOAEs from 18 children.[6] Although there is clearly much more variability in the children's data, especially in the first several years, from the age of 6 months

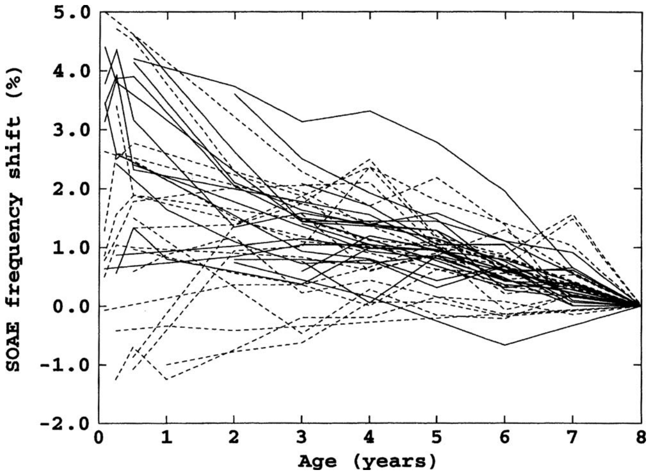


FIG. 7. Longitudinal frequency shifts for 21 SOAEs from 9 female subjects (solid lines) and 20 SOAEs from 9 male subjects (dashed lines) from ages 1 month to 8 years.

most SOAEs decreased in frequency and the average rate of decrease was roughly the same as that of the subjects in the present study.

Interestingly, in preterm infants the opposite occurs. Brienesse *et al.* (1997) found that SOAE frequencies increased in preterm infants. The rate of this increase declined from over 1%/week at 30 week conceptional age to 0%/week at 45–50 week conceptional age. Thus, the overall picture of SOAE frequencies over a lifespan is a rapid increase in frequency in the months just prior to term birth, which changes to a slow decrease starting in the months just after term birth.

The question engendered by the present data is the source of the apparently life-long decrease in SOAE frequencies, which begins at about 6 months of age. As noted, experimental manipulations that increase middle-ear stiffness result in increases in SOAE frequencies, in some cases by as much as 5% (e.g., Kemp, 1979). A relationship between SOAE frequency and middle-ear characteristics is predicted by the standing-wave model because changes in middle-ear stiffness alter the output impedance of the cochlea and thus alter the phase of reflected reverse-traveling waves at the stapes. Shera (2003) specifically addressed the frequency shifts associated with changes in middle-ear stiffness [see Eq. 19 and Fig. 5 in Shera, 2003]: Increases in middle-ear stiffness result in increases in SOAE frequencies, and decreases in middle-ear stiffness result in decreases in SOAE frequencies. Therefore, the slow decrease in SOAE frequencies with age observed in our subjects might be the result of a slow decrease in middle–ear stiffness from ages 6 to 60.

However, there are a number of arguments against this explanation. First, the SOAE frequency shifts produced by changes in middle-ear stiffness are strongly dependent on SOAE frequency. For example, a 100% increase in stiffness results in about a 1.5% increase in frequency for a 1000-Hz SOAE, but only about a 0.25% increase for a 4000-Hz SOAE (Shera, 2003), a sixfold effect. While there was a significant effect of SOAE initial frequency in our results, it was small; the difference in the percentage frequency shift between the lowest and highest SOAE frequencies in a given subject was at most a factor of 2, and was usually less.

Second, the magnitude of the frequency shifts in our subjects could only result from changes in middle-ear stiffness that also would result in low-frequency threshold shifts. For example, shifts in SOAE frequency greater than 2% are seen only for ear-canal pressure changes of greater than ±200 kPa (Kemp, 1979; Wilson and Sutton, 1981). Our subjects showed total frequency shifts of up to 9% with no apparent effects on hearing thresholds.[7]

Third, the rate of frequency shifts in our subjects was relatively uniform over ages ranges (6–12) where metrics of middle-ear function still show maturation effects (e.g., Okabe *et al.*, 1988). Frequency-shift rate was also uniform through adulthood, where no changes in middle-ear function have been reported, and continued to be uniform in the oldest subject who was age 60 at the end of the study. Feeney and Sanford (2004) reported aging effects on middle-ear function in a subject group including 60 year olds. These non-uniform changes in middle-ear function from ages 6 to 60 suggest it is unlikely that changes in middle-ear stiffness could account for the uniform frequency shifts over this age range.

The major determinate of SOAE frequency is the place-frequency map of the BM, which is, in turn, primarily a function of the exponentially varying stiffness of the BM from base to apex. A continual decrease in BM stiffness with age could account for the decline in SOAE frequency with age. According to current models (e.g., Talmadge *et al.*, 1998) there are two components to BM stiffness, the passive stiffness and an “active” component provided by the so-called cochlear amplifier. The passive component is mainly determined by the transverse fiber bands of the BM (e.g., Olson and Mountain, 1994). There are no reports in literature from which to assess any possible morphological changes in these bands in the mammalian BM over age ranges equivalent to those covered in our study, nor are there any reports on measurements of passive BM stiffness over this age range.

There are indirect measurements in humans, which could be interpreted as reflecting a change in BM stiffness with age. Ramotowski and Kimberley (1998) measured human BM traveling-wave delay in 91 subjects from 22 to 78 years of age and found a significant increase in delay as a function of age. This increase in delay, which is independent of hearing threshold differences, and is the opposite of what would be expected from an increase in tuning bandwidth with age, could be the result of a decrease in BM traveling-wave velocity due to a decrease in BM stiffness. Among the many caveats in comparing these data to our frequency-shift data are that the rate of increase in traveling-wave delay appears to become larger with age, which is not the case in the SOAE frequency-shift data, and there are no data on the age range from infancy to young adult. Also note that if these data do reflect decrease in BM stiffness, it could be due to changes in either the passive or the active component.

Because a component of stiffness is provided by the cochlear amplifier, loss of efficiency of the amplifier would presumably result in a reduction in stiffness and a concomitant decrease in SOAE frequencies. All types of otoacoustic emissions (OAEs) are manifestations of the existence of the cochlear amplifier, and their levels are assumed to be an indirect reflection of cochlear-amplifier efficiency. Therefore, studies of OAE levels with age should provide evidence regarding a possible decrease in BM stiffness due to a decrease in efficiency of the cochlear amplifier with age. However, the conductive mechanism also must be taken into account. For example, the well-documented high levels of OAEs in infants relative to adults are probably entirely attributable to immaturities in the ear canal and middle ear (Abdala and Keefe, 2006; Keefe and Abdala, 2007). A study comparing SOAE power (Burns and Keefe, 1997), a measurement which accounts for differences in probe impedance and placement as well as ear-canal size, showed no differences in the average power of SOAEs between 8-year-old children and adults.[8] The results of the present study showed a general decrease in SOAE levels with age, but not a consistent and uniform decrease in all SOAE levels as was the case for SOAE frequency. Studies on evoked OAE levels as a function of age in adults show little consensus in the results (e.g., Cilento *et al.*, 2003). Some studies show a decrease in OAE levels with age, which are independent of changes in hearing threshold, and others do not. In most cases where threshold-independent decreases in OAE levels are seen, high frequencies show significantly greater decreases than low frequencies (Dorn *et al.*, 1998).

Finally, the place-frequency map also could be shifted to lower frequencies by a uniform addition of BM mass (per unit length). For example, Long and Talmadge (1997) concluded, on the basis of modeling results, that the most likely explanation for the small modulations in the frequency of SOAEs that are correlated with heartbeat is the small increase in mass of the BM that occurs with the increase in blood flow with each heartbeat. As with BM stiffness, there are no measurements of the morphology of mammalian BMs over the age range equivalent to that in our study that might provide correlative evidence for a continuous increase in BM mass with age.

B. Long-term level shifts

In contrast with the totally consistent findings on frequency shifts with age, the findings for level shifts were much less consistent, both within and across subjects. Although there was an overall trend for SOAE levels to decrease with age, in most subjects individual SOAE levels might decrease, increase, or stay the same. A portion of this inconsistency is obviously related to the inherent variability of SOAE level measurements, both within and across sessions, relative to frequency measurements. SOAE levels are much more sensitive to external- and middle-ear acoustics: Probe calibration, probe placement, cerumen in the ear canal, changes in ear-canal size, and changes in middle-ear function all can have significant affects on SOAE levels. Large changes in the frequency stability of SOAEs also would affect their level measurements, given the constant binwidth of the measurements (12.5 Hz, 1983–1990; 10.8 Hz, 1991–2003).

Another factor is the interdependency often seen in the levels of SOAEs in ears with numerous SOAEs. As discussed above, often the decrease in level (or disappearance) of a particular SOAE is accompanied by the increase in level (or appearance) of a neighboring SOAE. Because we did not include SOAEs that appeared after the initial measurement session in the data analysis, this “energy conservation” aspect of SOAE levels was not totally taken into account. The significant correlation between initial level and level shift—higher initial level SOAEs showed greater level shifts—presumably reflects the obvious: Higher level SOAEs have more level to lose before disappearing below the noise floor.

Although in most subjects there was no evidence of an overall decline in SOAE levels, a few subjects did show a consistent decline in level with age. An example of this was shown in Fig. 4. This subject was first measured at age 6. Her SOAE levels were stable up to age 14, but all SOAEs in both ears showed a decrease in level from age 14 until the final measurement at age 26. The subject’s thresholds and tympanograms were normal at age 26, and her transient-evoked OAE levels were in the high end of the normal range. The frequency shifts of her SOAEs were consistent with those of the other subjects, and showed no obvious changes in slope between the periods, which corresponded to stable and decreasing SOAE levels. In short, there was nothing exceptional about this subject other than the uniform decrease in SOAE levels.

The obvious explanation for a decrease in SOAE levels with age would be a decrease in the efficiency of the cochlear amplifier with age. As discussed in Sec. IV A, the evidence for such a decrease is equivocal.

C. Long-term stability of frequency/level instabilities

The disappearance or appearance of individual SOAEs, both in the short term and the long term, is consistent with the standing-wave model because of the stochastic nature of the impedance irregularities that are the basis of the reflected waves, and the spatial variations in the nonlinear amplification necessary to maintain the waves (Talmadge *et al.*, 1998; Shera, 2003). That is, the model predicts the nominal frequencies at which SOAEs can occur, but whether a SOAE will be present at one of these frequencies at any particular time, and its level, depends on factors that can vary both in the short and long term. However, the energy-sharing instabilities noted in Sec. III A, i.e., noncontiguous-linked SOAEs, contiguous-linked SOAEs, and bimodal SOAEs, presumably reflect somewhat more complicated dynamics, perhaps interacting standing waves between tonotopic reflection sites. The fact that some of these instabilities are themselves quite stable over years suggests that at least a portion of the basis for these instabilities may be robust irregularities in BM morphology, for example, extra rows of outer hair cells at a particular location (Lonsbury-Martin *et al.*, 1988).

D. Threshold fine structure

The fact that the threshold fine structure of subject 18 shifted in frequency along with the shifts in his SOAE frequencies is completely consistent with both the empirical and modeling results (Talmadge *et al.*, 1998) on the relationship between SOAEs and threshold fine structure.

E. SOAE changes during pregnancy

McFadden (2008, 2009) presented strong evidence that prepartum exposure to male gonadal hormones directly affects the cochlear processes that produce SOAEs; for example, the exposure to high levels of male hormones in the womb "masculinizes" the ears of female opposite-sex-dizygotic twins such that they tend to have fewer and weaker SOAEs. There also are a number of auditory measurements, both physiological and psychophysical, which fluctuate during the menstrual cycle [summarized by McFadden (1998)]. It is therefore not unreasonable to suggest, as do Haggerty *et al.* (1993) and Penner (1995), that the monthly variations in frequency of female SOAEs could be directly related to fluctuating levels of the female gonadal hormones estrogen and progesterone over the menstrual cycle.[9]

The results of our measurements on two subjects during pregnancy do not lend much support to this idea, however. The only consistent frequency shift in the SOAEs in the two subjects was a small (about 0.6%) increase in frequency between the last prepartum measurement and the first postpartum measurement. This is within the range of shifts seen over the menstrual cycle for females; whereas estrogen and progesterone levels in pregnancy, and, in particular, from late postpartum to just prepartum, vary by almost two orders of magnitude more than they do during a menstrual cycle (Tulchincky *et al.*, 1972). In addition, both during a normal menstrual cycle and during pregnancy there are other changes occurring such as fluctuating levels of interstitial fluids that could affect middle-ear function (e.g., Cox, 1980), which as noted above, profoundly affects SOAE frequencies. In the one pregnancy subject where tympanograms were obtained at each measurement session, the pre-postpartum increase in frequency correlated with a decrease in peak compliance of about 0.1 ml. Also, as McFadden (1998) noted, most of the auditory measures that fluctuate over the menstrual cycle are either definitely, or most likely, mediated at post-cochlear levels of the auditory system. It thus seems more likely that the small frequency shifts seen in SOAEs in females over the menstrual cycle are related to middle ear-effects rather than levels of female hormones per se.

F. Relevance of SOAE frequency shifts to pitch coding

The following second-order correlation is presented for consideration by those readers interested in the apparently never-ending saga of whether pure-tone pitch is coded primarily by place or temporal information. The average SOAE frequency shift of 0.25%/year leads to a semitone shift in frequency after 24 years. Most possessors of absolute pitch (AP) report that their AP shifts by a semitone about every 20 years (Ward, 1999). The direction of pitch shift is, assuming place coding of pitch, consistent with a decrease in the frequency corresponding to a particular place along the BM as the result of, for example, a decrease in stiffness of the BM. The existence of a drug, which reversibly shifts AP (Chaloupka *et al.*, 1994), suggests an obvious experiment, using subjects who possess both AP and SOAEs, which would lead either to a first-order correlation between SOAE frequency shifts and shifts in AP, or would illustrate the irrelevance of these shifts to pitch coding.

V. CONCLUSIONS

SOAEs uniformly decrease in frequency at about 0.25%/year from shortly after birth to at least age 60. SOAEs also decrease in level with age, but unlike the decrease in frequency, decreases in level are not uniform or consistent either within or across subjects. Female gonadal hormones probably do not have a significant effect on SOAE frequencies.

ACKNOWLEDGMENTS

This research was supported by long-since-expired NIH Grant No. P01 DC00520. The author would like to thank Doug Keefe, Lynne Werner, and a particularly thorough anonymous reviewer for helpful comments on previous versions of this manuscript.

[1] Kohler and Fritze (1992) gave the average yearly shift as 1.4%. However, a perusal of their data shows that this was an error and the actual value is 0.14%.

[2] Super-emitters are arbitrarily defined as possessing 12 or more SOAEs in at least one ear; possessing at least one SOAE greater than 10 dB SPL; and, at some point, possessing one of the linked instabilities discussed in the Introduction.

[3] Some investigators prefer to base SOAE frequency measurements on the value obtained after the slow drift, which often follows placement of the microphone in the ear, has dissipated. However, this entails having the subjects sit quietly in the booth for up to 30 min, which is obviously not practical for young subjects, and therefore we did not do so. In any case, it is not clear that waiting for this apparently efferent-based shift, which is highly variable across subjects, would lead to less-variable repeated measurements across sessions.

[4] SOAEs in this paper are denoted by their nominal frequencies, i.e., their frequencies when they were first measured. By the time the SOAE at 1378 Hz appeared, in the sixth year of measurements, the 1469 Hz SOAE had declined in frequency to 1440 Hz. At this point, then, the separation between these SOAEs (about 4%) was smaller than the characteristic minimum spacing of about 6% in this frequency range, although still within the range of variability of characteristic minimum spacing (Shera, 2003). In this sense this SOAE pair was atypical for a bimodal instability, where the usual spacing is on the order of 1%–2%.

[5] Stable SOAEs are usually inaudible to their possessor or are very faint, even when measured at relatively high levels in the ear canal. Presumably this is because SOAEs are subject to the same type of adaptation that is measured psychophysically for low-level, pure-tone stimuli, i.e., "loudness adaptation" and "tone decay" (e.g., Scharf, 1983). SOAES that are unstable in level and/or frequency are, to varying degrees, released from this adaptation. For bimodal SOAEs, which apparently switch frequencies and are either "on" or "off" at a particular frequency, and for temporal conditions where the "on-times" of these SOAEs are long enough that loudness integration is complete (about 200 ms) and short enough so there is no loudness adaptation (about 500 ms), the equivalent level of the SOAE can be determined by comparing their level in the ear canal, obtained from STFT analyses, with the levels of tones matched in loudness to the SOAEs. For the 1378/1469 pair, and another bimodal pair in the same subject, the SOAEs were matched in loudness by tones whose levels in the ear canal were from 15 to 20 dB higher than the levels of the SOAEs in the ear canal. The value of 30 dB given in Burns, 1996 illustrates the danger of the practice of submitting abstracts based on preliminary data.

[6] Note that, because some of the SOAEs in infants were not seen in the earliest measurement sessions due to noisy measurement conditions, the reference for frequency shifts in this figure is the final (8-year-old) measurement rather than the initial measurement as in our other figures.

[7] The subjects who showed the largest cumulative frequency shift were those studied the longest, the author and two of his children (subjects 18,

3, and 6). Thresholds measured in these subjects at the end of the study were within normal limits with the exception of the high-frequency (8 kHz) loss in the author, which had also been present at time of initial measurements.

[8] For SOAEs measured in the ears of both adults and children, the power absorbed by the Etymotic ER-10C probe from the majority of SOAEs ranged from 0.1 to 10.0 aW, with the power absorbed from the highest-level SOAEs as large as 400 aW. The power was estimated from the SOAE SPL recorded by the probe microphone, the measured source impedance of the ER-10C probe, and the results calculated using an acoustic transmission-line model of the ear canal between the tympanic membrane and the probe. The model assumed that the ear canal was adequately represented by a finite cylindrical tube with rigid and loss-free walls; these assumptions are thought to be valid for older children and adults. The model was specified in terms of ear-canal cross-sectional area and length. The ear-canal volume was estimated using a tympanometric measurement, assuming the same insertion depth for the tympanometry probe and the ER-10C probe. Any small differences in insertion depth would produce only small errors in the calculated power because of the absence of standing-wave effects in the ear canal in the frequency range of the SOAE measurements. The ear-canal cross-sectional area was estimated acoustically based on reflectance measurements (Keefe and Abdala, 2007), and the ear-canal length was calculated as the ratio of volume to area.

[9] Haggerty *et al.* (1993) also suggested the possibility that the pineal hormone melatonin, the levels of which correlate with the circadian rhythm, might control the menstrual-cycle variations in SOAE frequency. However, more recent results (Parry *et al.*, 2006) indicate that melatonin levels are stable over the menstrual cycle in normal subjects.

Abdala, C., and Keefe, D. H. (**2006**). "Effects of middle-ear immaturity on distortion product otoacoustic emission suppression tuning in infant ears," J. Acoust. Soc. Am. **120**, 3832–3842.

Bell, A. (**1992**). "Circadian and menstrual rhythms in frequency variations of spontaneous otoacoustic emissions from human ears," Hear. Res. **58**, 91–100.

Brienesse, P., Anteunis, L. J. C., Maertzdorf, W., Blanco, C. E., and Manni, J. J. (**1997**). "Frequency shift of individual spontaneous otoacoustic emissions in preterm infants," Pediatr. Res. **42**, 478–483.

Burns, E. M. (**1996**). "Equivalent levels of SOAEs estimated from loudness matches to unstable SOAEs," in *Abstracts of the 19th Midwinter Research Meeting of the ARO*, edited by R. Popelka (Association of Research in Otolaryngology, Des Moines, IA), p. 182.

Burns, E. M. (**1999**). "Longitudinal measurements of SOAEs in children revisited, for the last time, really," in *Abstracts of the 22nd Midwinter Research Meeting of the ARO*, edited by R. Popelka (Association of Research in Otolaryngology, Des Moines, IA), p. 92.

Burns, E. M., Campbell, S. L., Arehart, K. H., and Keefe, D. H. (**1993a**). "Long-term stability of spontaneous otoacoustic emissions," in *Abstracts of the 16th Midwinter Research Meeting of the ARO*, edited by D. Lim (Association of Research in Otolaryngology, Des Moines, IA), p. 98.

Burns, E. M., Harrison, W. A., Bulen, J. C., and Keefe, D. H. (**1993b**). "Voluntary contraction of middle ear muscles: Effects on input impedance, energy reflectance and spontaneous otoacoustic emissions," Hear. Res. **67**, 117–127.

Burns, E. M., and Keefe, D. H. (**1992**). "Intermittent tinnitus resulting from unstable otoacoustic emissions," in *Tinnitus 91: Proceedings of the Fourth International Tinnitus Seminar*, edited by J. M. Aran and R. Dauman (Kugler, Amsterdam).

Burns, E. M., and Keefe, D. H. (**1997**). "SOAEs and power transfer in the middle and external ears of children and adults," in *Abstracts of the 20th Midwinter Research Meeting of the ARO*, edited by R. Popelka (Association of Research in Otolaryngology, Des Moines, IA), p. 168.

Burns, E. M., and Pitton, J. W. (**1993**). "Time-frequency analyses of coherent frequency fluctuations among spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **93**, 2314 (Abstract).

Burns, E. M., Strickland, E. A., Tubis, A., and Jones, K. L. (**1984**). "Interactions among spontaneous otoacoustic emissions. I. Distortion products and linked emissions," Hear. Res. **16**, 271–278.

Chaloupka, V., Mitchell, S., and Muirhead, R. (**1994**). "Observation of a reversible, medication-induced change in pitch perception," J. Acoust. Soc. Am. **96**, 145–149.

Cilento, B. W., Norton, S. J., and Gates, G. A. (**2003**). "The effects of aging and hearing loss on distortion product otoacoustic emissions," Otolaryngol.-Head Neck Surg. **129**, 382–389.

Cox, J. R. (**1980**). "Hormonal influence on auditory function," Hear. Res. **1**, 219–222.

De Kleine, E., van Dijk, P., and Avan, P. (**2000**). "The behavior of spontaneous otoacoustic emissions during and after postural changes," J. Acoust. Soc. Am. **107**, 3308–3316.

Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (**1998**). "On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions," J. Acoust. Soc. Am. **104**, 964–971.

Feeney, P. M., and Sanford, C. A. (**2004**). "Age effects in the human middle ear: Wideband acoustical measures," J. Acoust. Soc. Am. **116**, 3546–3558.

Fritze, W. (**1983**). "Registration of spontaneous cochlear emissions by means of Fourier transformation," Eur. Arch. Otorhinolaryngol. **238**, 189–196.

Haggerty, H. S., Lusted, H. S., and Morton, S. C. (**1993**). "Statistical quantification of 24-hour and monthly variabilities of spontaneous otoacoustic emission frequency in humans," Hear. Res. **70**, 31–49.

Harrison, W. A., and Burns, E. M. (**1993**). "Effects of contralateral acoustic stimulation on spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **94**, 2649–2658.

Hauser, R., Probst, R., and Harris, F. P. (**1993**). "Effects of atmospheric pressure variation on spontaneous, transiently evoked, and distortion product otoacoustic emissions in normal human ears," Hear. Res. **69**, 133–145.

Keefe, D. H., and Abdala, C. (**2007**). "Theory of forward and reverse middle-ear transmission applied to otoacoustic emissions in infant and adult ears," J. Acoust. Soc. Am. **121**, 978–993.

Keefe, D. H., Burns, E. M., Ling, R., and Laden, B. (**1990**). "Chaotic dynamics of otoacoustic emissions," in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. Geisler, J. Mathews, M. Ruggero, and C. Steele (Springer-Verlag, Berlin), pp. 194–201.

Kemp, D. T. (**1979**). "The evoked cochlear mechanical response and the auditory microstructure—Evidence for a new element in cochlear mechanics," Scand. Audiol. Suppl. **9**, 35–47.

Kohler, W., and Fritze, W. (**1992**). "A long-term observation of spontaneous oto-acoustic emissions (SOAEs)," Scand. Audiol. **21**, 55–58.

Long, G. L., and Talmadge, C. L. (**1997**). "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831–2848.

Long, G. R. (**1989**). "Modification of the frequency and level of otoacoustic emissions by contralateral stimulation, in a subject with no acoustic reflexes in one ear," in *Abstracts of the 12th Midwinter Research Meeting of the ARO*, edited D. Lim (Association of Research in Otolaryngology, Columbus, OH), p. 228.

Long, G. R., and Tubis, A. (**1988**). "Investigations into the nature of the association between threshold microstructure and otoacoustic emissions," Hear. Res. **36**, 125–138.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1988**). "Spontaneous otoacoustic emissions in the nonhuman primate. II. Cochlear anatomy," Hear. Res. **33**, 69–94.

McFadden, D. (**1998**). "Sex differences in the auditory system," Dev. Neuropsychol. **14**, 261–298.

McFadden, D. (**2008**). "What do sex, twins, spotted hyenas, ADHD, and sexual orientation have in common?," Perspect. Psychol. Sci. **3**, 309–323.

McFadden, D. (**2009**). "Masculinization of the mammalian cochlea," Hear. Res.

Mott, J. B., Norton, S. J., Neely, S. T., and Warr, W. B. (**1989**). "Changes in otoacoustic emissions produced by acoustic stimulation of the contralateral ear," Hear. Res. **38**, 229–242.

Okabe, K., Tanaka, S., Hamada, H., Miura, T., and Funai, H. (**1988**). "Acoustic impedance measurement of normal ears of children," J. Acoust. Soc. Jpn. **9**, 287–294.

Olson, E. S., and Mountain, D. C. (**1994**). "Mapping the cochlear partition's stiffness to its cellular architecture," J. Acoust. Soc. Am. **95**, 395–400.

Parry, B., Martinez, L. F., Maurer, E., López, A., Sorenson, D., and Meliska, C. (**2006**). "Sleep rhythms and women's mood. Part I: Menstrual cycle, pregnancy and postpartum," Sleep Med. Rev. **10**, 129–144.

Penner, M. J. (**1995**). "Frequency variation of spontaneous otoacoustic emissions during a naturally occurring menstrual cycle, amenorrhea, and oral contraception: A brief report," Ear Hear. **16**, 428–432.

Rabinowitz, W. M., and Widen, G. P. (**1984**). "Interaction of spontaneous oto-acoustic emissions and external sounds," J. Acoust. Soc. Am. **76**, 1713–1720.

Ramotowski, D., and Kimberley, B. (**1998**). "Age and the human cochlear traveling wave delay," Ear Hear. **19**, 111–119.

Scharf, B. (**1983**). "Loudness adaptation," in *Hearing Research and Theory*, edited by J. V. Tobias and E. D. Schubert (Academic, San Diego, CA), Vol. **2**, pp. 1–56.

Schloth, E. (**1983**). "Spectral composition of spontaneous oto-acoustic emissions and fine structure of the threshold in quiet," Acustica **53**, 250–256.

Schloth, E., and Zwicker, E. (**1983**). "Mechanical and acoustical influences on spontaneous otoacoustic emissions," Hear. Res. **11**, 285–293.

Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.

Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). "New off-line method for detecting spontaneous otoacoustic emissions in human subjects," Hear. Res. **71**, 170–182.

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.

Tulchincky, D., Hobel, C. J., Yeager, E., and Marshall, J. R. (**1972**). "Plasma estrone, estraiol, estroil, progesterone, and 17-hydroxprogesterone in human pregnancy," Am. J. Obstet. Gynecol. **112**, 1095–1100.

Van Dijk, P., and Wit, H. P. (**1990**). "Amplitude and frequency fluctuations of spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **88**, 1779–1793.

Ward, W. D. (**1999**). "Absolute pitch," in *The Psychology of Music*, edited by D. Deutsch (Academic, San Diego, CA), pp. 265–298.

Whitehead, M. L. (**1988**). "Some properties of otoacoustic emissions in vertebrate ears, and their relationship to other hearing mechanisms," Ph.D. thesis, University of Keele (Keele, UK).

Whitehead, M. L. (**1991**). "Slow variations of the amplitude and frequency of spontaneous otoacoustic emissions," Hear. Res. **53**, 269–280.

Wilson, J. P. (**1986**). "The influence of temperature on frequency-tuning mechanisms," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, A. Hubbard, S. St. Neely, and A. Tubis (Springer-Verlag, Berlin), pp. 229–236.

Wilson, J. P., and Sutton, G. J. (**1981**). "Acoustic correlates of tonal tinnitus," in *Tinnitus*, edited by D. Evered and G. Lawrenson (Pitman Medical, London), pp. 82–107.

Wit, H. P. (**1985**). "Diurnal cycle for spontaneous oto-acoustic emission frequency," Hear. Res. **18**, 197–199.

Zurek, P. M. (**1981**). "Spontaneous narrowband acoustic signals emitted by human ears," J. Acoust. Soc. Am. **69**, 514–523.

Representation of the vocal roughness of aperiodic speech sounds in the auditory cortex

Santeri Yrttiaho[a)]

Department of Signal Processing and Acoustics, Helsinki University of Technology, P.O. Box 3000, FI-02015 TKK, Finland; Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, P.O. Box 3310, FI-02015 TKK, Finland; and BioMag Laboratory, Hospital District of Helsinki and Uusimaa HUSLAB, Helsinki University Central Hospital, P.O. Box 340, FI-00029 HUS, Finland

Paavo Alku

Department of Signal Processing and Acoustics, Helsinki University of Technology, P.O. Box 3000, FI-02015 TKK, Finland

Patrick J. C. May and Hannu Tiitinen

Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, P.O. Box 3310, FI-02015 TKK, Finland and BioMag Laboratory, Hospital District of Helsinki and Uusimaa HUSLAB, Helsinki University Central Hospital, P.O. Box 340, FI-00029 HUS, Finland

(Received 17 September 2008; revised 4 February 2009; accepted 12 February 2009)

Aperiodicity of speech alters voice quality. The current study investigated the relationship between vowel aperiodicity and human auditory cortical N1m and sustained field (SF) responses with magnetoencephalography. Behavioral estimates of vocal roughness perception were also collected. Stimulus aperiodicity was experimentally varied by increasing vocal jitter with techniques that model the mechanisms of natural speech production. N1m and SF responses for vowels with high vocal jitter were reduced in amplitude as compared to those elicited by vowels of normal vocal periodicity. Behavioral results indicated that the ratings of vocal roughness increased up to the highest jitter values. Based on these findings, the representation of vocal jitter in the auditory cortex is suggested to be formed on the basis of reduced activity in periodicity-sensitive neural populations. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097471]

PACS number(s): 43.64.Qh, 43.64.Sj, 43.64.Ri, 43.71.Gv [WPS] Pages: 3177–3185

I. Introduction

A major acoustic characteristic of a healthy human voice is the periodicity of voiced speech sounds such as vowels and sonorants. The periodicity of voiced speech sounds originates from the periodic vibration of the vocal folds in the larynx and distinguishes them from unvoiced speech sounds (e.g., fricatives) that are produced without vibrating the vocal folds. This natural vibratory pattern of the vocal folds may, however, be disrupted by laryngeal anomalies such as tumor growths. Such conditions are referred to as dysphonia as they involve pathological alterations in the phonation of speech sounds giving rise to voice aperiodicities that can be described with qualities such as vocal roughness. The effects that laryngeal parameters, including vocal fold anomalies, have on speech acoustics (e.g., Lieberman, 1963) and perceived vocal quality (see Kreiman *et al.*, 1993 for a review) have been studied extensively. However, the human auditory physiological processes relevant to perception of phonation, especially in the case of dysphonic voice, are poorly understood. The lack of brain research focusing on the processing of these aspects of vocal quality may partly be attributable to the challenges in creating realistic-sounding and highly-controlled speech stimuli. This major obstacle has, however, been lifted by advances in acoustic modeling of speech production (see Alku *et al.*, 1999) which enable the generation of natural, yet fully controllable speech stimuli for the research of speech perception using behavioral and brain measures. Now the task of the auditory physiological research on dysphonic voice perception rests on finding the crucial relationships between the parameters of voice production and the cerebral sensory processes of voice quality perception.

The periodicity of natural voiced speech is not perfect and is referred to as quasi-periodicity. That is, the fundamental frequency (F0) of voiced speech, which is determined by the period of the glottal flow waveform, varies randomly on a cycle-to-cycle basis. This random perturbation is called vocal jitter, which may be quantified as the average absolute difference between successive periods relative to the fundamental period, and is generally less than 1% for normal voices (e.g., Horii, 1979; Muñoz *et al.*, 2003). In pathological voices, this cycle-to-cycle variation in F0 may, however, be unusually large (Lieberman, 1963; Iwata and von Leden, 1970; Deal and Emanuel, 1978; Murry and Doherty, 1980) and, for example, over ten-fold perturbation magnitudes have been reported for pathologic versus normal voices (Murry and Doherty, 1980).

[a)]Author to whom correspondence should be addressed. Electronic mail: santeri.yrttiaho@tkk.fi

Perceptually, increased jitter is linked to the quality of vocal roughness often associated with a sore or dry throat. Hillenbrand (1988) presented listeners with synthetic speech sounds with varying percentages of vocal jitter. The synthetic stimuli enabled an investigation of the perception of vocal jitter free from the confounding effects of other sources of aperiodicity, such as shimmer and additive noise which naturally co-occur in dysphonic voices. The results showed a monotonic increase in ratings of perceived roughness as the jitter was increased. Thus, vocal jitter represents a perceptually relevant form of aperiodicity that is strongly related to voice quality. Despite the development of suitable brain research techniques, human auditory physiological correlates of voice quality in general and vocal jitter, in particular, have not yet been investigated.

To date, one of the most viable techniques for studying cortical processing of speech sounds is multi-channel magnetoencephalography (MEG) which allows non-invasive measurements of cortically-induced magnetic fields [see Hämäläinen *et al.* (1993) for a review]. With a temporal resolution in the order of milliseconds, MEG captures detailed information about brain activation elicited by auditory stimuli. Among the most prominent auditory cortical responses are the N1m response that peaks around 100 ms after sound onset (Näätänen and Picton, 1987) and the sustained field (SF) which builds in amplitude for up to 400 ms after stimulus onset during continuous auditory stimulation (Hari *et al.*, 1980; Pantev *et al.*, 1994; see also Picton *et al.*, 1978a, 1978b). An important characteristic of these two responses is their sensitivity to the acoustic features of sound stimuli in terms of response amplitude, latency, and source location which may be estimated with techniques such as equivalent current dipole (ECD) modeling. Therefore, N1m and SF responses provide useful indices for studying the auditory cortical dynamics of periodicity and speech processing.

Auditory MEG studies that have systematically varied the degree of stimulus periodicity (Krumbholz *et al.*, 2003; Soeta *et al.*, 2005; Gutschalk *et al.*, 2007) may, potentially, be relevant for uncovering the brain dynamics involved in the processing of aperiodicities that result from abnormal phonation. Krumbholz *et al.* (2003) and Soeta *et al.* (2005) presented subjects with iterated rippled noise (IRN) stimuli, which were produced by delaying a noise, adding it back to the original, and iterating this delay-and-add process. Their results indicated a positive correlation between the amplitude of cortical transient responses and the degree of periodicity of the IRN stimuli. Similarly, Gutschalk *et al.* (2007) found a positive correlation between the degree of periodicity of click train stimuli and cortical transient and sustained responses. However, as only non-speech stimuli were used in these studies, the results might not be generalizable to the processing of human voice. Therefore, experiments with realistic speech stimuli with plausible amounts of F0 perturbation are needed in order to investigate the cortical basis of the perception of vocal jitter-induced roughness.

Initial MEG studies investigating the cortical sensitivity to speech sound periodicity were carried out by Alku *et al.* (2001), Tiitinen *et al.* (2005), and Yrttiaho *et al.* (2008). Using semi-synthetic speech generation (SSG) (Alku *et al.*, 1999), which is based on the source-filter model of speech production, it is possible to manipulate the glottal excitation source of the speech sounds independent of the filtering effect of the vocal tract transfer function. The above-mentioned studies applied these techniques of speech generation in creating fully-controlled periodic and aperiodic speech sounds. Aperiodic reproductions of naturally-uttered vowels were produced by replacing the periodic glottal source with a random noise excitation while preserving the features of the spectrum envelope. It was found that the cortical N1m (Alku *et al.*, 2001; Tiitinen *et al.*, 2005; Yrttiaho *et al.*, 2008) and SF (Yrttiaho *et al.*, 2008) responses elicited by vowel stimuli were larger in amplitude for the periodic vowels than for their aperiodic counterparts. Similar results were also found by Hertrich *et al.* (2000), who contrasted N1m responses to periodic and aperiodic artificial speech-like stimuli. These studies, thus, indicate that the human auditory cortex is sensitive to the periodicity of speech sounds. However, the vowel stimuli used in these investigations were categorically either periodic, as in healthy phonation, or aperiodic, as characterized by random noise excitation. No intermediate degrees of speech periodicity were used. Consequently, the question of whether auditory cortical responses are sensitive to more subtle reductions in the degree of speech periodicity that occur, for example, in dysphonic voices is left open.

The aim of the current study was to investigate whether auditory cortical activity reflects the degree of speech periodicity, quantified as vocal jitter, and its perceptual counterpart, the vocal quality of roughness. To this end, vowel stimuli along the periodic-aperiodic continuum were presented to human subjects during non-invasive MEG measurements. The periodicity of the vowel stimuli was varied in a controlled manner by either increasing the amount of vocal jitter or by replacing the periodic glottal excitation by an aperiodic random sequence. A realistic manipulation of the vocal jitter was accomplished with a separately modeled glottal source, vocal tract transfer function, and lip radiation components and by introducing the cycle-to-cycle perturbation at the level of the glottal source. The present univariate manipulation of stimulus periodicity in terms of vocal jitter is in principle analogous to that of Hillenbrand (1988) apart from the significant difference in the implementation level: Here the manipulations in periodicity are realized at the level of the voicing source, the glottal flow generated by the vocal folds, rather than at the level of the speech waveform where the characteristics of the source and the vocal tract filter are combined. Since the real source of aperiodicity in dysphonic speech is the abnormal fluctuation of the vocal folds, the present approach enables a highly justified means to investigate the auditory processes underlying the perception of aperiodic speech. In order to study the perceptual consequences of these manipulations, the vocal quality of the quasi-periodic speech stimuli was studied with a behavioral scaling experiment. Together the results were expected to reveal the relationships between the acoustic, cortical, and perceptual indices of vocal jitter.

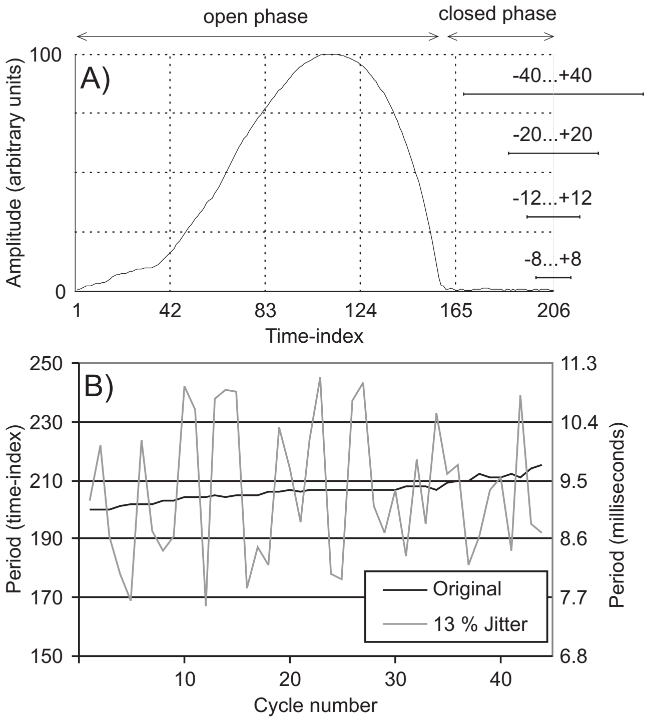


FIG. 1. Generation of speech stimuli of varying jitter by manipulating the duration of the glottal flow pulse. A glottal pulse extracted from a natural vowel is shown in (A) by indicating the open and the closed phase of the vocal fold cycle. Four new glottal flow excitation waveforms with increasing jitter were generated by concatenating copies of the glottal pulse with their durations randomly manipulated. Manipulation was accomplished by changing the length of the closed phase by a random number between $-R$ and R. R values of 8, 12, 20, and 40 samples were used and are depicted with interval markers in (A). Cycle-by-cycle period lengths are shown for the original vowel (with jitter <1%) and for the vowel with the largest jitter value (13% corresponding to $R=40$ samples) in (B).

II. METHODS

A. Subjects

Fifteen subjects (average age 25 years, standard deviation 6 years; 9 females) participated in the study with written informed consent. All subjects reported having normal hearing and being right-handed. The experiment was approved by the Ethical Committee of Helsinki University Central Hospital. During the experiment, the subjects, instructed not to pay attention to the auditory stimuli, concentrated on watching a silent video. The subjects were also instructed to avoid eye movements and blinks during MEG data acquisition.

B. Stimuli

In natural voice production, the periodic vocal fold vibration consists of two separate phases, the closed and the open phase. After the closed phase, the glottis typically starts to open gradually. As a result, the air flow through the vocal folds, the glottal flow pulse, starts to increase from the zero level toward the peak value which takes place when the vocal folds are maximally open [see Fig. 1(A)]. In the current vowel syntheses, vocal jitter was simulated by cycle-to-cycle manipulations in the period length of the stimuli. The objective was to generate these time-domain perturbations in a highly-controlled manner without causing artifactual changes in the shape of the glottal pulse during the open phase of the vocal fold vibration cycle when the flow is nonzero. Because any modifications of the pulse waveform during the open phase would result in artifacts such as high-frequency clicks or other changes in the pulse spectrum, the manipulations were strictly restricted to the closed phase of the vocal fold cycle when the amplitude of the flow is zero. Due to this restriction, it is impossible to generate realistic speech stimuli, having typical F0-values of 100 Hz, with local jitter values much more than approximately 15%. It is worth emphasizing that this is clearly a speech-specific issue and does not hold true for non-ecologic stimuli such as click trains whose periodic structure can be easily manipulated in any arbitrary manner. Therefore, in understanding the cortical basis of vocal jitter perception, the role of accurate modeling of the real voice production mechanism is very important as it enables reliable experimentation with speech sounds of plausible amounts of F0 perturbation.

The present stimuli were generated by manipulating the excitation source that was extracted by inverse filtering a vowel sound (/a/) uttered by a male Finnish speaker. New excitation waveforms were generated by concatenating copies of a representative glottal pulse that was estimated from the original glottal waveform. The starting values for the cycle lengths of the new sequences were derived from the original F0 history of the naturally-uttered vowel (mean F0 =107 Hz). The cycle lengths of the new sequences were then manipulated by changing the length of the closed phase of each glottal pulse by a random number that was uniformly distributed between $-R$ and R. R values of 8, 12, 20, and 40 samples were used in order to obtain four new glottal flow excitations of increasing jitter. In addition, an aperiodic excitation source was produced by generating a noise sequence from zero-mean random numbers of uniform distribution. The excitation waveforms, with matched overall spectral envelopes, were then passed back through the vocal tract filters that were extracted from naturally-uttered vowels /a/ and /e/. The vowel stimuli were easily identifiable despite the manipulations in their periodicity.

The jitter of the stimuli was calculated as the average difference between the cycle lengths of successive cycles and was divided by the average of all cycle lengths. The following percent jitter values were obtained: <1% (in the original utterance), 3% ($R=8$), 4% ($R=12$), 8% ($R=20$), and 13% ($R=40$). The amount of amplitude perturbation, shimmer, which might result from the glottal source manipulations, was inspected from the vowel speech waveforms. The obtained shimmer values ranged from 0.02 to 0.36 dB which are highly unlikely to contribute to the vocal roughness perception [cf. Hillenbrand (1988) where roughness ratings were unaffected by shimmer values <0.4 dB in the case of the uncorrelated amplitude sequences; the present aperiodic sequences were uncorrelated as well]. The degree of periodicity of the vowel stimuli was inspected by calculating the rms magnitudes of the autocorrelation series derived from the stimulus waveforms. This analysis showed that the degree of periodicity decreased monotonically with increasing

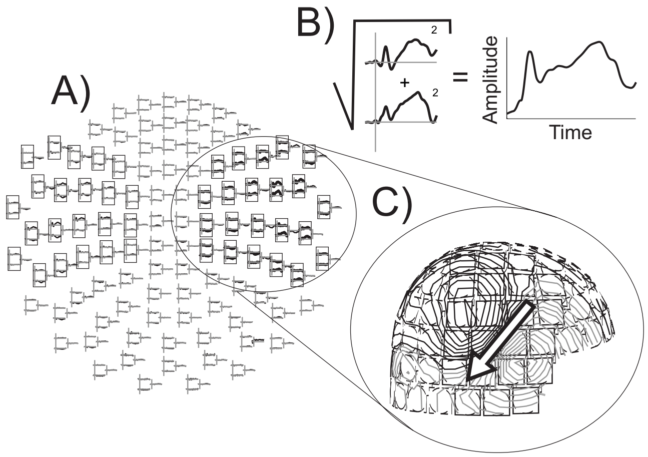


FIG. 2. Examples of MEG data analyses. The measurements consisted of 306 MEG channel waveforms (A) of which 44 gradiometer channel waveforms, obtained over the temporal area of either hemisphere (marked with rectangles), were used for the construction of vector sum waveforms (B) and for the estimation of ECDs (C).

jitter for the quasi-periodic vowels and that the noise-excited stimuli had the smallest degree of periodicity.

In order to obtain raw material for the SSG synthesis of the present study, two sustained vowels, /a/ and /e/, produced by a male Finnish speaker were recorded in an anechoic chamber using a high-quality condenser microphone (Bruel&Kjaer 4188). Sounds were both recorded and further processed with a sampling frequency of 22 050 Hz and a resolution of 16 bits. The duration of the vowel stimuli was set to 400 ms and their onsets and offsets were smoothed with a 10-ms Hanning window. The stimuli were presented to both ears of the subjects at 75-dB(A) sound pressure level with plastic tubes and earpieces, characterized by a 3-dB pass-band frequency response from 70 Hz to 4 kHz. The same audio equipment and experimental control software (PRESENTATION®, Neurobehavioral Systems, Inc.) were used for both behavioral and MEG experiments.

C. MEG data acquisition

Cortical activation elicited by the stimuli was registered with a 306-channel [Fig. 2(A)] whole-head MEG measurement device (Elekta Neuromag Oy, Finland) in a magnetically shielded room. The data were acquired with a recording bandwidth of 0.1–200 Hz and sampled at 600 Hz. Each of the 12 stimuli was presented in its own sequence in a Latin-square design order. At the beginning of each stimulus sequence, the head position with respect to the sensor array was determined by using head position indicator coils attached to the subject's scalp, with the locations of the coils with respect to the left and right preauricular points and the nasion having been determined prior to the measurement. The stimuli were presented at an onset-to-onset rate of 1500 ms, and 150 artifact-free evoked responses per stimulus were averaged over a period of 700 ms including a 100-ms pre-stimulus baseline. The epoch rejection criteria for the MEG and the electro-oculography sensors were set to 3000 fT/cm and 150 μV, respectively.

D. MEG data analysis

The MEG waveforms were baseline-corrected with respect to the 100-ms pre-stimulus interval and low-pass filtered at 20 Hz prior to MEG data analysis. The amplitudes of the N1m and the SF responses in both the left and the right hemisphere were investigated with vector sums from gradiometer channel pairs exhibiting maximum amplitude values. Latency analysis was restricted to the transient N1m responses because these (unlike SF responses) have well-defined peaks. The maximum amplitude points of the auditory N1m and SF were selected from the 85–150- and 300–400-ms post-stimulus latency ranges, respectively. Data from all 15 subjects indicated prominent N1m and SF responses in channels located over the auditory cortices and were thus included in the analysis of the response amplitudes and in the analysis of the N1m latency. The construction of a vector sum waveform from a pair of gradiometer waveforms is illustrated in Fig. 2(B).

The generator locations of the N1m and the SF responses were investigated with the ECD modeling technique in each hemisphere separately, with the assumption of a single dipole in a spherical volume conductor. The ECDs were fitted to the maximum amplitude points of the auditory N1m and SF, in the 85–150- and 300–400-ms post-stimulus latency ranges, respectively. The locations of the ECDs are reported in three dimensional coordinates where the x-, y-, and z-axes represent the lateral-medial, anterior-posterior, and superior-inferior dimensions, respectively. The head center is referenced by coordinates $x=0$, $y=0$, and $z=40$. Conditions with poor ECD fits were defined as dipoles with goodness-of-fit values <60%, anomalous locations, or orientations. Any poor ECD fits were considered missing values in the statistical analysis stage and led to subject rejection. The average goodness of fit of ECDs was over 85% for the ten (N1m response) and nine (SF response) subjects that were included in the statistical analyses. An example of ECD modeling results from the right-hemispheric sensor array is shown in Fig. 2(C).

E. Roughness rating experiment

The vocal quality of the dysphonic vowels created with SSG was studied with a perceptual scaling experiment, the aim of which was to investigate the relationship with vocal jitter and the perceived roughness of the vowel sounds. In this experiment, all the quasi-periodic vowel stimuli defined in Sec. II B were used. The vocal quality was assessed with an anchored perceptual scaling procedure where listeners assigned a numerical scale value to each test stimulus with the help of two reference (i.e., anchor) stimuli having the smallest (<1%) and the largest (13%) amount of vocal jitter in the current set of vowel stimuli. Each trial of the experiment consisted of three successive vowel stimuli with 400-ms silent periods between the stimuli. The first and the last stimulus were always the anchor stimuli, and the middle stimulus was selected randomly from the stimulus set as a test stimulus to be judged by the listener using a five-point rating scale. The smallest scale value (1) indicated the least amount of vocal roughness and the largest scale value (5) indicated the largest amount of perceived vocal roughness. Both vowels /a/ and /e/ were assessed, and the anchor stimuli used in each trial were always matched to the test stimuli with respect to vowel identity. Each test stimulus was repeated ten times. The subjects were allowed unlimited response time in each trial, and the roughness magnitude estimation judgments were entered with a computer keyboard.

F. Statistical analyses

The means of the MEG vector sum amplitudes and the ECD coordinates obtained from different stimulus conditions were compared with repeated measures analyses of variance (ANOVA). Mauchley sphericity tests were run in order to test the assumption of sphericity of data, and Greenhouse–Geisser corrections on the degrees of freedom were made when the assumption of sphericity was violated. The amplitude data were analyzed with a response × hemisphere × vowel × periodicity ANOVA where "response" comprised response types N1m and SF, "hemisphere" comprised the left and the right hemisphere, "vowel" comprised vowels /a/ and /e/, and the levels of factor "periodicity" consisted of the five quasi-periodic conditions as well as the aperiodic condition. The N1m latencies calculated from the vector sums were analyzed with a hemisphere × vowel × periodicity ANOVA. Similar three-way ANOVAs were performed for the ECD locations separately for the N1m and the SF responses.

The perceptual scaling scores of vocal roughness were formed on the basis of average ratings over the stimulus repetitions. The relationship between the vocal jitter and the roughness perception scores was investigated with a vowel × jitter ANOVA where the factor vowel was defined by the vowels /a/ and /e/ and the factor "jitter" consisted of the five vocal jitter values (<1%, 3%, 4%, 8%, and 13%).

Both the main and the interaction effects of the ANOVAs were investigated, and all statistically significant effects are reported. Newman–Keuls tests were used as a means of *post-hoc* analysis for pairwise differences in the data. The vector sum amplitudes, ECD coordinates, and perceptual scaling scores reported in Sec. III are mean values.

III. RESULTS

A. MEG experiment

Prominent N1m and SF responses were recorded in both cerebral hemispheres and were elicited by all the quasi-periodic and the aperiodic instances of the two vowels /a/ and /e/. Figure 3 shows grand-average vector sum waveforms from two pairs of MEG gradiometer sensors, the one yielding maximum response amplitudes over the left and the other over the right auditory cortex. The significant ANOVA results from MEG data are summarized in Table I.

The N1m responses were observed at the average latencies of 104 and 109 ms and with amplitudes of 54 and 60 fT/cm over the left and the right hemisphere, respectively. The SF responses reached their maxima at around 400 ms, with average amplitudes of 66 and 71 fT/cm in the left and the right hemisphere, respectively. The N1m and the SF amplitudes are shown in Fig. 4. While the N1m latencies were unvarying across the stimulus conditions, the N1m and the SF amplitudes were highly affected by the manipulations in the vowel periodicity ($F_{2.1,29.5}=21.23,\ p<0.001$). *Post-hoc* analyses for the aggregated values of the N1m and the SF responses of both hemispheres indicated that the amplitudes in the noise-excited aperiodic vowel condition (50.8 fT/cm) were reduced in comparison to the quasi-periodic conditions (60.0–68.1 fT/cm; p-values <0.001). Importantly, the amplitudes in the condition with the most severe vocal jitter of 13% (60.0 fT/cm) were reduced relative to the conditions with smaller vocal jitter values (64.1–68.1 fT/cm; p-values <0.05 for all comparisons except for the comparison with the 4% condition having a p-value <0.051).

The effect of periodicity manipulation on response amplitude differed somewhat between the N1m and the SF, as

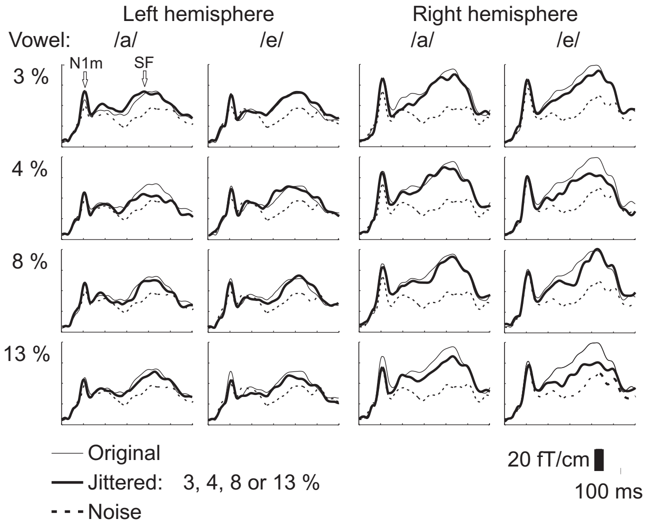


FIG. 3. Grand-average vector sum waveforms from the left- and the right-hemispheric gradiometer pairs yielding maximum response amplitudes and located directly above the left and the right auditory cortices of 15 subjects. The N1m and the SF response amplitudes were reduced for the aperiodic condition relative to the quasi-periodic conditions (original and 3%–13% jittered vowels). For the quasi-periodic conditions, the N1m and the SF amplitudes remained approximately constant for jitter values up to 8% but declined for the 13% jitter condition. The amount of jitter in the quasi-periodic vowel conditions is indicated with the percentages on the left side of the response waveforms. The measurements from the original vowel condition (thin line) and the aperiodic noise-excited vowel condition (dotted line) are drawn as references for the waveforms of the quasi-periodic conditions (thick line) where the vowel jitter was 3%, 4%, 8%, or 13%.

TABLE I. ANOVA results for MEG response amplitude (data from N1m and SF responses combined) and source location (data from N1m and SF responses separately). Only statistically significant ANOVA effects are shown.

Effect	*F*	df1	df2	*P*
Response amplitude (N1m and SF)				
Periodicity	21,23	2,1	29,5	<0.001
Response × periodicity	4,11	2,9	40,7	<0.05
N1m source *y*-coordinate				
Periodicity	7,27	2,6	23,1	<0.001
Hemisphere	26,25	1,0	9,0	<0.01
SF source *y*-coordinate				
Hemisphere	29,31	1,0	8,0	<0.01

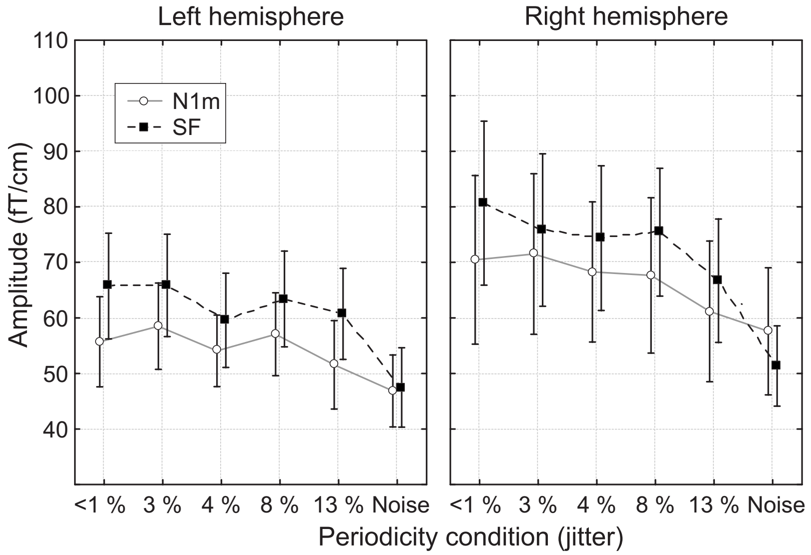


FIG. 4. The effect of periodicity manipulation on the N1m and the SF amplitudes in both hemispheres. The amplitudes were larger in quasi-periodic conditions relative to the aperiodic noise-excited vowel condition. Importantly, the amplitudes were reduced in the 13% jitter condition as compared with the conditions with smaller jitter values. The effect of periodicity manipulation was stronger for the SF than for the N1m responses. Responses for the two vowels, /a/ and /e/, were identical, and are therefore combined in the figure. The periodicity condition is expressed on the *x*-axis and the response amplitudes are indicated on the *y*-axis. The data are based on MEG responses of 15 subjects. Error bars represent standard error of the mean.

indicated by an interaction effect between response type and periodicity condition ($F_{2.9,40.7}=4.11$, $p<0.05$). From Fig. 4 it can be seen that the largest drop in amplitude occurred in the case of N1m already in the 13% jitter condition whereas in the case of SF the largest reduction in response amplitude occurred in the aperiodic noise condition. Thus, the N1m amplitude was reduced to the level of the aperiodic condition already in the 13% condition. Accordingly, *post-hoc* analyses showed that the difference between the 13% jitter condition and the noise-excited vowel condition was significant only for the SF response but not for the N1m.

The ECD locations of the recorded responses indicated activation in the temporal regions of both hemispheres, in the vicinity of the auditory cortical areas. As shown in Fig. 5, the ECDs for the N1m response were on the average located at *x*-coordinates −52 and 54 mm and at *z*-coordinates 52 and 53 mm in the left and the right hemisphere, respectively. Interestingly, a periodicity-specific effect was revealed in the *y*-axis position of the N1m source as the source locations in the quasi-periodic conditions were located anterior to those obtained in the case of the noise-excited vowels ($F_{2.6,23.1}=7.27$, $p<0.001$; *post-hoc* test *p*-values <0.001 for all comparisons between the aperiodic condition against the quasi-periodic conditions). In the case of the quasi-periodic and the aperiodic vowel stimuli, the ECDs were located at *y*-coordinates −4 and −7 mm in the left hemisphere and at 3 and 0 mm in the right hemisphere, respectively. The manipulations in vowel periodicity had no effect on the *x*- or *z*-coordinates of the ECDs. No significant differences in ECD locations were found among the quasi-periodic conditions.

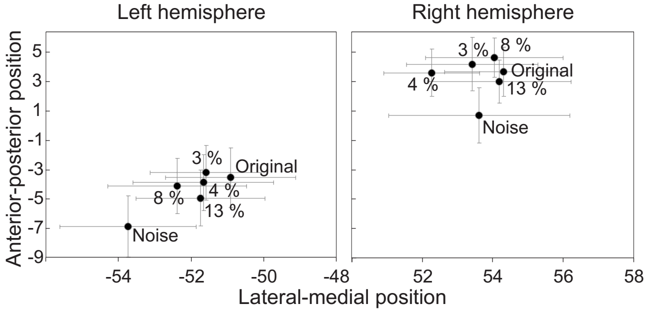


FIG. 5. N1m source locations on the lateral-medial and anterior-posterior axes. The source locations were more anterior in the quasi-periodic conditions than in the aperiodic noise condition. Responses for the two vowels /a/ and /e/ were identical, and are therefore combined in the figure. The periodicity condition is indicated next to the data points and error bars represent standard error of the mean. The data are based on ECDs from ten subjects.

The ECDs for the SF response were located at *x*-coordinates −48 and 49 mm, *y*-coordinates −2 and 7 mm, and *z*-coordinates 47 and 47 mm in the left and the right hemisphere, respectively. The periodicity condition did not affect the ECD coordinates of the SF response. The right-hemispheric responses were generated anterior to their left-hemispheric counterparts in case of both the N1m ($F_{1,9}=26.25$, $p<0.01$) and the SF ($F_{1,8}=29.31$, $p<0.01$) responses.

B. Roughness rating experiment

In the perceptual roughness scaling experiment, the listeners' roughness ratings followed the amount of F0 perturbation in the vowel stimuli without excessive uncertainty (Fig. 6). The roughness scores increased monotonically as a function of vowel jitter ($F_{2.1,29.8}=180.71$, $p<0.001$; *post-hoc* test *p*-values <0.001 for all pair-wise comparisons).

C. Correlations between cortical responses and roughness ratings

In order to investigate the relationship between the amplitudes of the cortical responses and the behavioral ratings of perceived roughness, the correlations between these two measures were evaluated. The correlation coefficients were calculated between the roughness ratings and the N1m and SF amplitudes, normalized subject wise. The coefficients between the N1m and the roughness rating were −0.23 and −0.34 in the left and the right hemisphere, respectively. The corresponding coefficients between the SF and the roughness score were −0.12 and −0.33. All these correlations were statistically significant, except for the correlation between

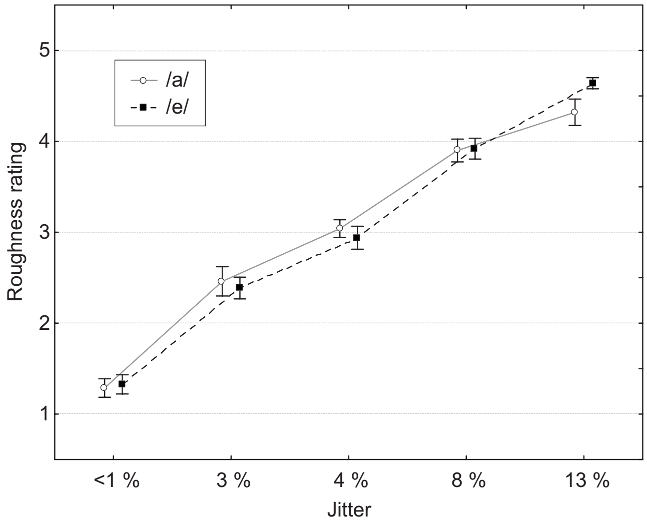


FIG. 6. The effect of stimulus vocal jitter on roughness perception scores. The roughness scores increased with every increase in vocal jitter although the rate of this increase was reduced for the high jitter values (8%–13%). The jitter is expressed on the x-axis and the roughness scores on the y-axis. The scores are based on the ratings of 15 subjects. Error bars represent standard error of the mean.

the roughness score and the left-hemispheric SF amplitude. Together this analysis shows that an increased roughness rating is related to a decreased amplitude of the cortical responses.

IV. DISCUSSION

In the current experiments, both cortical neuromagnetic responses and behavioral scaling scores to acoustically modeled dysphonic voice of varying periodic structure were studied. In the MEG measurements, the N1m and the SF responses, both sensitive to speech periodicity (Yrttiaho *et al.*, 2008), were recorded. The largest difference in amplitude occurred between the responses elicited by the quasi-periodic and the aperiodic noise-excited speech sounds. Importantly, the amount of vocal jitter was reflected in the cortical responses: the N1m and SF amplitudes decreased when the random F0 perturbation of the vowel stimuli was increased to a jitter value of 13%. These dependencies of the auditory N1m and SF amplitudes on vocal jitter were observed for both of the two vowels, /a/ and /e/, used in the current study. The perceived rough vocal quality of the quasi-periodic vowels, measured with the behavioral rating experiment, was shown to increase as a function of vocal jitter. Thus, the present manipulations in vocal jitter of the vowel sounds, realized at the level of the glottal source excitation, successfully produced the rough perceptual quality associated with aperiodic abnormal phonation. Together, the present measurements reveal a significant relationship between the activation of auditory areas and the roughness perception of dysphonic speech sounds.

As the amplitudes of the currently recorded N1m and SF responses decreased in the face of both complete and partial reductions in vowel periodicity, it seems that vocal aperiodicity is represented in the auditory cortex with decreased neural activity in periodicity-sensitive cell populations. This interpretation fits well with the emerging picture of general cortical periodicity sensitivity in the domain of both speech (Alku *et al.*, 2001; Tiitinen *et al.*, 2005; Yrttiaho *et al.*, 2008) and non-speech (e.g., Krumbholz *et al.*, 2003; Soeta *et al.*, 2005; Gutschalk *et al.*, 2002, 2004) related auditory processing. According to this, periodic sounds lead to higher levels of activation than aperiodic sounds. Importantly, the current study fills in a significant gap in generalizing this previously observed periodicity-specificity of cortical responses to the field of aperiodic dysphonic voice quality, previously unexplored by human electrophysiology. However, it remains to be explored whether similar periodicity-sensitive cortical dynamics hold for different kinds of aperiodicities. For example, the current study utilized both noise- and jitter-based aperiodicities, which result in distinct perceptual qualities. Here, a larger reduction in cortical response amplitude was observed in the case of noise excitation than in the case of jitter. While this difference is most likely due to the larger random aperiodicity in the noise-excited vowels, the possible additional contribution of the type of aperiodicity (noise versus jitter) cannot be ruled out. To this end, further experiments are needed in which stimuli with different types of aperiodicities are varied along a general index of the degree of periodicity such as autocorrelation.

In the current study, the effects of two different vowel aperiodicities on the cortical transient and the sustained responses were investigated. Relative to the condition of healthy phonation, both 13% jitter and noise excitation resulted in decreased amplitudes of the N1m and SF responses. However, the amount of amplitude decrement differed between the N1m and SF. Namely, the SF was relatively more affected by the noise excitation whereas the N1m was approximately equally affected by jitter and noise excitation. Thus, the SF appears to be more vulnerable to noise aperiodicity than the N1m, and the N1m is more vulnerable to vocal jitter than the SF.

The larger reduction in SF than in N1m responses in the case of aperiodic noise-excited speech sounds, as compared to the periodic conditions, was also observed in Yrttiaho *et al.* (2008). A plausible explanation for this difference might be found in the mixture of activation in sound onset-sensitive and periodicity-sensitive cell populations (see Gutschalk *et al.*, 2004; Schönwiesner and Zatorre, 2008): as activity generated by onset-selective cells may be relatively smaller in the SF than in the N1m, the SF is more dependent on sound periodicity than the N1m where the activity of onset-sensitive and periodicity-sensitive sources is presumably combined. Some evidence for the smaller onset-dependency of SF as opposed to that of the initial transient N1m response may be found in Mäkinen *et al.* (2004) where the amplitude of the transient response appears to vary with the rise-time of the auditory stimulus while the SF remains roughly constant across different stimulus rise-times. Unfortunately, as the focus of Mäkinen *et al.* (2004) was response timing, precise amplitude analyses were not reported.

The reverse case of greater vulnerability of the N1m to vocal jitter relative to the SF resembles the results of Gutschalk *et al.* (2007) who found that increases in the aperiodicity of click-trains affected the transient responses more than the sustained responses. This effect could in principle be explained by the differing durations of periodicity information that can be accumulated up to the peak level of these responses. Increasing stimulus duration enhances the N1 only up to 50 ms (Näätänen and Picton, 1987), whereas the SF reaches its maximum at 400 ms of continued stimulation. Therefore, the N1m is elicited upon less periodicity information than the SF, whose generators may be able to extract periodicity information over a longer time window and therefore may be able to detect this periodicity even when it is partially corrupted by vocal jitter. Taken together, the current results tentatively suggest that cortical coding of stimulus periodicity could vary along the time axis of neural activity as revealed by the N1m and the SF responses. However, further experiments are needed to adequately test the explanations suggested here.

Vowel periodicity was also reflected in the ECD location of the N1m response. As in the previous studies investigating cortical responses to speech sound periodicity (Alku *et al.*, 2001; Tiitinen *et al.*, 2005; Yrttiaho *et al.*, 2008), the ECDs of the N1m in the quasi-periodic conditions (comprising here the original and the jitter-manipulated vowels) were located anterior to those obtained in the aperiodic noise-excited conditions. By using depth electrode recordings in a human patient, Schönwiesner and Zatorre (2008) provided direct evidence for the existence of a periodicity-sensitive area close to the lateral tip of Heschl's gyrus and which has previously been suggested by both MEG (Krumbholz *et al.*, 2003; Gutschalk *et al.*, 2004) and functional magnetic resonance imaging (Griffiths *et al.*, 1998; Patterson *et al.*, 2002; Penagos *et al.*, 2004) results. A plausible explanation for the location difference between the N1m recorded in the aperiodic and quasi-periodic vowel conditions of the current study may therefore be based on selective activation of the periodicity-sensitive source by the quasi-periodic but not by the aperiodic vowels. Interestingly, all the quasi-periodic vowels elicited N1m responses which had sources anterior to the source of the N1m to the aperiodic vowel. Consistent with the explanation based on a periodicity-specific generator, the quasi-periodic vowels were perhaps still sufficiently periodic to activate the brain region(s) specialized in the processing of auditory periodicity. In contrast to the N1m source location effect, no such difference in the ECD locations of the SF responses arising out of vowel periodicity manipulations was observed here. This absence of an ECD location difference between periodic and aperiodic conditions also runs contrary to the previous studies using speech (Yrttiaho *et al.*, 2008) and non-speech sounds (Gutschalk *et al.*, 2002, 2004) which report that the SF source location is sensitive to sound periodicity. It should, however, be noted that the SF responses recorded in the aperiodic condition of the current experiment were substantially reduced in amplitude relative to those recorded in the quasi-periodic conditions. This reduction in the signal-to-noise ratios in the neuromagnetic responses may have rendered the previously observed location differences undetectable in the current data.

In the present behavioral measures of perceived vocal quality, two basic effects were observed. First, the roughness ratings increased as a function of vocal jitter in the stimuli. Second, the rate of this increase diminished when the vocal jitter was increased. In these two respects, the current roughness scaling results were similar to those of Wendahl (1966), who used saw-tooth analogs of vowel sounds, and of Hillenbrand (1988), who used synthetic vowels. The current results, however, differ from the findings of Hillenbrand (1988) where the roughness ratings saturated already at jitter values of 2%. In contrast, the currently obtained roughness curve reached a deceleration point at jitter values of 8%–13%, which is near the corresponding threshold value of 6%–10% of Wendahl (1966). These discrepancies may be explained by both task- and stimulus-related factors. The most significant task-related difference between the studies was that listeners in the Hillenbrand (1988) study rated vowels without external references, whereas, in the current study, two reference stimuli were presented to the listeners for each test vowel to be rated. In this respect, the procedure used by Wendahl (1966) was in principle similar to the current rating task, as the roughness estimates were derived from pair-wise comparisons of the stimuli. According to Kreiman *et al.* (2007), the rating performance is significantly facilitated by providing a reference stimulus to the listener. Therefore, the earlier deceleration in roughness scores observed in Hillenbrand (1988) as compared to Wendahl (1966) and the current results may arise from the increased listener uncertainty related to the use of an implicit standard as opposed to an external reference stimulus. The saw-tooth stimuli used by Wendahl (1966) were obviously different from the synthetic speech sounds used in the current experiments as well as from those used by Hillenbrand (1988). The manipulation of vocal jitter also differed between the approach based on the manipulation of the speech waveform used by Hillenbrand (1988) and the current semi-synthetic approach which enabled the vocal jitter manipulations at the level of the glottal source. Finally, earphones were used in the current study whereas Hillenbrand (1988) used loudspeakers which, as he noted, may make the stimuli sound rougher and thus shift the roughness curve toward the smaller jitter values.

Given that the roughness perception and the listener rating behavior may vary considerably between studies using different rating tasks and stimuli, the perceptual evaluation of the vocal aperiodicity of the speech stimuli is critical for the inferences to be made concerning the cortical basis of roughness perception. If the relationship between cortical responses and the stimulus jitter parameter is not paralleled by perceptual measures, the observed jitter-dependent cortical dynamics are unlikely to reflect the perception of vocal roughness quality. Therefore, in addition to cortical measurements, a simultaneous demonstration that the same manipulations in the jitter of vowel stimuli result in increased roughness of the perceived vocal quality is needed. In the current study, the manipulations of vocal jitter were, indeed, shown to affect both the perception of vocal roughness and the cortical responses. Furthermore, the roughness ratings were shown to be negatively correlated with the normalized amplitudes of the cortical responses of individual subjects. This parallel relationship between vocal jitter, roughness sensation, and N1m and SF amplitude reduction suggests a role for the generators of these auditory evoked responses in the cortical basis of roughness perception.

The current study represents an initial investigation of the representation of dysphonic vocal quality in the human brain by providing both cortical and behavioral measures of aperiodic F0 perturbation of speech sounds. The results suggest that the vocal quality of roughness is represented in the auditory cortex by reduced activation of periodicity-sensitive populations which contribute to the generation of the N1m and the SF responses. The initial success in applying MEG to study the cortical processing of F0 perturbation is encouraging and paves the way for further investigations into the link between the physiology of the human auditory system and the perception of voice quality.

ACKNOWLEDGMENTS

This study was supported by the Academy of Finland (Project Nos. 111848, 217114, 217082, and 217113).

Alku, P., Sivonen, P., Palomäki, K., and Tiitinen, H. (**2001**). “The periodic structure of vowel sounds is reflected in human electromagnetic brain responses,” Neurosci. Lett. **298**, 25–28.

Alku, P., Tiitinen, H., and Näätänen, R. (**1999**). “A method for generating natural-sounding speech stimuli for cognitive brain research,” Clin. Neurophysiol. **110**, 1329–1333.

Deal, R. E., and Emanuel, F. W. (**1978**). “Some waveform and spectral features of vowel roughness,” J. Speech Hear. Res. **21**, 250–264.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., and Patterson, R. D. (**1998**). “Analysis of temporal structure in sound by the human brain,” Nat. Neurosci. **1**, 422–427.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., and Scherg, M. (**2002**). “Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex,” Neuroimage **15**, 207–216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., and Rupp, A. (**2004**). “Temporal dynamics of pitch in human auditory cortex,” Neuroimage **22**, 755–766.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., and Rupp, A. (**2007**). “The effect of temporal context on the sustained pitch response in human auditory cortex,” Cereb. Cortex **17**, 552–561.

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., and Lounasmaa, O. V. (**1993**). “Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of signal processing in the human brain,” Rev. Mod. Phys. **65**, 413–497.

Hari, R., Aittoniemi, K., Järvinen, M. L., Katila, T., and Varpula, T. (**1980**). “Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators,” Exp. Brain Res. **40**, 237–240.

Hertrich, I., Mathiak, K., Lutzenberger, W., and Ackermann, H. (**2000**). “Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields,” NeuroReport **11**, 4017–4020.

Hillenbrand, J. (**1988**). “Perception of aperiodicities in synthetically generated voices,” J. Acoust. Soc. Am. **83**, 2361–2371.

Horii, Y. (**1979**). “Fundamental-frequency perturbation observed in sustained phonation,” J. Speech Hear. Res. **22**, 5–19.

Iwata, S., and von Leden, H. (**1970**). “Pitch perturbations in normal and pathologic voices,” Folia Phoniatr. (Basel) **22**, 413–424.

Kreiman, J., Gerratt, B. R., and Ito, M. (**2007**). “When and why listeners disagree in voice quality assessment tasks,” J. Acoust. Soc. Am. **122**, 2354–2364.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). “Perceptual evaluation of voice quality—Review, tutorial, and a framework for future-research,” J. Speech Hear. Res. **36**, 21–40.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., and Lütkenhöner, B. (**2003**). “Neuromagnetic evidence for a pitch processing center in Heschl’s gyrus,” Cereb. Cortex **13**, 765–772.

Lieberman, P. (**1963**). “Some acoustic measures of fundamental periodicity of normal and pathologic larynges,” J. Acoust. Soc. Am. **35**, 344–353.

Mäkinen, V., May, P., and Tiitinen, H. (**2004**). “Transient brain responses predict the temporal dynamics of sound detection in humans,” Neuroimage **21**, 701–706.

Muñoz, J., Mendoza, E., Fresneda, M. D., Carballo, G., and López, P. (**2003**). “Acoustic and perceptual indicators of normal and pathological voice,” Folia Phoniatr. Logop. **55**, 102–114.

Murry, T., and Doherty, E. T. (**1980**). “Selected acoustic characteristics of pathologic and normal speakers,” J. Speech Hear. Res. **23**, 361–369.

Näätänen, R., and Picton, T. (**1987**). “The N1 wave of the human electric and magnetic response to sound—A review and an analysis of the component structure,” Psychophysiology **24**, 375–425.

Pantev, C., Eulitz, C., Elbert, T., and Hoke, M. (**1994**). “The auditory-evoked sustained field—Origin and frequency-dependence,” Electroencephalogr. Clin. Neurophysiol. **90**, 82–90.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (**2002**). “The processing of temporal pitch and melody information in auditory cortex,” Neuron **36**, 767–776.

Penagos, H., Melcher, J. R., and Oxenham, A. J. (**2004**). “A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging,” J. Neurosci. **24**, 6810–6815.

Picton, T. W., Woods, D. L., and Proulx, G. B. (**1978a**). “Human auditory sustained potentials. 1. Nature of response,” Electroencephalogr. Clin. Neurophysiol. **45**, 186–197.

Picton, T. W., Woods, D. L., and Proulx, G. B. (**1978b**). “Human auditory sustained potentials. 2. Stimulus relationships,” Electroencephalogr. Clin. Neurophysiol. **45**, 198–210.

Schönwiesner, M., and Zatorre, R. J. (**2008**). “Depth electrode recordings show double dissociation between pitch processing in lateral Heschl’s gyrus and sound onset processing in medial Heschl’s gyrus,” Exp. Brain Res. **187**, 97–105.

Soeta, Y., Nakagawa, S., and Tonoike, M. (**2005**). “Auditory evoked magnetic fields in relation to iterated rippled noise,” Hear. Res. **205**, 256–261.

Tiitinen, H., Mäkelä, A. M., Mäkinen, V., May, P. J. C., and Alku, P. (**2005**). “Disentangling the effects of phonation and articulation: Hemispheric asymmetries in the auditory N1m response of the human brain,” BMC Neurosci. **6**, 62–70.

Wendahl, R. W. (**1966**). “Some parameters of auditory roughness,” Folia Phoniatr. (Basel) **18**, 26–32.

Yrttiaho, S., Tiitinen, H., May, P. J. C., Leino, S., and Alku, P. (**2008**). “Cortical sensitivity to periodicity of speech sounds,” J. Acoust. Soc. Am. **123**, 2191–2199.

Sensitivity of the human auditory system to temporal fine structure at high frequencies

Brian C. J. Moore[a)]
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Aleksander Sęk
Institute of Acoustics, Adam Mickiewicz University, 85 Umultowska, 61-614 Poznań, Poland and Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 11 September 2008; revised 14 February 2009; accepted 27 February 2009)

The frequency of sounds is coded partly by phase locking to the temporal fine structure (TFS) of the waveform evoked on the basilar membrane. On the basis of data obtained using sinusoids, it is usually assumed that in mammals, including humans, TFS information becomes unusable for frequencies above about 5000 Hz. Here, sensitivity to the TFS of complex sounds up to much higher frequencies is demonstrated. Subjects discriminated a harmonic complex tone, with a fundamental frequency F0, from a tone in which all harmonics were shifted upwards by the same amount in hertz. The phases of the components were selected randomly for every stimulus. Both tones had an envelope repetition rate equal to F0, but the tones differed in their TFS. To prevent discrimination based on spectral cues, the tones were passed through a fixed bandpass filter, centered at 14F0. A background noise was used to mask combination tones. Performance was well above chance for most subjects when F0 was 800 or 1000 Hz and all audible components were above 8000 Hz. Supplementary experiments confirmed that performance was not based on changes in the excitation pattern or on the discrimination of partially resolved components.
© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106525]

PACS number(s): 43.66.Fe, 43.66.Hg [RYL] Pages: 3186–3193

I. INTRODUCTION

Complex sounds are analyzed within the cochlea into a series of bandpass-filtered signals. Each such signal has a relatively slowly varying temporal envelope superimposed on a higher-frequency carrier [the temporal fine structure (TFS)]. Information about the envelope is conveyed in the auditory nerve mainly by changes in firing rate over time, while information about TFS is conveyed by the detailed timing of the nerve impulses (in other words, phase locking) (Rose *et al.*, 1967). Phase locking weakens at high frequencies (Johnson, 1980; Palmer and Russell, 1986), and it is usually assumed that in mammals the information becomes unusable for frequencies above about 5000 Hz, although physiological measurements show that residual phase locking is present at higher frequencies (Heinz *et al.*, 2001; Recio-Spinoso *et al.*, 2005). The assumption that sensitivity to TFS in humans is lost above 5000 Hz is mainly based on behavioral data showing changes in the perception of sinusoids when their frequencies fall above 5000 Hz (Moore, 2003). For example, thresholds for detecting a change in frequency, expressed as Weber fractions, increase above about 5000 Hz (Moore, 1973), and the ability to identify musical intervals (Ward, 1954) or melodies (Attneave and Olson, 1971) worsens for frequencies above 5000 Hz. Also, people with absolute pitch have a reduced ability to name notes whose frequencies fall above 5000 Hz (Ohgushi and Hatoh, 1991).

While there is considerable evidence supporting the idea that the perception of the pitch of sinusoids changes as their frequency is increased above 5000 Hz, it is less clear that the perception of complex tones changes when all of their components fall above 5000 Hz. Ritsma (1962, 1963) reported that complex tones did not evoke a clear residue pitch when all of their components fell above 5000 Hz, but a recent study of Oxenham and Keebler (2008) showed that complex tones with all components at 6000 Hz or above were capable of evoking a musical pitch. In the present experiments, we assessed whether sensitivity to the TFS of complex tones persists for frequencies above 5000 Hz. Three experiments were performed. In the main experiment, the task was similar to that described by Hopkins and Moore (2007) and Moore and Sęk (2009). Subjects were required to discriminate two complex tones which differed in their TFS but not in their envelope repetition rate. We believe that this task provides a direct measure of sensitivity to the TFS of complex tones, provided that the frequency components of the tones are unresolved. In the first supplementary experiment, we assessed the possible role of excitation-pattern cues in the main experiment, by examining the smallest detectable change over a restricted region of the excitation pattern; this was done by measuring thresholds for detecting a change in level of a single component in one of the complex tones. In the second

[a)]Author to whom correspondence should be addressed. Electronic mail: bcjm@cam.ac.uk.

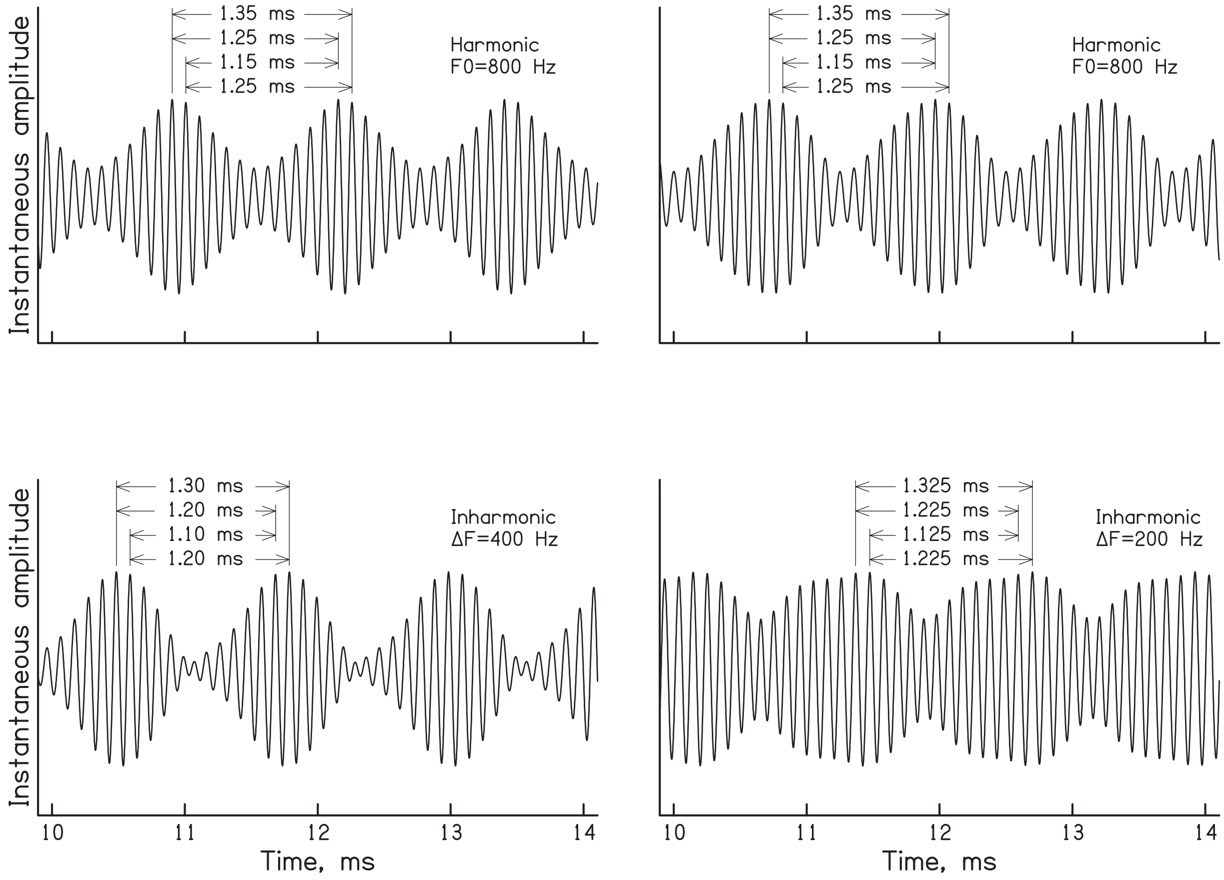


FIG. 1. Illustration of the waveforms of harmonic stimuli (top) and inharmonic stimuli (bottom) after the waveforms have been passed through a simulated auditory filter centered at 9600 Hz. For each panel, a different random selection of starting phases was used.

supplementary experiment, we assessed whether discrimination in the main experiment might have been based on discrimination of partially resolved components; this was done by comparing thresholds obtained in the main experiment with thresholds for discriminating the frequency of an isolated high-frequency sinusoid. We conclude that sensitivity to the TFS of complex tones persists for frequencies up to at least 8000 Hz.

II. MAIN EXPERIMENT

A. Stimuli

The task and stimuli were similar to those described by Hopkins and Moore (2007) and Moore and Sęk (2009). Subjects had to discriminate a harmonic complex tone (H), with a fundamental frequency F0, from a tone in which all harmonics were shifted upward by the same amount in hertz, ΔF, resulting in an inharmonic tone (I). Usually, tone I was heard as having a higher pitch than tone H, for reasons that are explained below. Both tones had an envelope repetition rate equal to F0, but the tones differed in their TFS. The phases of the components were selected randomly for every stimulus. As a result, the shape of the envelope was different for every stimulus, which prevented the use of cues based on envelope shape. The value of F0 was either 800 or 1000 Hz. To prevent discrimination based on spectral cues, related to changes in the excitation pattern (Glasberg and Moore, 1990), all tones were passed through a fixed bandpass filter, centered on the high (unresolved) components. The filter (designed using the FIR2 function in MATLAB, with 512 taps at a 50-kHz sampling rate) had a central flat region with a width of 5F0 and was centered at 14F0. The skirts of the filter decreased in level at a rate of 30 dB/octave. The use of a relatively shallow slope helps to reduce differences between the excitation patterns for the H and I tones; such differences are discussed in more detail later.

Figure 1 shows examples of the waveforms of H and I tones with F0=800 Hz after they have been passed through a simulated auditory filter centered at 9600 Hz, which is at the lower end of the passband of the filter used to generate the tones. It is often assumed that, for complex tones containing only unresolved harmonics, but with some harmonics below the 14th, the pitch is determined from the time intervals between peaks in the TFS close to adjacent envelope maxima (de Boer, 1956; Schouten *et al.*, 1962; Moore and Moore, 2003). For the samples of tone H shown in the top two panels, the most prominent time interval (the interval that occurs most often) is 1.25 ms, but other intervals such as 1.15 and 1.35 ms also occur. For the I tone with ΔF =400 Hz (left panel), the most prominent intervals are 1.1, 1.2, and 1.3 ms. This value of ΔF corresponds to 0.5F0, the value for which the tones H and I are most different. For ΔF=200 Hz (right panel), the most prominent intervals are 1.125, 1.225, and 1.325 ms. The differences in intervals between the H and I tones can, in principle, account for the difference in pitch between the two tones. Note that the envelope modulation depth depends on the specific random selection of starting phases of the frequency components of the tones and varies randomly from one tone burst to the next, for both the H and I tones. Thus, the envelope modulation depth could not be used as a cue to perform the task.

A background threshold-equalizing noise (TEN, Moore *et al.*, 2000) was used to mask combination tones and to limit the audibility of components of the complex tones falling outside of the passband of the filter. The TEN level is specified as the level in a 1-ERB_N wide band centered at 1 kHz, where ERB_N is the equivalent rectangular bandwidth of the auditory filter as determined for young normally hearing listeners at moderate sound levels (Glasberg and Moore, 1990). The level of the TEN was either 10 or 15 dB below the overall level of the complex-tone signal.

All stimuli were generated using a personal computer with a Sound Blaster Audigy 2 24-bit sound card, using a sampling rate of 50 kHz. The H and I tones were presented at an overall level that was 20 dB above the absolute threshold for sinusoidal tones at the center frequency of the passband. Stimuli were presented using one earpiece of a Sennheiser HD580 headset.

B. Procedure

A two-interval two-alternative forced-choice (2AFC) method was used to measure the value of ΔF required for threshold. A standard 2AFC method, with the H tone in one interval and the I tone in the other, is not suitable for these stimuli, since the I tone is usually heard as higher in pitch than the H tone, but is sometimes heard as lower in pitch (Moore and Sęk, 2009). We used the task described by Moore and Sęk (2009), which they found to give stable results and very small learning effects. In one interval (selected randomly), there were four successive bursts of tone H, separated by 100 ms. In the other interval, tones H and I alternated, with the same 100-ms inter-burst interval, giving the pattern HIHI. The phase of the components was selected randomly for every H tone and every I tone. The task of the subject was to choose the interval in which the sound changed across the four tone bursts within an interval. Feedback as to the correct answer was given on the computer screen. The duration of each tone burst was 200 ms (including 20-ms raised-cosine rise/fall ramps) and the two intervals were separated by 300 ms. A two-down one-up procedure was used to track the 70.7% correct point on the psychometric function. Following two correct responses in a row, the value of ΔF was decreased, while following one incorrect response it was increased. The procedure continued until eight turnpoints had occurred. The value of ΔF was changed by a factor of 1.953 (1.25^3) until one turnpoint had occurred, by a factor of 1.5625 (1.25^2) until the second turnpoint had occurred, and by a factor of 1.25 thereafter. The "threshold" was estimated as the geometric mean of the values of ΔF at the last six turnpoints.

The value of ΔF was not allowed to exceed 0.5F0. If the adaptive procedure called for a value of ΔF greater than 0.5F0 three times during a run, the value of ΔF was fixed at 0.5F0 and 20 more trials were presented. The score was then given as the percentage correct out of 20. At least three runs were obtained for each condition. If the adaptive procedure could not be completed for all three runs, then one or more extra runs were completed.

It should be noted that the repetition of the H and I tones within each interval makes the task easier than if there were only two tone bursts within each interval (HI in one and HH in the other). The method used here tracks the value of ΔF required for discrimination of the HIHI stimulus from the HHHH stimulus with a level of performance corresponding to $d'=0.78$. Making some reasonable assumptions from signal detection theory (Macmillan and Creelman, 1991), the value of d' that would be obtained for the same value of ΔF if there were only two tone bursts in each interval would be $\sqrt{2}$ smaller, i.e., about 0.55, which is rather small. This should be borne in mind when interpreting the results that are presented below.

C. Measurement of absolute thresholds

Absolute thresholds for detecting a sinusoid at the center frequency of the bandpass filter used to generate the complex tones were measured using a two-interval forced-choice task. The sinusoidal signal could occur either in interval 1 or interval 2, selected at random. The signal lasted 200 ms, including 20-ms raised-cosine rise-fall times, and the intervals were separated by 500 ms. The intervals were indicated by boxes on a computer screen (labeled 1 and 2), each of which was lit up in blue during the appropriate interval. The adaptive procedure was similar to that described above. Six turnpoints were obtained. The step size was initially 6 dB. It was changed to 4 dB after one turnpoint and to 2 dB after the second turnpoint. The threshold was taken as the mean signal level at the last four turnpoints.

D. Subjects

Nine subjects with normal hearing (audiometric thresholds better than 20 dB HL from 250 to 8000 Hz) were tested with F0=800 Hz. Their ages ranged from 22 to 63 years. Their absolute thresholds at the center frequency of 11 200 Hz (used with F0=800 Hz) ranged from 8 to 42 dB SPL. Eight subjects with normal hearing were tested with F0=1000 Hz, four of whom had previously been tested with F0=800 Hz. Their ages ranged from 24 to 50 years. Their absolute thresholds at the center frequency of 14 000 Hz (used with F0=1000 Hz) ranged from 9 to 50 dB SPL.

E. Results

We present first the results for F0=800 Hz. Initially, subjects were tested with the TEN level/ERB_N set 15 dB below the overall level of the complex tone. With this level, each component within the passband would have been about 10 dB above its masked threshold if it were presented alone. The lowest audible component on the lower skirt of the filter would have been at 8000 Hz. Seven of the nine subjects were able to complete all three adaptive runs, i.e., they could consistently perform the task. The geometric mean threshold value of ΔF for these subjects was 143 Hz, which corresponds to 0.18F0. The individual thresholds (in ascending order) were 64, 106, 110, 158, 168, 223, and 279 Hz. The remaining two subjects each completed one adaptive run, and scored 60% and 90% correct with $\Delta F = 0.5F0$ for the remaining runs (including the results for one or two extra runs for these subjects). Even for the subject who scored most poorly, the score was significantly above chance ($p < 0.02$), based on a one-tailed binomial test.

To check that the subjects were not using combination tones to perform the task, they were re-tested with the TEN level/ERB_N set only 10 dB below the overall level of the complex tone. This did lead to poorer performance, as would be expected. However, three of the nine subjects were able to complete all three adaptive runs, achieving mean thresholds of 77, 190, and 308 Hz. For the remaining subjects, the runs where the adaptive procedure could be completed were assigned a percent correct score of 71% (the value tracked by the adaptive procedure), and this score was averaged with the scores for the runs where the non-adaptive procedure was used (there were between 4 and 5 such runs). The resulting percent correct scores were 58, 60, 60, 64, 68, and 73. All scores were significantly above chance ($p < 0.005$) based on a one-tailed binomial test. Thus, all subjects showed some ability to perform the task even at this very adverse signal-to-noise ratio.

For F0 = 1000 Hz, the TEN level/ERB_N was set 15 dB below the overall level of the complex tone. The lowest audible component on the lower skirt of the filter would have been at 10 000 Hz. Four of the eight subjects were able to complete all three adaptive runs. The geometric mean thresholds for these subjects were 110, 163, 259, and 274 Hz. Of the remaining subjects, one completed two out of three of the adaptive runs and two completed one adaptive run. For these subjects, the runs where the adaptive procedure could be completed were assigned a percent correct score of 71%, and this score was averaged with the scores for the runs where the non-adaptive procedure was used. The resulting percent correct scores were 51, 53, 67, and 69. The first two of these scores did not differ significantly from chance at $p = 0.05$, whereas the latter two scores were significantly greater than chance ($p < 0.005$). Thus, six of the eight subjects were able to perform the task at above-chance levels.

It is noteworthy that the subject who performed most poorly had the highest absolute threshold of the group at 14 000 Hz (50 dB SPL).

Overall, the results show that all subjects could perform the task at above-chance levels when F0 was 800 Hz and the lowest audible component fell close to 8000 Hz, and six of the eight subjects could perform the task at above-chance levels when F0 was 1000 Hz and the lowest audible component fell at about 10 000 Hz.

III. SUPPLEMENTARY EXPERIMENT 1: ASSESSING THE ROLE OF EXCITATION-PATTERN DIFFERENCES

A. Rationale

One concern was whether subjects might be performing the task using residual place cues, i.e., cues related to the distribution of vibration along the basilar membrane. To assess this, excitation patterns (Glasberg and Moore, 1990) were calculated for the H and I tones with F0 = 800 Hz and ΔF set to 143 Hz, which was the mean threshold determined in the main experiment for the seven subjects who completed all three adaptive runs when the TEN level/ERB_N was 15 dB below the overall level of the complex tones. The calculated excitation patterns took the effect of the TEN into account. The patterns are shown in the upper panel of Fig. 2 (plotted only over the frequency range where there were differences between the patterns for the H and I tones). The largest difference between the excitation patterns was 0.6 dB, and this occurred over a restricted region of the patterns. To assess whether subjects could have detected such a small difference, a supplementary experiment was performed, in which the task was to detect an increase in level of the component at 8800 Hz in the H tone, which was the component associated with the largest change in excitation level between the H and I tones. All other components were fixed in level.

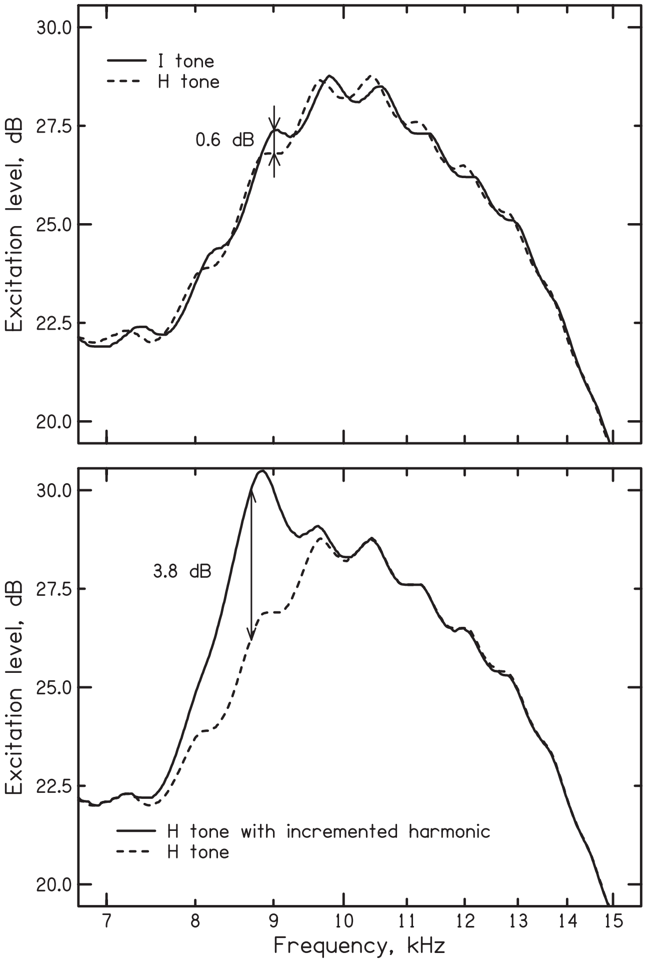


FIG. 2. The top panel shows excitation patterns (Glasberg and Moore, 1990) for a harmonic tone and an inharmonic tone with $\Delta F = 143$ Hz. The effect of the background noise is included. The signal level was the mean level used for the subjects. The lower panel shows the effect on the excitation pattern of incrementing the level of the 8800-Hz component in the H tone by 5.3 dB.

B. Procedure and subjects

The adaptive procedure was the same as described above for discrimination of the H and I tones. In one interval, there were four bursts of the H tone with F0=800 Hz. The H tone had exactly the same timing, spectral characteristics, and level as for the main experiment. In the other interval, the H tone alternated with a tone H(+), in which the component at 8800 Hz was incremented in level, giving the pattern HH(+)HH(+). The task was to identify the interval in which the alternation of level occurred. The starting increment in level, ΔL, was typically 9 dB, which allowed good performance for most subjects. The value of ΔL was changed by 3 dB until one reversal occurred, then by 2 dB until one more reversal occurred, and by 1 dB for the final six reversals. The threshold was taken as the mean increment in level at the last six reversals.

The same background TEN was used as in the main experiment, and its level/ERB_N was 15 dB below the overall level of the complex tone. The seven subjects who completed all three adaptive runs in the first experiment were tested and each subject completed three adaptive runs. The mean change in level of the 8800-Hz harmonic at threshold was 5.3 dB. The range across subjects was 3.1–11.6 dB. The mean threshold is higher than the mean threshold of about 3 dB reported by Moore *et al.* (1984) for changes in level of the 11th harmonic of a complex tone with F0=200 Hz. The difference may result from an effect of center frequency, since detection of an increment in a spectral profile tends to be worse at high frequencies than at low to medium frequencies (Moore *et al.*, 1989).

The change in excitation pattern produced by a 5.3 dB change in level of the harmonic at 8800 Hz is shown in the lower panel of Fig. 2. The maximum change was 3.8 dB and a change larger than 3 dB occurred over a relatively large frequency range (8.5–9.0 kHz). Thus, the threshold change in excitation level (3.8 dB) was about a factor of 6 larger than the maximum difference in excitation level (0.6 dB) produced by the change from the H tone to the I tone when ΔF was at its threshold value. It is possible that the auditory filters at high frequencies are sharper than assumed in our calculation of excitation patterns (Oxenham and Shera, 2003), in which case the differences in excitation patterns for the H and I tones would be larger, but it seems unlikely that the discrepancy would be large enough to account for the factor of 5 difference noted above.

We conclude that the ability to discriminate between the H and I tones in the main experiment was probably not based on the use of excitation-pattern cues.

IV. SUPPLEMENTARY EXPERIMENT 2: ASSESSING THE POSSIBLE ROLE OF PARTIALLY RESOLVED COMPONENTS

A. Rationale

Although the results of the first supplementary experiment suggest that performance in the main experiment was not based on the use of excitation-pattern cues, we decided to conduct a further experiment to check on the possibility that performance might have been based on the frequency discrimination of partially resolved harmonics. It is usually assumed that, for complex tones with equal-amplitude harmonics, only the lowest five to eight harmonics are resolvable (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Moore *et al.*, 2006). However, Bernstein and Oxenham (2003) proposed a somewhat higher limit, based on an experiment in which the “target” harmonic in a complex tone was pulsed on and off. The method of Bernstein and Oxenham (2003) has been criticized (Hartmann and Goupell, 2006; Moore *et al.*, 2009), but nevertheless it seemed possible that one or two of the components in the complex tones in the main experiment were partially resolved, and that discrimination of the H and I tones was based on the discrimination of these partially resolved components.

To check on this possibility, we measured thresholds for detecting a change in frequency of an isolated sinusoid with mean frequency equal to the center frequency of the passband used for the H and I tones. It is known that the threshold for discriminating the frequency of a harmonic presented within a complex tone is larger than the threshold for discriminating the frequency of an isolated sinusoid of the same frequency as the harmonic (Moore *et al.*, 1984; Gockel *et al.*, 2007). For harmonics above the fifth, the threshold is typically a factor of 5 or more higher for a harmonic within a complex tone than for the harmonic in isolation (Moore *et al.*, 1984; Gockel *et al.*, 2007). Hence, if thresholds in the main experiment (discrimination of the H and I tones) were not markedly larger than thresholds for discriminating the frequency of the isolated sinusoid, this would provide strong evidence that performance in the main experiment was not based on the discrimination of partially resolved components.

B. Subjects and method

The subjects who completed two or three adaptive runs using F0=1000 Hz, with the filter centered at 14 000 Hz, were also used in this supplementary experiment. The task was to detect a change in frequency of a single sinusoid with a reference frequency of 14 000 Hz. The procedure was very similar to that used to measure discrimination of the H and I tones. In one interval, four sinusoidal tones bursts with a frequency of 14 000 Hz were presented. In the other interval, the frequency of the four tones alternated between 14 000 Hz and 14 000+ΔF Hz. The value of ΔF was adapted in the same way as for the main experiment. The value of ΔF at threshold is denoted ΔF_{sin}. The sinusoid was presented at 20 dB SL and no background noise was present.

It is known that, at high frequencies, changes in frequency can result in changes in loudness that might be used as a cue for frequency discrimination (Henning, 1966). To reduce the salience of loudness as a cue, frequency discrimination thresholds were measured with the level randomly varied (roved) over a range of ±3 and ±6 dB. The roving was applied to each and every tone; thus the level varied randomly across the four tone bursts within an interval. A very large rove in level was not used because large changes in level can result in changes in pitch (Terhardt, 1974; Verschuure and van Meeteren, 1975; Emmerich *et al.*, 1989; Moore and Glasberg, 1989) and also because a large rove range would have resulted in some of the stimuli falling below absolute threshold. For comparison, frequency discrimination thresholds were also measured with no roving. Three adaptive runs were obtained for each subject for each condition.

C. Results

The geometric mean thresholds (ΔF_{sin}) across the five subjects were 185 Hz with no roving of level, and 226 and 268 Hz with ±3 and ±6 dB of roving, respectively. The geometric mean threshold value of ΔF for the same five subjects for discrimination of the H and I tones in the main experiment was 184 Hz. Thus, discrimination of the H and I tones was comparable to the discrimination of the isolated sinusoid, even when no level roving was applied to the sinusoid. With level roving, discrimination of the isolated sinusoid was slightly worse than discrimination of the H and I tones. However, a one-way within-subjects analysis of variance (ANOVA) conducted on the logarithms of the values of ΔF for discrimination of the H and I tones and on the logarithms of the values of ΔF_{sin} for the three conditions of this supplementary experiment did not give a significant effect of condition: $F(3,12)=1.55$, $p=0.251$.

If discrimination of the H and I tones in the main experiment were based on the discrimination of partially resolved components, thresholds for frequency discrimination of the isolated sinusoid should have been markedly lower than thresholds for discriminating the H and I tones. Clearly, this was not the case. Overall, the results strongly suggest that discrimination of the H and I tones in the main experiment was not based on the discrimination of partially resolved components. Consistent with this, good performance in the main experiment was not associated with good performance in discrimination of the isolated sinusoid. One subject gave values of ΔF_{sin} that were consistently about a factor of 2 larger than the threshold value of ΔF for discrimination of the H and I tones, while another subject showed the opposite pattern. These individual differences account for the lack of significant effect of condition in the ANOVA.

V. DISCUSSION

The results of the main experiment showed that most subjects could discriminate the H and I tones when F0 was 800 Hz and the bandpass filter was centered at 11 200 Hz. Several subjects could discriminate the tones when F0 was as high as 1000 Hz and the bandpass filter was centered at 14 000 Hz. The results of the supplementary experiments indicate that the discrimination of the H and I tones in the main experiment was probably not based on differences in excitation patterns between the H and I tones or on the frequency discrimination of partially resolved components. Given that no envelope cues were available to allow discrimination of the H and I tones, it seems very likely that discrimination was based on the use of TFS. Therefore, our results suggest that human listeners are sensitive to TFS in complex tones up to higher frequencies than believed hitherto. Although our task did not require the subjects to discriminate the H and I tones on the basis of their pitch, all subjects reported that they performed the task by listening for the fluctuation in pitch that was associated with the interval containing the alternating HIHI tones.

As described in Sec. I, several aspects of the perception of sinusoids change as the frequency is increased above 5000 Hz; for a review see Moore (2003). One way to interpret these changes is to assume that the perception of the pitch of sinusoids depends on a temporal code (phase locking) for frequencies below about 5000 Hz and on a place code for higher frequencies. However, an alternative possibility is that pitch is derived from residual temporal information for frequencies above 5000 Hz (Goldstein and Srulovicz, 1977; Heinz *et al.*, 2001), and that the changes in perception simply result from that information become less precise as the frequency is increased above 5000 Hz.

If the frequency discrimination of sinusoids depended on the use of temporal information for low and medium frequencies, and on the use of "place" information for high frequencies, then one might expect that the Weber fraction for frequency discrimination would increase over the frequency range where there was a transition between use of the two types of information, but would then reach a plateau above a certain frequency. The plateau is expected since the relative bandwidth of the auditory filters (bandwidth as a proportion of center frequency) is almost invariant with frequency at high frequencies (Glasberg and Moore, 1990), and since the sharpness (Q value) of neural tuning curves is approximately constant at high frequencies (Palmer, 1995). In fact, the Weber fraction for frequency discrimination tends to increase progressively with increasing frequency above 4000–5000 Hz (Henning, 1966; Moore, 1973; Emmerich *et al.*, 1989; Sęk and Moore, 1995; Rose and Moore, 2005), although the data are sparse for frequencies above 8000 Hz. The data from the present study showed a Weber fraction for frequency discrimination of about 0.014 for frequency discrimination of a 14000-Hz sinusoid without roving of level. This is comparable to the Weber fraction typically reported for a frequency of 8000 Hz, but our task was easier than the 2AFC task (with one tone in each interval) that is typically used to measure frequency discrimination. The threshold would have been higher if we had measured the Weber fraction using the more typical method. Overall, the results are consistent with the idea that the frequency discrimination of pure tones with frequencies above 5000 Hz depends on the use of residual phase-locking information rather than on the use of place information (Goldstein and Srulovicz, 1977; Heinz *et al.*, 2001).

A possible reason why residual phase-locking information may be used more effectively for the discrimination of complex tones than for the discrimination of pure tones is related to refractory effects in auditory neurons. Generally, when a single neuron has generated a spike, it cannot "fire" again for a certain time (the absolute refractory period, typically about 1 ms), and it has a reduced probability of firing for a somewhat longer time (the relative refractory period) (Kiang *et al.*, 1965). When a tone with very high frequency is presented, say, at 10 000 Hz, the shortest time between successive spikes will correspond to many periods of the sound, but the exact number of elapsed periods will vary randomly from one inter-spike interval to the next. For example, ignoring the effects of jitter in the exact timing of the nerve spikes, successive inter-spike intervals might be 1.5, 1.8, 2.7, 1.2, and 3.5 ms; any integer multiples of 0.1 ms above about 1 ms might be present. No doubt information can be combined across neurons, but even then there would be a dominance of intervals that were very long compared with the 0.1-ms period of the sound. Given that there would actually be significant jitter in the spike timing, it might be very difficult for the auditory system to estimate the period of the sound from the inter-spike intervals.

Consider now the case of discrimination of H and I tones, as in the main experiment. We take as an example the condition in which F0=800 Hz, with the bandpass filter centered at 11 200 Hz. Assume that the subject monitors an auditory filter centered at 10 000 Hz. For the H tone, the most prominent time intervals between peaks in the TFS close to adjacent envelope maxima are 1.15, 1.25, and 1.35 ms. Assume that, for the I tone, ΔF is 200 Hz (which is a little larger than the mean threshold found in the experiment). The most prominent time intervals between peaks in the TFS close to adjacent envelope maxima are now 1.125, 1.225, and 1.325 ms. The representation of these intervals would also be affected by refractory effects, but the time intervals to be discriminated (e.g., 1.25 versus 1.125 ms), and low-numbered integer multiples of them (e.g., 2.5 versus 2.25 ms), would be represented in inter-spike intervals. Thus, discrimination of the H and I tones might be easier than discrimination of a high-frequency sinusoid because the ambiguity of the information in the inter-spike intervals is smaller for the former and inter-spike intervals corresponding to the TFS intervals to be discriminated are actually present. It should be noted that the time intervals to be discriminated, 1.25 and 1.225 ms in the above example, differ by about 2%. In other words, the Weber fraction is about 0.02. This is only a little larger than measured for lower center frequencies using the same task (Moore and Sęk, 2009).

Some other research has led to the suggestion that TFS information might be used for frequencies above 5000 Hz. For example, Alves-Pinto and Lopez-Poveda (2008) suggested that TFS information may be used in the discrimination of spectral shape (specifically, detection of a spectral notch in broadband noise) at high frequencies. Hopkins (2009) suggested that TFS information at high frequencies is helpful in identifying the speech of a target talker in the presence of a background talker. The extent to which TFS information at frequencies above 5000 Hz is useful in other tasks remains to be determined. Further research in this area is clearly needed.

VI. SUMMARY AND CONCLUSIONS

In the main experiment, subjects were required to discriminate harmonic (H) and inharmonic (I, frequency shifted) tones that were passed through a fixed bandpass filter centered at 14F0. The H and I tones had different TFS but had the same envelope repetition rate. Further, the exact envelope shape and modulation depth varied randomly from one tone to the next, depending on the random selection of starting phases. Thus, there were no cues in the envelope that would support discrimination. All of nine normally hearing subjects were able to perform the task at above-chance levels when F0 was 800 Hz and the lowest audible component fell close to 8000 Hz. Six out of eight normally hearing subjects could perform the task at above-chance levels when F0 was 1000 Hz and the lowest audible component fell close to 10 000 Hz.

The results of supplementary experiment 1 showed that the largest difference in excitation patterns between the H and I tones, when the frequency shift was set to the threshold value, was markedly smaller than the smallest detectable change in excitation level, as determined using a task in which subjects had to detect a change in level of a single harmonic in the H tone. This suggests that subjects did not use changes in excitation level over a limited frequency region to discriminate the H and I tones in the main experiment.

The results of supplementary experiment 2 showed that the threshold for detecting a change in frequency of an isolated sinusoid with frequency equal to the center frequency of the passband was not markedly smaller than the threshold frequency shift measured in the main experiment. This suggests that discrimination of the H and I tones in the main experiment was not based on the frequency discrimination of partially resolved components.

Taken together, the results suggest sensitivity to the TFS of complex tones up to at least 8000 Hz, which is higher than the limit that is usually assumed for the use of TFS information.

ACKNOWLEDGMENTS

This work was supported by Deafness Research (UK) and the Medical Research Council (UK). We thank Tom Baer, Alain de Cheveigné, Hedwig Gockel, and Kathryn Hopkins for helpful comments on an earlier version of this paper. We also thank Christophe Micheyl and Fan-Gang Zeng for helpful reviews.

Alves-Pinto, A., and Lopez-Poveda, E. A. (**2008**). “Psychophysical assessment of the level-dependent representation of high-frequency spectral notches in the peripheral auditory system,” J. Acoust. Soc. Am. **124**, 409–421.

Attneave, F., and Olson, R. K. (**1971**). “Pitch as a medium: A new approach to psychophysical scaling,” Am. J. Psychol. **84**, 147–166.

Bernstein, J. G., and Oxenham, A. J. (**2003**). “Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?,” J. Acoust. Soc. Am. **113**, 3323–3334.

de Boer, E. (**1956**). “Pitch of inharmonic signals,” Nature (London) **178**, 535–536.

Emmerich, D. S., Ellermeier, W., and Butensky, B. (**1989**). “A reexamination of the frequency discrimination of random-amplitude tones, and a test of Henning’s modified energy-detector model,” J. Acoust. Soc. Am. **85**, 1653–1659.

Glasberg, B. R., and Moore, B. C. J. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hear. Res. **47**, 103–138.

Gockel, H., Moore, B. C. J., Carlyon, R. P., and Plack, C. J. (**2007**). “Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency,” J. Acoust. Soc. Am. **121**, 373–382.

Goldstein, J. L., and Srulovicz, P. (**1977**). “Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,” in *Psychophysics* *and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Hartmann, W. M., and Goupell, M. J. (**2006**). "Enhancing and unmasking the harmonics of a complex tone," J. Acoust. Soc. Am. **120**, 2142–2157.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.

Henning, G. B. (**1966**). "Frequency discrimination of random amplitude tones," J. Acoust. Soc. Am. **39**, 336–339.

Hopkins, K. (**2009**). "The role of temporal fine structure information in the perception of complex sounds for normal-hearing and hearing-impaired subjects," Ph.D. thesis, University of Cambridge, Cambridge, England.

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, England).

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego.

Moore, B. C. J., and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," J. Acoust. Soc. Am. **86**, 1722–1732.

Moore, B. C. J., Glasberg, B. R., and Jepsen, M. L. (**2009**). "Effects of pulsing of the target tone on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. (to be published).

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.

Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820–836.

Moore, B. C. J., and Sęk, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. (to be published).

Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.

Ohgushi, K., and Hatoh, T. (**1991**). "Perception of the musical pitch of high frequency tones," in *Ninth International Symposium on Hearing: Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).

Oxenham, A. J., and Keebler, M. V. (**2008**). "Complex pitch perception above the "existence region" of pitch," in *ARO 31st Mid-Winter Research Meeting* (ARO, Phoenix, AZ).

Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.

Palmer, A. R. (**1995**). "Neural signal processing," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner haircells," Hear. Res. **24**, 1–15.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am. **43**, 764–767.

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 auditory-nerve fibers," J. Neurophysiol. **93**, 3615–3634.

Ritsma, R. J. (**1962**). "Existence region of the tonal residue. I," J. Acoust. Soc. Am. **34**, 1224–1229.

Ritsma, R. J. (**1963**). "Existence region of the tonal residue. II," J. Acoust. Soc. Am. **35**, 1241–1245.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.

Rose, M. M., and Moore, B. C. J. (**2005**). "The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects," Hear. Res. **204**, 16–28.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.

Sęk, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479–2486.

Terhardt, E. (**1974**). "Pitch of pure tones: its relation to intensity," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin).

Verschuure, J., and van Meeteren, A. A. (**1975**). "The effect of intensity on pitch," Acustica **32**, 33–44.

Ward, W. D. (**1954**). "Subjective musical pitch," J. Acoust. Soc. Am. **26**, 369–380.

Effects of pulsing of the target tone on the audibility of partials in inharmonic complex tones

Brian C. J. Moore[a] and Brian R. Glasberg
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Morten L. Jepsen
Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

(Received 27 June 2008; revised 9 March 2009; accepted 10 March 2009)

The audibility of partials was measured for complex tones with partials uniformly spaced on an $\mathrm{ERB_N}$-number scale. On each trial, subjects heard a sinusoidal "probe" followed by a complex tone. The probe was mistuned downwards or upwards (at random) by 3% or 4.5% from the frequency of one randomly selected partial in the complex (the "target"). The subject indicated whether the target was higher or lower in frequency than the probe. The probe and the target were pulsed on and off and the ramp times and inter-pulse intervals were systematically varied. Performance was better for longer ramp times and longer inter-pulse intervals. In a second experiment, the ability to detect which of two complex tones contained a pulsed partial was measured. The pattern of results was similar to that for experiment 1. A model of auditory processing including an adaptation stage was able to account for the general pattern of the results of experiment 2. The results suggest that the improvement in ability to hear out a partial in a complex tone produced by pulsing that partial is partly mediated by a release from adaptation produced by the pulsing, and does not result solely from reduction of perceptual confusion.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3109997]

PACS number(s): 43.66.Fe, 43.66.Hg, 43.66.Mk [RLF] Pages: 3194–3204

I. INTRODUCTION

If attention is directed in an appropriate way, human listeners have some ability to "hear out" individual partials in complex tones (Helmholtz, 1954). Partials that can be heard out in this way are referred to as "resolved." It is often assumed that the ability to hear out partials depends at least partly on the sharpness of the auditory filters (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Moore *et al.*, 2006c), although factors such as the musical experience of the listeners (Soderquist, 1970; Fine and Moore, 1993) and phase locking (Moore and Ohgushi, 1993; Hartmann *et al.*, 1990; Hartmann and Doty, 1996; Moore *et al.*, 2006c) may also play a role. It has been proposed (Moore and Ohgushi, 1993; Moore, 2003) that, for a complex tone with equal-amplitude partials, each "inner" partial (i.e., excluding the highest and lowest partials) can be heard out with 75% accuracy when it is separated from neighboring partials by about 1.25 $\mathrm{ERB_N}$, where $\mathrm{ERB_N}$ refers to the equivalent rectangular bandwidth of the auditory filter as determined for young normally hearing listeners at moderate sound levels (Glasberg and Moore, 1990; Moore, 2003). This is consistent with the finding of several studies that, for harmonic complex tones with fundamental frequencies in the range 100–400 Hz, harmonics with numbers up to about 5–8 can be heard out (Plomp, 1964; Plomp and Mimpen, 1968; Fine and Moore, 1993).

Recently, Bernstein and Oxenham (2003, 2006, 2008) suggested that the ability to hear out harmonics extends to harmonics with numbers up to about 10. They used a two-alternative forced-choice task similar to that employed by Roberts and Bregman (1991) and Moore and Ohgushi (1993). On each trial, subjects had to indicate whether a sinusoidal "probe" was higher or lower in frequency than the nearest harmonic in the complex tone that was presented after the probe; this is referred to as the "target" harmonic. Bernstein and Oxenham (2003, 2006, 2008) suggested that the results of previous studies using this or similar methods might have been influenced by cognitive or attentional factors. For example, subjects may have difficulty in deciding which harmonic in the complex to compare with the probe. Subjects may also have difficulty in overcoming perceptual fusion of the harmonics in the complex caused by their harmonicity and by their synchronous gating (Moore *et al.*, 1986). For brevity, we will refer to all such effects as "confusion" effects. Musically trained listeners may be less susceptible to such confusion effects, which could explain why they perform better than non-musicians in tasks requiring them to hear out partials (Soderquist, 1970; Fine and Moore, 1993). To overcome these difficulties, in the experiment of Bernstein and Oxenham (2003) both the probe and the target harmonic were pulsed on and off. This was done "to help overcome any nonperipheral limitations and to encourage

[a] Author to whom correspondence should be addressed. Electronic mail: bcjm@cam.ac.uk

perceptual segregation, while not affecting peripheral resolvability" (p. 3325). Performance on this task was very good (>90%) for the lower harmonics, but decreased with increasing harmonic number, reaching 75% correct for about the tenth harmonic, for fundamental frequencies of 100 and 200 Hz. Thus, the upper limit found by Bernstein and Oxenham (2003) was markedly higher than found in earlier studies, in which the target harmonic was not pulsed.

Bernstein and Oxenham (2003) assumed that the pulsing would not affect peripheral factors. However, it is not clear that this assumption is valid (Hartmann and Goupell, 2006; Moore *et al.*, 2006c). When a tone is presented continuously, the response of single neurons in the auditory nerve to that tone declines over time, an effect called adaptation (Smith, 1977). For a continuous harmonic complex tone, the responses to all harmonics would show adaptation. However, when a tone is turned off and then back on again, the firing rate of single neurons often shows an initial peak when the tone is turned back on, presumably as a consequence of recovery from adaptation (Smith, 1977). Similarly, if one partial in a complex tone is turned on after the remaining partials, the discharge rate of neurons tuned to the frequency of the delayed-onset partial shows a distinct increase relative to the rate obtained when all partials are turned on at the same time (Palmer *et al.*, 1995). Palmer *et al.* (1995, p. 1787) concluded that "the action of adaptation of the discharge of auditory-nerve fibers can increase the spectral contrast of an introduced component." It is possible therefore that the ability of subjects to hear out the target harmonic in the experiments of Bernstein and Oxenham (2003, 2006, 2008) was enhanced by a recovery from adaptation produced by pulsing the target harmonic on and off.

Hartmann and Goupell (2006) studied the ability of subjects to make pitch matches to a pulsed harmonic in a complex tone. They found that the highest harmonic for which a match could be made reliably depended on the relative phases of the harmonics. When subjects were allowed to listen to harmonic complexes with different starting phases of the harmonics, they were able to select a "favorable" set of phases that allowed a selected (pulsed) harmonic to be heard out more easily and matched more accurately than for other phase selections. For favorable phase selections, subjects could make reliable pitch matches for harmonics up to the 20th. Hartmann and Goupell (2006) argued that harmonics above the tenth would not be spectrally resolved, and that the pulsing enabled pitch matching to harmonics that would not normally be resolved.

In the present experiment, we assessed the possible role of recovery from adaptation by using a similar task to that of Bernstein and Oxenham (2003, 2006, 2008), but exploring the effect of varying the interval between pulses and the rise/fall time of the pulses. If the pulsing produces a release from adaptation that makes the task easier, then shortening the interval between pulses or decreasing the rise/fall time of the pulses should lead to a worsening in performance, because both of these manipulations would lead to a reduced release from adaptation (Smith, 1977; Smith, 1979; Westerman and Smith, 1984).

If pulsing the target mainly influences confusion effects, for example, by helping the subject to determine which harmonic to attend to, then changing the interval between pulses or changing the rise/fall time of the pulses might also have some influence on the results, by making the pulsed component more or less salient. Bernstein and Oxenham (2003) did not discuss this possibility, implicitly assuming that the interval between pulses and the rise/fall time of the pulses used by them were sufficient to remove any effects of confusion. If resolution of confusion is the sole effect of pulsing the target, then one would expect performance to reach a plateau if the inter-pulse interval is sufficiently long. For closely spaced partials, for which peripheral resolvability should limit performance, this plateau should occur for a performance level well below 100%. On the other hand, if part of the effect of pulsing of the target is to produce a release from adaptation, one might expect to see a progressive improvement in performance with increasing inter-pulse interval, even when the partials are closely spaced. To test which of these predictions was closer to the truth, we included inter-pulse intervals longer than those employed by Bernstein and Oxenham (2003, 2006, 2008).

As a further way of assessing the relative importance of peripheral adaptation as opposed to central confusion, we used two approaches. First, we conducted a second experiment, in which we measured the ability to detect which of two complex tones contained a pulsed partial, varying the interval between pulses and changing the rise/fall time of the pulses in the same way as for experiment 1. We argued that central factors should play a much smaller role in this case, as the stimuli were chosen so that there was minimal uncertainty about which component in the complex tone to attend to. If the pattern of results was similar for the two experiments, this would support the idea that peripheral factors such as release from adaptation were the main cause of the effects of varying inter-pulse interval and rise/fall time. As a further test of this idea, we used the stimuli as input to a model of auditory processing that includes an adaptation stage (Dau *et al.*, 1996a), to explore whether the pattern of results could be predicted by the internal representation (IR) produced by the model.

In contrast to Bernstein and Oxenham (2003) and Hartmann and Goupell (2006), who used harmonic complex tones, we used complex tones with partials uniformly spaced on an ERB$_N$-number scale (see below for details), which were therefore inharmonic. This was done for three reasons. First, it reduced the tendency for the components to fuse based on their harmonicity (Moore *et al.*, 1986). Second, it avoided the possibility that subjects might infer which harmonic in a complex tone was closest in frequency to the probe without actually hearing out the target partial, based on a conscious or implicit knowledge of the frequencies of tones that form a harmonic series. Finally, the waveform of the inharmonic tones was not periodic, so it was unlikely that the results would be influenced by the specific set of random starting phases chosen for the partials.

TABLE I. Frequencies (Hz) of the partials in the complex tones for each spacing used. The frequency of the middle partial is given in bold type.

	Partial number										
Spacing	1	2	3	4	5	6	7	8	9	10	11
1.5*E*	408	520	652	806	988	**1201**	1452	1747	2094	2501	2980
1.0*E*	605	700	806	924	1055	**1201**	1364	1545	1747	1972	2222
0.75*E*	726	806	893	988	1090	**1201**	1322	1452	1594	1747	1913

II. EXPERIMENT 1: EFFECT OF PULSING ON THE ABILITY TO HEAR OUT PARTIALS

A. General method

The method was similar to that used by Roberts and Bregman (1991), Moore and Ohgushi (1993), Bernstein and Oxenham (2003), and Moore *et al.* (2006c). On each trial subjects were presented with a sinusoidal tone followed by a complex tone. The sinusoid will be referred to as the probe. Subjects were told that the probe was close in frequency to one of the partials in the complex tone (the target), but was actually slightly higher or lower in frequency than the target. On half the trials, chosen at random, the probe was higher in frequency than the target by Δf, and on the other half it was lower by Δf. Subjects were asked to indicate, by pressing the appropriate button on the response box, whether the "closest" partial in the complex was higher or lower in frequency than the probe. Correct-answer feedback was provided after each trial by lights on the response box. The partial that was "probed" was varied randomly from trial to trial. The frequencies of all partials in the complex tone were randomly varied (roved) from trial to trial by multiplying them by a factor randomly chosen within the range 0.9–1.1, while keeping the frequency ratios between partials fixed. The frequency of the probe was multiplied by the same factor.

Before testing started, subjects were given training, starting with easy conditions and working toward more difficult conditions. Training started with a "complex" tone containing a single sinusoid with a nominal frequency of 1000 Hz. In this case, the task was a simple frequency discrimination task (but with roving, as described above). Then, the number of partials in the complex tone was increased to 2, with widely separated frequencies. When subjects scored better than 90% with this complex, the number of partials was increased to 3 with widely spaced frequencies, and then to 5. Some subjects who found the task to be easy skipped the training with intermediate numbers of partials. Subjects were then given training runs with the complex tones to be used in the main experiment that contained 11 partials. After this training, performance appeared to remain largely stable.

In the main experiment, each partial in a complex was probed ten times in a given run, five times with the probe lower in frequency than the relevant partial in the complex, and five times with it higher. Five runs were obtained for each complex tone, giving a total of 50 judgments for each partial. In a few cases, when scores for the first few runs were close to 100% for all partials, testing stopped after fewer than five runs.

B. Stimuli

The partials in each complex tone were equally spaced on an $\mathrm{ERB_N}$-number scale. The relationship between $\mathrm{ERB_N}$-number, E, and frequency, f (Hz), was assumed to be as suggested by Glasberg and Moore (1990):

$$E = 21.4 \log_{10}(0.00437f + 1). \qquad (1)$$

The spacings used were 0.75E, 1.0E, and 1.5E. The mean frequency of the central partial was always 1201 Hz, corresponding to $E=17$. The mean frequencies of all partials for each spacing used are given in Table I. All of the complex tones contained 11 partials. The values of Δf were 3% of the frequency of the target partial for the spacing of 0.75E and 4.5% of the frequency of the target partial for the other spacings. The value of Δf was made smaller for the spacing of 0.75E to ensure that the probe tone was always closer in frequency to the target partial than to any other partial. For example, when the ninth partial (1594 Hz) in the complex tone was the target, a value of Δf of +3% led to the probe being 3% higher in frequency than the target partial and 6% lower than the tenth partial (1747 Hz).

The probe tone and all partials in the complex tone each had an overall duration of 1000 ms, with 20-ms raised-cosine rise/fall ramps. The inter-stimulus interval was 300 ms. The non-target partials in the complex were always uninterrupted for the 1000-ms duration. In condition 1, the probe tone and the target partial were also uninterrupted for the 1000-ms duration. This condition is very similar to that used previously by Moore *et al.* (2006c). In the remaining conditions, both the probe tone and the target partial were pulsed off and on, with the same off-on pattern for the probe and target. The following conditions were used, all with raised-cosine ramps (note that these ramp times refer to the ramps *within* the pulsing pattern; the very first and very last ramps always lasted 20 ms):

(2) 10-ms ramps, 0-ms off time, 306.7-ms steady state pulse duration,
(3) 20-ms ramps, 0-ms off time, 293.3-ms steady state pulse duration,
(4) 40-ms ramps, 0-ms off time, 266.7-ms steady state pulse duration,
(5) 20-ms ramps, 50-ms off time, 260-ms steady state pulse duration [as used by Bernstein and Oxenham (2003)], and
(6) 20-ms ramps, 100-ms off time, 226.7-ms steady state pulse duration.

The pulsing of the probe was always relatively easy to

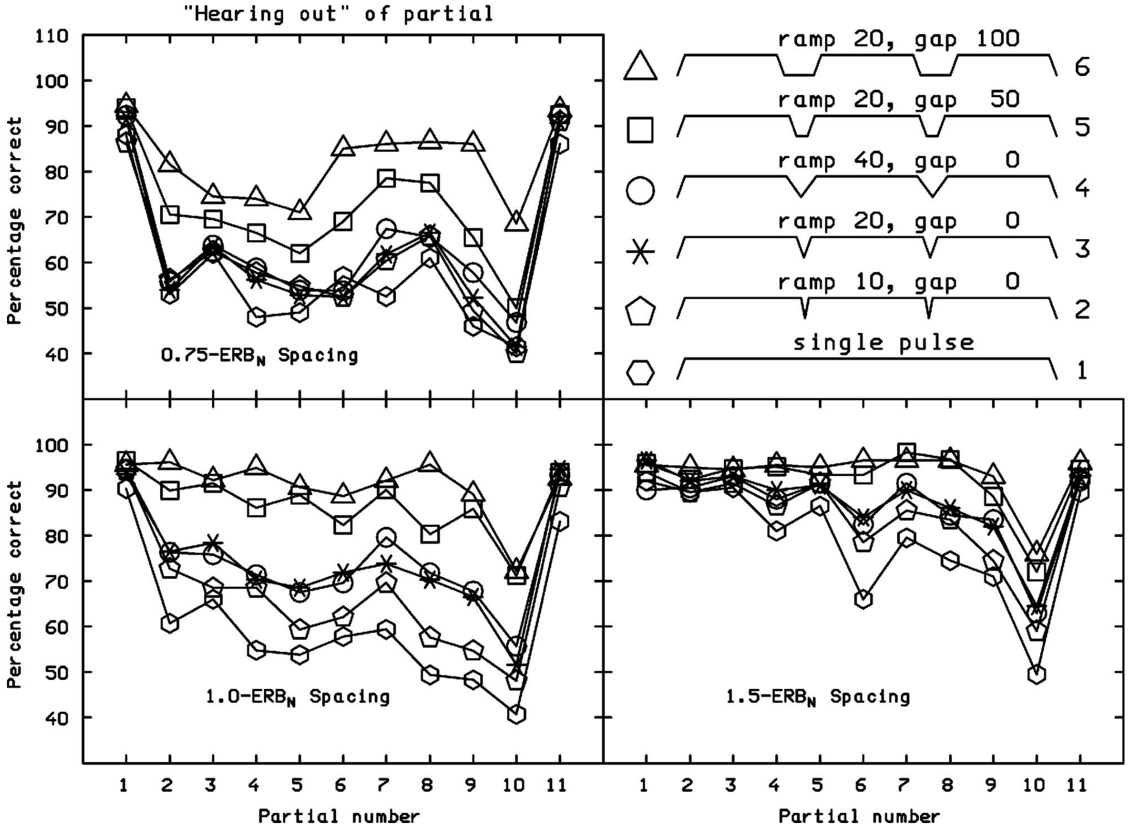


FIG. 1. Scores averaged across subjects for experiment 1. The percentage correct is plotted as a function of partial number (see Table I for frequencies of partials). Within each panel, the parameter is the pulsing pattern of the probe tone and the target tone within the complex (illustrated schematically on the top-right). Each panel shows results for one spacing of the partials, as indicated in the key.

hear, even for condition 2, which used the shortest ramps and off time. The probe tone and each partial in the complex tone all had the same level, which was 50 dB sound pressure level (SPL). The starting phases of the partials in each complex tone were chosen randomly for each trial.

Stimuli were generated digitally on-line using a Tucker-Davis Technologies (TDT) system II. The stimuli were played through a 16-bit digital-to-analog converter (TDT, DD1) at a 50-kHz sampling rate, lowpass filtered at 8 kHz (Kemo VBF8/04), attenuated (TDT, PA4), and presented via a headphone buffer (TDT, HB6), a manual attenuator (Hatfield 2125), and one earpiece of a Sennheiser HD580 headphone, which has a diffuse-field response. Levels specified are equivalent diffuse-field levels. Levels at the eardrum would have been higher for frequencies around 3000 Hz (Moore *et al.*, 1998). Subjects were tested individually in a double-walled sound-attenuating chamber.

C. Subjects

Four subjects (one male, three female) were used, all with no reported history of hearing disorders. One was author BG. Their absolute thresholds were better than 20 dB HL for audiometric frequencies from 250 to 8000 Hz (ISO 389-8, 2004). Their ages ranged from 21 to 62 years and all had some degree of musical training. Musically trained subjects were chosen, since subjects without such training often have difficulty with this task, especially when the probe and target tone are not pulsed (Fine and Moore, 1993). All subjects except author BG were paid for their participation.

D. Results and discussion

Although there were some individual differences, as have also been found in previous similar studies, the general pattern of results was similar across subjects. Mean results are shown in Fig. 1. The scores are plotted as a function of partial number and are averaged across the cases when the probe was lower in frequency than the target and when it was higher in frequency. Each panel shows results for one spacing. The different conditions of pulsing are indicated schematically at the right of the figure. The scores for the lowest and highest (edge) partials were high for all spacings and for all pulsing patterns. These high scores are consistent with other research showing that edge partials are easier to hear out from complex tones than inner partials (Plomp, 1964; Moore, 1973; Moore *et al.*, 1984; Moore and Ohgushi, 1993; Moore *et al.*, 2006c; Gockel *et al.*, 2007).

Scores for the inner partials generally worsened when the partial number was above 8, corresponding to a frequency of approximately 1500–1700 Hz. A similar trend was found by Moore and Ohgushi (1993) and by Moore *et al.* (2006c); the trend seems to be related to absolute frequency rather than to partial number. The trend for worse performance at higher frequencies is consistent with the finding of Plomp (1964) and of Plomp and Mimpen (1968) that the frequency spacing necessary for a given partial to be heard out was greater than a critical band at high frequencies but less than a critical band at low frequencies. This trend may indicate a role of phase locking in the ability to hear out partials, since phase locking becomes less precise for frequencies above about 1500 Hz (Johnson, 1980; Palmer and Russell, 1986).

The scores for the second-highest partial in the complex tone (partial 10) were especially low and for the two smaller spacings were below the chance level of 50% for some conditions, indicating systematic errors. Such a pattern has been observed previously (Moore *et al.*, 2006c). These low scores can be partly explained in terms of the high salience of the pitch of the highest partial. This high salience may be produced by phase locking to that partial in neurons tuned above the frequency of the partial (Moore, 2003). It appears that, when the frequency of the probe was above the frequency of the second-highest partial, the pitch of the probe was often judged relative to that of the highest partial rather than that of the closest partial (the target). Hence, subjects consistently and erroneously responded that the partial in the complex was higher in pitch. It is of interest that a dip in performance for the tenth partial occurred even for the pulsing pattern with the longest inter-pulse interval (condition 6). It appears that the extra salience of the target partial produced by pulsing it on and off was not sufficient to overcome the effect of the intrinsic high salience of the highest partial in the complex.

The results for individual subjects showed systematic irregularities in the pattern of results as a function of partial number. For example, for the 1-ERB$_N$ spacing, one subject showed especially low scores for the sixth partial and another subject showed especially low scores for the fifth partial. Such irregularities have been observed previously, and it has been proposed that they result from irregularities in middle-ear transmission, which may make some partials easier to hear out than others (Moore and Ohgushi, 1993; Moore *et al.*, 2006c).

Performance generally improved with increasing spacing, as expected. Importantly, performance was also strongly influenced by the pattern of pulsing of the target partial. Performance was worst overall when the target was presented as a single long pulse (hexagons) and was best overall for the condition where the inter-pulse interval was 100 ms (triangles). For example, for the 1-ERB$_N$ spacing, scores for partials 2–8 were close to chance for the former and above 90% for the latter.

To illustrate the overall effect of spacing and of the pulsing pattern of the target tone, the data were averaged across all inner partials, except for the tenth (since the latter led to anomalous results, as discussed above). The outcome is shown in Fig. 2. For each pulsing pattern, the scores increased with increasing spacing, and for each spacing the scores were strongly influenced by the pulsing pattern.

To assess the statistical significance of the effects described above, the data were transformed to rationalized arcsine units (RAU; Studebaker, 1985) and then to a within-subjects analysis of variance (ANOVA), with factors type of pulsing (six types), partial number (2–9), and spacing ($0.75E$, $1E$, or $1.5E$). The main effect of type of pulsing was significant: $F(5,15)=67.13$, $p<0.001$. The main effect of partial number was not significant: $F(7,21)=1.96$, $p=0.11$. The main effect of spacing was not significant: $F(2,6)=4.04$, $p=0.077$. The interaction of type of pulsing and spacing was significant: $F(10,30)=2.30$, $p=0.038$, but accounted for only about 2% of the variance in the data. The interaction of partial number and pulsing pattern was also significant: $F(35,105)=1.89$, $p=0.007$, but again accounted for less than 2% of the variance in the data. No other interactions were significant.

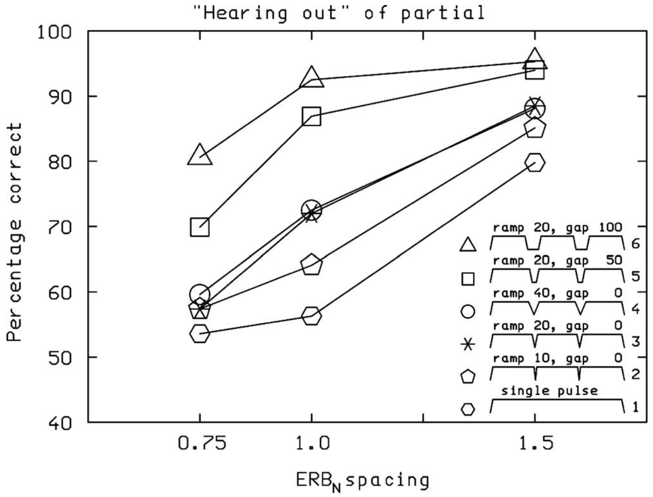


FIG. 2. Scores for experiment 1 averaged across all inner partials and plotted as a function of the spacing of the partials. The parameter is the pulsing pattern of the probe tone and the target tone within the complex, as indicated in the key.

The large effect of the pulse pattern of the probe and target tones indicates that the effect of pulsing is not simply to remove confusion effects in an all-or-none manner. If the effect of the pulsing was mainly to reduce confusion effects, one might expect that performance would initially improve with increasing inter-pulse interval or rise/fall time, and then reach an asymptote when confusion was completely resolved. The data show no sign of an asymptote, except when performance was close to ceiling. For the 0.75-E spacing, performance was close to chance when the target component was not pulsed, as would be expected if the components were completely unresolved for this spacing. However, when the target was pulsed, the mean scores improved from about 67% to 80% when the inter-pulse interval was increased from 50 to 100 ms.

The pattern of results is consistent with an explanation based on recovery from adaptation, as described in Sec. I. According to this explanation, turning the target tone off momentarily leads to a recovery from adaptation, so that the neural response to the target is increased relative to that for adjacent (non-pulsed) partials when the target is turned back on. The amount of recovery would increase as the on-off ramps were made longer and as the inter-pulse interval was increased. However, it is hard to rule out a contribution from resolution of confusion caused by an increase in the salience of the target component as the inter-pulse interval or rise/fall time were made longer.

III. EXPERIMENT 2: DETECTION OF WHICH COMPLEX TONE CONTAINS A PULSED PARTIAL

According to the explanation of the results based on recovery from adaptation, as described above, the pulsing of the target is beneficial because it leads to an enhanced neural response to the target, such that the response of neurons tuned close to the frequency of the target becomes higher than the response of neurons tuned to the frequencies of adjacent partials. If this explanation is correct, then the pulsing should affect the ability to detect the pulsed target in the complex tone in a similar way that it affects the ability to "hear out" the target partial, as measured in experiment 1. Experiment 2 was conducted to test this prediction. The stimuli and task were designed so that there was little uncertainty about what was to be detected. In experiment 1, there was a potential perceptual confusion because the frequency of the probe had to be compared with the frequency of the closest partial in the complex, and it may not have been obvious to the listener which one was the closest. In experiment 2, the probe frequency coincided exactly with the frequency of one of the partials in the complex. The other source of confusion mentioned in Sec. I, namely, perceptual fusion of the partials, should have been small in both experiments, for the following reasons: (1) The stimuli were inharmonic, eliminating the tendency for perceptual fusion caused by harmonicity; (2) the stimuli were relatively long, whereas perceptual fusion caused by synchronous gating is strong mainly for short-duration signals (Moore *et al.*, 1986). Thus the results of experiment 2 should have been minimally affected by confusion effects. We reasoned that, if the pattern of results was similar for experiments 1 and 2, this would support the idea that the effects of pulsing found in experiment 1 were mainly due to recovery from adaptation rather than to resolution of confusion.

A. Stimuli, procedure, and subjects

A two-interval two-alternative forced-choice procedure was used. In each interval a probe tone with 1000-ms overall duration (including 20-ms raised-cosine ramps) was followed by a complex tone. The complex tones were the same as those used in experiment 1, and had the same overall duration of 1000 ms. The frequency of the probe tone was *equal* to the frequency of one of the partials in the complex tone, the target. Thus, there was no uncertainty about which partial in the complex the probe should be compared to. In both intervals of a trial, the probe tone was pulsed on and off with one of the pulse patterns corresponding to conditions 2–6 of experiment 1. In both intervals, all of the components in the complex tone except the target were uninterrupted over the 1000-ms duration. In one interval, selected randomly, the target tone was pulsed on and off with the same pattern as the probe. In the other interval, the target was uninterrupted over the 1000-ms duration. The task of the subject was to identify the interval in which the target was pulsed. The frequency of the probe tone and the target tone was randomly selected on each trial from 1 of the 11 possible frequencies, as indicated in Table I. In this experiment, the probe served to indicate the frequency region in which to listen for the pulsing and also the pulsing pattern to listen for. To make the task as similar as possible to that for experiment 1, the frequencies of all partials in the complex tone were randomly varied (roved) from trial to trial by multiplying them by a factor randomly chosen within the range 0.9–1.1, while keeping the frequency ratios between partials fixed. The frequency of the probe was multiplied by the same factor. The subjects were the same as for experiment 1. Each partial in a complex was probed five times in a given run. Five runs were obtained for each complex tone, giving a total of 25 judgments for each partial.

B. Results and discussion

Again, the general pattern of results was similar across subjects. Mean results are shown in Fig. 3. The scores are plotted as a function of partial number, with pulse pattern as parameter. Each panel shows results for one spacing. Although the results were somewhat "noisy," the overall pattern was similar to that found in experiment 1. This is illustrated in Fig. 4, which shows results averaged across all of the inner partials except the tenth and plotted as a function of spacing, with pulse pattern as a parameter. Scores improved with increasing spacing and performance was strongly affected by the pulse pattern, being worst for condition 2 (10-ms ramps, no gaps) and best for condition 6 (20-ms ramps and 100-ms gaps).

To assess the statistical significance of these effects, the data were subjected to a rationalized arcsine transform and then to a within-subjects ANOVA, with factors type of pulsing (five types), partial number (2–9), and spacing (0.75, 1, or 1.5ERB$_N$). The main effect of type of pulsing was significant: $F(4,12)=47.38$, $p<0.001$. The main effect of partial number was significant: $F(7,21)=2.80$, $p=0.032$. The main effect of spacing was significant: $F(2,6)=21.78$, $p=0.002$. No interactions were significant.

C. Comparison of results for experiments 1 and 2

To compare the results for the two experiments, a within-subjects ANOVA was conducted on the RAU-transformed data with factors task ("hearing out" of partial as in experiment 1, or "detection of pulsing" as in experiment 2), spacing, pulsing type, and partial number (2–9). The data for condition 1 from experiment 1 (where the probe and target were not pulsed) were excluded, as there was no corresponding condition in experiment 2.

The effect of spacing was significant, $F(2,6)=34.01$, $p<0.001$, as were the effects of pulsing pattern, $F(4,12)=90.16$, $p<0.001$, and partial number, $F(7,21)=3.04$, $p=0.023$. The effect of task was not significant: $F(1,3)=3.46$, $p=0.16$. In other words, a given partial could be heard out (experiment 1) with about the same accuracy as the interval containing the pulsed target could be identified in experiment 2. The grand mean scores in RAU were 82.4 for experiment 1 and 89.2 for experiment 2. However, it should be noted that two subjects did show somewhat better performance for experiment 2 than for experiment 1. The other two subjects showed similar performance for the two experi-

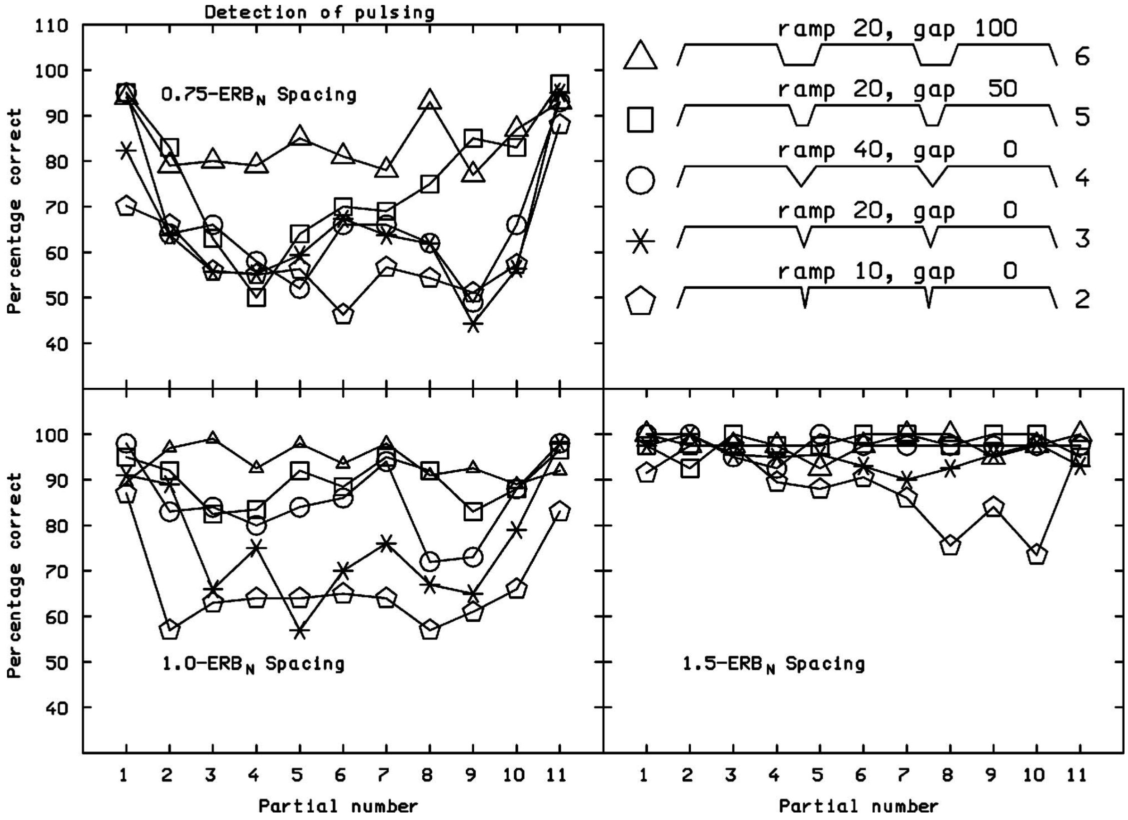


FIG. 3. As Fig. 1, but showing scores for experiment 2, in which subjects had to identify which of two complex tones contained a pulsed partial.

ments. There was a significant interaction between task and pulsing pattern; $F(4,12)=4.2$, $p=0.024$, although this accounted for less than 2% of the variance in the data.

IV. INTERPRETING THE RESULTS USING A MODEL OF AUDITORY PROCESSING

The similar pattern of results for the two experiments could be explained in terms of the increasing perceptual salience of the gaps in the pulsed partial as the gaps were made longer. The increasing salience could draw attention to the pulsed partial, making it both easier to hear out and to detect. This explanation would be consistent with previous data on the detection of gaps or decrements in level of sounds, which have been modeled in terms of the detection of dips in the output of a sliding temporal integrator (Moore *et al.*, 1988; Plack and Moore, 1991; Moore *et al.*, 1993), without taking into account processes associated with adaptation or a release from adaptation. However, the stimuli in those experiments mostly involved gaps or decrements in narrowband sounds or sounds in which the gap or decrement occurred across the whole stimulus spectrum. For multiple-component sounds with closely spaced components in which only one component has gaps, as in the present experiments, the fact that the response to the pulsed component would momentarily exceed the response to adjacent components (as a result of release from adaptation), or exceed the on-going response within the channel tuned to the pulsed component, may be highly perceptually relevant.

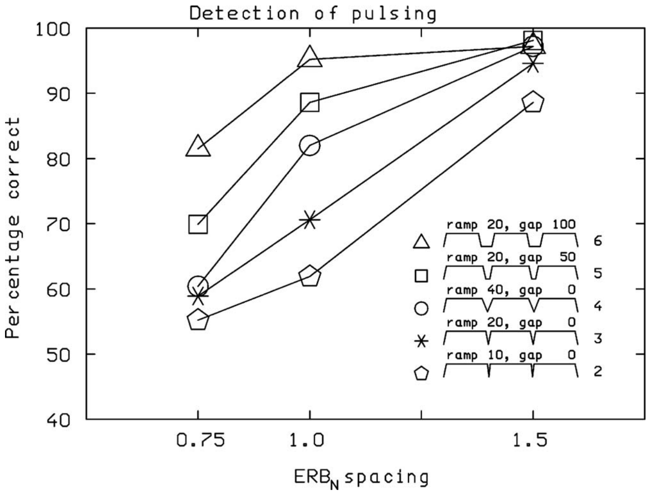


FIG. 4. As Fig. 2, but showing results for experiment 2.

In this section, we use a model of auditory processing to illustrate the effects that adaptation might have on the detection of the pulsed partial, as tested in experiment 2. We use the model described by Dau *et al.* (1996a), which has been shown to account for a broad range of spectral and temporal masking data (Dau *et al.*, 1996b). This model includes a linear auditory filter bank and an adaptation stage, which is important for accounting for temporal masking effects (Dau *et al.*, 1996b). The original version of the model has been modified to include a modulation filter bank (Dau *et al.*, 1997a, 1997b) and a more complex basilar-membrane (BM) stage to simulate cochlear compression (Jepsen *et al.*, 2008). However, we used the original version of the model (Dau *et al.*, 1996a) for simplicity, since it was thought that BM

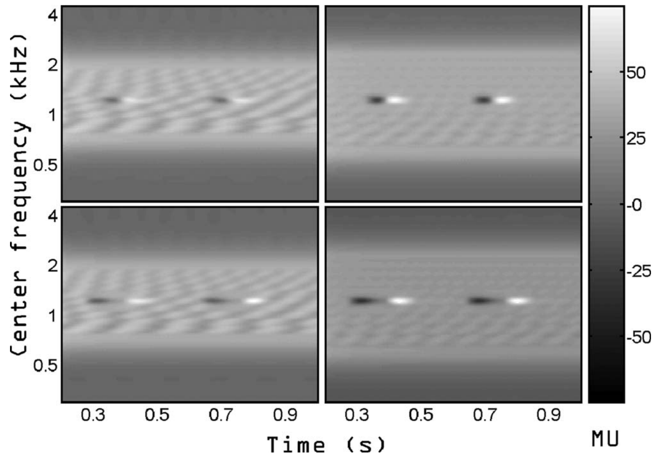


FIG. 5. IR for the complex tones used in four different conditions, with the sixth partial pulsed. Time is plotted on the *x*-axis, channel CF is plotted on the *y*-axis, and darkness-brightness shows excitation in MUs (see scale on the right). The spacing of the partials was either 0.75*E* (left) or 1*E* (right). The two pulsing conditions were 40-ms ramps, 0-ms gaps (top) or 20-ms ramps, 100-ms gaps (bottom).

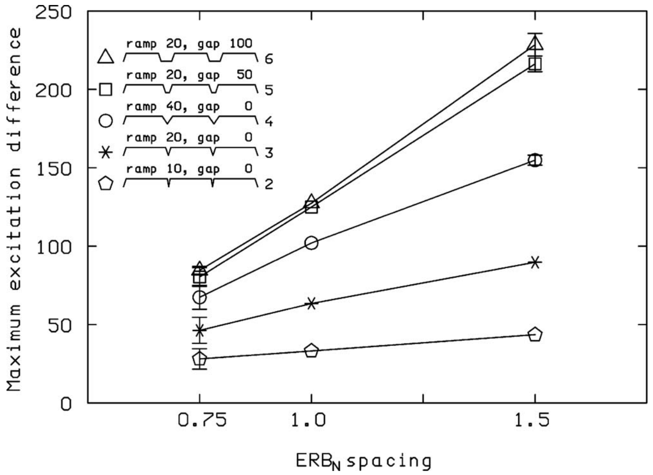


FIG. 6. The MED plotted as a function of ERB$_N$-spacing, with pulsing condition as parameter. The data points are based on the mean values of the MED obtained when partial numbers 3, 6, and 9 were pulsed. Error bars show the standard deviation of the MED across the three partials.

compression and processing in the modulation filter bank would not have a major influence on the detection of a pulsed partial.

The model consists of the following.

(1) An array of linear gammatone filters (Patterson *et al.*, 1995), to simulate BM processing. The bandwidths of the filters had the values suggested by Glasberg and Moore (1990).
(2) Half-wave rectification followed by a lowpass filter with a cut-off frequency at 1 kHz, to simulate the action of the inner hair cells.
(3) A non-linear adaptation stage that consists of five feedback loops with time constants ranging from 5 to 500 ms.
(4) A lowpass filter with a cutoff frequency at 8 Hz. This acts to emphasize energy from low-frequency modulation in the envelope of the signal.

For full details of each stage, the reader is referred to Dau *et al.* (1996a). It should be noted that the model as used here does not take confusion effects into account.

We focus on illustrating the effects of adaptation within the model, by using displays of its internal representation (IR). Hence we disregard the decision mechanism of the original model. The IR display is a spectrogram-like plot of the internal signal excitation in model units (MUs) as a function of filter center frequency (CF) and time. Figure 5 shows IRs for the complex tones used in four different conditions, with the sixth partial pulsed. The contrast of the display has been chosen such that a change in level of about 1 dB would just be visible. The initial part of the IR, from 0 to 0.2 s, is not shown. The spacing of the partials was either 0.75*E* (left) or 1*E* (right). The two pulsing conditions were 40-ms ramps, 0-ms gaps (condition 4, top) or 20-ms ramps, 100-ms gaps (condition 6, bottom). In each case, the channel tuned to the frequency of the pulsed partial shows a decrease in excitation when the partial is turned off, and then a distinct increase in excitation, above the amount of excitation during the steady parts of the sound, just after the partial is turned back on. This increase corresponds to a release from adaptation. The release from adaptation is greater for the wider ERB$_N$ spacing (compare the left and right panels) and is greater for the longer gap duration (compare the top and the bottom panels).

For the 0.75*E* and 1*E* spacings, the IR showed only small across-CF variation when no partial was pulsed; the "ripple" in the IR, which provides an indication of the degree to which the partials were resolved, was about 1.5 MU for the 0.75*E* spacing and 1.9 MU for the 1*E* spacing, while the mean excitation was about 43 MU (the ripple was larger, at about 6 MU for the 1.5*E* spacing). These small variations are consistent with the idea that, for the spacings of 0.75*E* and 1*E*, the partials were barely, if at all, resolved. Pulsing a partial off and on produced more substantial decreases and increases in excitation in the channel tuned to the frequency of the pulsed partial; the increases above the amount of excitation during the steady part of the sound are the result of a release from adaptation. For example, for the condition with 20-ms ramps and 100-ms gaps, the increases in excitation produced by the pulsing relative to the excitation during the steady parts of the stimulus (averaged across channels tuned to the third, sixth, and ninth partials) were 28 MU for the 0.75*E* spacing, 57 MU for the 1*E* spacing, and 107 MU for the 1.5*E* spacing.

To quantify the effect of pulsing of the partial on the IR, we considered only the output of the channel tuned to the frequency of the pulsed partial. This seems reasonable, since in experiment 2 the pulsed partial was presented in isolation before the complex tone, and this would have directed attention to the relevant channel. The first 200 ms of the IR was disregarded, as this was strongly affected by the onset response to all partials. We used as a measure the maximum excitation difference (MED), defined as the difference in excitation in MU between the minimum excitation in the channel (which occurred just after the partial was turned off) and the peak excitation in the channel (which occurred just after the partial was turned back on). Figure 6 shows the MED plotted as a function of ERB$_N$-spacing, with pulsing condi-

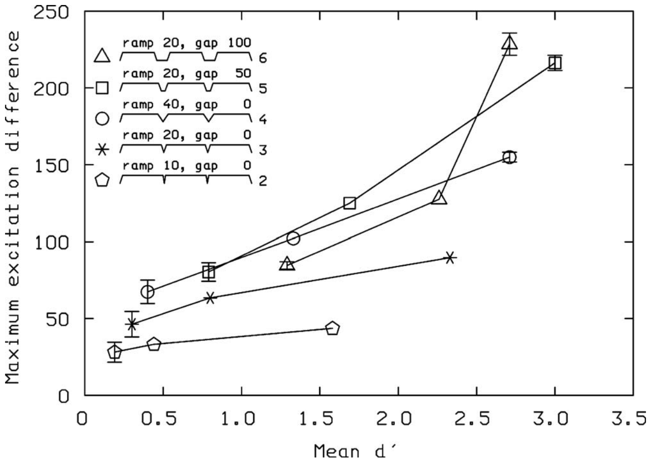


FIG. 7. The MED plotted against the corresponding mean d' value obtained for each pulsing condition and spacing in experiment 2. Error bars show the standard deviation of the MED across partials 3, 6, and 9.

tion as parameter. The data points are based on the mean values of the MED obtained when partial numbers 3, 6, and 9 were pulsed. Error bars show ±1 standard deviation. For each pulsing condition, the MED increases as the spacing of the partials increases. More importantly, the MED increases as the ramp duration and/or the gap duration increases. The ordering of the functions for the different pulsing conditions is the same as for the experimental data plotted in Fig. 4.

Figure 7 shows the MED plotted against the corresponding mean d' value obtained for each pulsing condition and spacing in experiment 2. If performance were perfectly accounted for by the MED, then the functions for the different pulsing conditions would all lie on top of one another. Clearly, this is not the case. In particular, the MEDs for the pulsing conditions with 0-ms gaps and ramps of 10 or 20 ms fall below those for the other pulsing conditions. Hence, the model predicts that the salience of the pulsing in these two conditions would be smaller relative to the other conditions than is actually the case. This may reflect a limitation in the accuracy of the model in simulating adaptation effects. It may also reflect some other aspects of auditory processing that are not simulated by the model, such as suppression. Nevertheless, the modeling clearly illustrates how pulsing a partial on and off can provide cues for identifying the channel corresponding to the pulsed partial, even when the partial is not resolved by the auditory filters.

V. DISCUSSION

We argued in Sec. I that if resolution of confusion were the sole effect of pulsing the target in experiment 1, then one would expect performance to reach a plateau if the inter-pulse interval was sufficiently long. For closely spaced partials, for which peripheral resolvability should limit performance, this plateau should occur for a performance level well below 100%. In fact, in experiment 1, performance increased progressively with increasing inter-pulse interval, and there was no sign of a plateau, even for the closely spaced components. However, a similar pattern of performance was observed in experiment 2, indicating that the ability to detect the pulsing also increased progressively with increasing inter-pulse interval. Since the pulsing was never perfectly detectable at the smallest spacing of the partials, the cue provided by the pulsing in experiment 1 would not have been fully effective in resolving confusion. This makes it difficult to draw any strong conclusions from the lack of a plateau in the data for the smallest spacing.

Experiment 2 was designed so that there was little uncertainty about what was to be detected; the probe tone indicated the frequency and pulsing pattern of the target tone in the complex. Also, the partials in the complex were not harmonically related, so perceptual fusion based on harmonicity could not have played a role. Thus, the results should have been minimally affected by attentional or cognitive factors. The pattern of results for experiment 2 was very similar to that for experiment 1, and, excluding the special case of the tenth (second highest) partial, overall performance was similar for the two experiments. This suggests that the change in performance with pulsing pattern in experiment 1 was not a consequence of the pulse patterns varying in the extent to which they reduced uncertainty about what to listen for. Rather, the results are consistent with the idea that, in both experiment 1 and experiment 2, increasing the rise/fall time or inter-pulse interval of the target tone produced a greater recovery from adaptation, and that this led to improved performance. The results of the modeling described in Sec. IV support this interpretation.

If this interpretation is correct, it means that measurement of the ability to hear out a partial in a complex tone when the target partial is pulsed on and off does not give a valid indication of how well that partial can be heard out in the more common situation when the target tone is not pulsed. Rather, the method using a pulsing target leads to an over-estimate of the ability to hear out a partial. The present results are consistent with earlier results (Moore and Ohgushi, 1993; Moore *et al.*, 2006c) indicating that, in the absence of pulsing, a target partial needs to be separated from neighboring partials by between $1E$ and $1.5E$ to be heard out with 75% accuracy, although a somewhat greater separation is needed as the frequency of the target partial increases above about 1500 Hz, probably because the contribution of phase locking information decreases at high frequencies (Hartmann *et al.*, 1990; Hartmann and Doty, 1996; Moore *et al.*, 2006c). For harmonic complex tones with equal-amplitude (non-pulsed) harmonics, the results imply that only the lowest 5–8 harmonics can be heard out for complex tones with fundamental frequencies in the range 100–400 Hz, as concluded earlier by Plomp (1964) and Plomp and Mimpen (1968). For example, for a fundamental frequency of 100 Hz, the seventh and eighth harmonics are separated by slightly less than $1E$, and so would be barely, if at all, resolved.

This conclusion has important implications for theories of pitch perception. It has been shown in several studies that the ability to discriminate the fundamental frequency of harmonic complex tones is worse when the tones contain only high harmonics than when they contain low harmonics (Hoekstra and Ritsma, 1977; Hoekstra, 1979; Moore and Glasberg, 1988; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994). The transition from good to poor performance as the rank, N, of the lowest harmonic increases, occurs roughly over the range $N=8-14$ (Hoekstra and Ritsma, 1977; Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003, 2008; Moore *et al.*, 2006b). One interpretation of the "transition region" is that it reflects a progressive loss of sensitivity to the temporal fine structure of the stimuli (Moore and Moore, 2003b; Moore *et al.*, 2006a; 2006b; Hopkins and Moore, 2007; Ives and Patterson, 2008). An alternative interpretation is that it reflects a transition from resolved to unresolved harmonics (Hoekstra and Ritsma, 1977; Hoekstra, 1979; Carlyon and Shackleton, 1994), or from harmonics that would usually be resolved to harmonics that would usually be unresolved (Bernstein and Oxenham, 2003, 2008). The present results support the former interpretation, since it appears that harmonics above the eighth are not usually resolved.

These results also have implications for the interpretation of a test of sensitivity to temporal fine structure described by Hopkins and Moore (2007) and Moore and Sek (2009). The validity of this test depends on the partials in the stimuli being unresolved. The task involves discriminating a harmonic complex tone (H), with fundamental frequency F0, from a similar tone in which all components are shifted up in frequency by the same amount in hertz, ΔF, so as to create an inharmonic tone (I). The shift does not change the envelope repetition rate, but it results in a change in the temporal fine structure of the sound. As argued above, for tones with equal-amplitude harmonics, harmonics with numbers above 8 are not resolved. However, a harmonic higher than the eighth may be resolved when it has a higher amplitude than adjacent harmonics or when it is the highest or lowest harmonic in the complex tone. To generate complex tones that contained only unresolved harmonics, Hopkins and Moore (2007) and Moore and Sek (2009) passed the complex tones through a fixed bandpass filter with moderately steep slopes of 30 dB/octave. A similar strategy has been used by previous researchers (Hoekstra, 1979; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Moore and Moore, 2003a). To prevent detection of harmonics well down on the skirts of the resulting spectrum, and to prevent the detection of combination tones, the stimuli were presented in a background noise.

An important question is: where does the bandpass filter need to be positioned to ensure that no harmonics are resolved? In one of the conditions of Hopkins and Moore (2007), the filter was centered on the 11th harmonic, and the lowest harmonic within the passband was the 9th. It is likely that one or two harmonics falling on the lower skirt of the filter would have been audible in the presence of the background noise. Thus the lowest audible harmonic would have been the seventh or the eighth. These harmonics would have been barely, if at all, resolved, but it is just possible that they provided usable spectral cues. However, our more recent work (Moore *et al.*, 2009) has shown that it is possible to perform the test (albeit with poorer performance), when the bandpass filter is centered on the 13th or 15th harmonic. In the latter case, the lowest audible harmonic would have been the 11th, and this would certainly have been unresolved. Thus, performance of the test does not seem to depend on information from resolved harmonics when the CF of the bandpass filter is sufficiently high.

VI. CONCLUSIONS

The ability to hear out a partial in a complex tone was strongly affected by pulsing the target partial on and off, and by varying the ramp times and off times of the pulses; longer ramp times and longer inter-pulse intervals led to improved performance. The pattern of results was similar to that obtained when the task was to detect which of two complex tones contained a pulsed partial. The results are more consistent with an explanation for the effect of the pulsing in terms of recovery from adaptation than with an explanation in terms of attentional or cognitive factors. Results obtained using a model of auditory processing are consistent with this interpretation. Overall, the results suggest that measurement of the ability to hear out a partial in a complex tone when the target partial is pulsed on and off does not give a valid indication of how well that partial can be heard out in the more common situation when the target tone is not pulsed.

ACKNOWLEDGMENTS

This work was supported by the MRC (U.K.). We thank Hedwig Gockel, Kathryn Hopkins, Richard Freyman, Andrew Oxenham, and an anonymous reviewer for helpful comments on an earlier version of this paper.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," J. Acoust. Soc. Am. **120**, 3916–3928.

Bernstein, J. G., and Oxenham, A. J. (**2008**). "Harmonic segregation through mistuning can improve fundamental frequency discrimination," J. Acoust. Soc. Am. **124**, 1653–1667.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?" J. Acoust. Soc. Am. **95**, 3541–3554.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.

Fine, P. A., and Moore, B. C. J. (**1993**). "Frequency analysis and musical ability," Music Percept. **11**, 39–53.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gockel, H., Moore, B. C. J., Carlyon, R. P., and Plack, C. J. (**2007**). "Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency," J. Acoust. Soc. Am. **121**, 373–382.

Hartmann, W. M., and Doty, S. L. (**1996**). "On the pitches of the components of a complex tone," J. Acoust. Soc. Am. **99**, 567–578.

Hartmann, W. M., and Goupell, M. J. (**2006**). "Enhancing and unmasking the harmonics of a complex tone," J. Acoust. Soc. Am. **120**, 2142–2157.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). "Hearing a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc.

Am. **88**, 1712–1724.

Helmholtz, H. L. F. (**1954**). *On the Sensations of Tone* (Dover, New York).

Hoekstra, A. (**1979**). "Frequency discrimination and frequency analysis in hearing," Ph.D. thesis, Institute of Audiology, University Hospital, Groningen, The Netherlands.

Hoekstra, A., and Ritsma, R. J. (**1977**). "Perceptive hearing loss and frequency selectivity," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London, England).

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.

ISO 389-8 (**2004**). Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (International Organization for Standardization, Geneva).

Ives, D. T., and Patterson, R. D. (**2008**). "Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics," J. Acoust. Soc. Am. **123**, 2670–2679.

Jepsen, M. L., Ewert, S. D., and Dau, T. (**2008**). "A computational model of human auditory signal processing and perception," J. Acoust. Soc. Am. **124**, 422–438.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

Moore, B. C. J. (**1973**). "Some experiments relating to the perception of complex tones," Q. J. Exp. Psychol. **25**, 451–475.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Moore, B. C. J., Alcántara, J. I., and Dau, T. (**1998**). "Masking patterns for sinusoidal and narrowband noise maskers," J. Acoust. Soc. Am. **104**, 1023–1038.

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments," in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).

Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006b**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006a**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure information," Hear. Res. **222**, 16–27.

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006c**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.

Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.

Moore, B. C. J., and Moore, G. A. (**2003a**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res. **182**, 153–163.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., Hopkins, K., and Cuthbertson, S. J. (**2009**). "Discrimination of complex tones with unresolved components using temporal fine structure information," J. Acoust. Soc. Am. (in press).

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). "Detection of temporal gaps in sinusoids: Effects of frequency and level," J. Acoust. Soc. Am. **93**, 1563–1570.

Moore, B. C. J., and Sek, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. (in press).

Moore, G. A., and Moore, B. C. J. (**2003b**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner haircells," Hear. Res. **24**, 1–15.

Palmer, A. R., Summerfield, Q., and Fantini, D. A. (**1995**). "Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement," J. Acoust. Soc. Am. **97**, 1786–1799.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.

Plack, C. J., and Moore, B. C. J. (**1991**). "Decrement detection in normal and impaired ears," J. Acoust. Soc. Am. **90**, 3069–3076.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am. **43**, 764–767.

Roberts, B., and Bregman, A. S. (**1991**). "Effects of the pattern of spectral spacing on the perceptual fusion of harmonics," J. Acoust. Soc. Am. **90**, 3050–3060.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

Smith, R. L. (**1977**). "Short-term adaptation in single auditory-nerve fibers: Some poststimulatory effects," J. Neurophysiol. **49**, 1098–1112.

Smith, R. L. (**1979**). "Adaptation, saturation and physiological masking in single auditory-nerve fibers," J. Acoust. Soc. Am. **65**, 166–178.

Soderquist, D. R. (**1970**). "Frequency analysis and the critical band," Psychonomic Sci. **21**, 117–119.

Studebaker, G. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Westerman, L. A., and Smith, R. L. (**1984**). "Rapid and short-term adaptation in auditory nerve responses," Hear. Res. **15**, 249–260.

An investigation of auditory dimensional interaction in a bivariate bilateral conditioning paradigm in the rabbit

Kristin N. Mauldin
Department of Neurosciences, University of California, San Diego, 340 Stein Clinical Research Building, La Jolla, California 92093

Robin D. Thomas, Stephen D. Berry, and William P. O'Brien
Psychology Department, Miami University, 90 North Patterson Avenue, Oxford, Ohio 45056

Christen W. Hoedt
School of Medicine, Boston University, 715 Albany Street, Boston, Massachusetts 02118

(Received 13 June 2008; revised 15 January 2009; accepted 17 January 2009)

The current study adapted the Garner paradigm for diagnosing separable versus integral perceptual dimensions to the eye-blink classical conditioning paradigm using rabbits. Specifically, this study examined the ability of rabbits to categorize stimuli based on one auditory dimension while ignoring a second, irrelevant dimension by displaying an appropriate eye-blink for bilaterally conditioned discriminative responses. Tones used in training varied along two dimensions, starting frequency and magnitude of frequency sweep upwards from the start. Rabbits first learned to categorize along a single dimension (blinking one eye for one category response and the other eye for the other response) and then continued to categorize tones in a second phase in which the irrelevant dimension was varied. The variation of the irrelevant dimension did not disrupt performance, indicating that rabbits perceive these dimensions as separable.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081387]

PACS number(s): 43.66.Gf, 43.66.Lj [BLM] Pages: 3205–3213

I. INTRODUCTION

Knowing how dimensions of a stimulus interact in a multidimensional situation is essential before proper inferences can be made from performance (Palmer, 1978). Aspects of performance attributed to "higher level" cognitive constructs may actually result from interactions among stimulus dimensions early in perception (Gibson, 1979; MacMillan and Creelman, 2004). The failure of an animal to generalize along one dimension may not be a problem in learning or memory but may be the result of an experimental (design) assumption that the dimensions that the animal perceived are *separable,* that is, they afford selective attention, rather than *integral*, or holistically processed (Garner, 1974a, 1974b). Although other types of dimensional interaction have been defined (e.g., Garner, 1976; Ashby and Townsend, 1986; Kadlec and Townsend, 1992a, 1992b; Thomas, 2001), these are the most studied in the human perceptual literature. When humans are the research participants, our knowledge of how a stimulus is perceived is no doubt aided by our ability to use language. In nonverbal organisms, whose behavioral repertoire is limited, discovering which stimulus components are perceived separably presents a challenge. For instance, upon discovering that the dimensions of groove width and force interacted with monkeys' perception of roughness, Pruett *et al.* (2001) stated that a clear interpretation of their classification of neural responses was impossible.

A multitude of instrumental conditioning studies have been conducted on the interaction between task design and working memory capacity and its effect on selective attention, especially with avians (for review, see Zentall, 2005), but very few studies have explored the issue of perceived dimensional interaction (e.g., Appeltants *et al.*, 2005; Hulse *et al.*, 1997; Pruett *et al.*, 2001; Wisniewski and Hulse, 1997; Yokoyama *et al.*, 2006). Relevant studies have found that canaries and starlings are able to continue discriminating birdsongs successfully when songs from birds of different species or social structures are overlaid onto the original songs (Appeltants *et al.*, 2005; Hulse *et al.*, 1997; Wisniewski and Hulse, 1997). In the Pavlovian conditioning literature, parameters involved in the learning of bisensory, compound stimuli have undergone intensive investigation (see Kehoe and Gormezano, 1980 and Wasserman and Miller, 1997 for review). Because bisensory stimuli are of different modalities, it is often assumed that they are perceived as separable (Garner, 1974a; Treisman, 1969; but see Marks, 2004; Melara and O'Brien, 1987; and Melara, 1989). However, the study of dimensional interaction with bivariate unimodal stimuli has received much less attention in the animal literature. In the case of such unimodal stimuli, the assumption of separability cannot be made *a priori*, and, therefore, must be tested empirically. It is the goal of this study to explore the potential for sensory interactions between the dimensions of auditory bivariate unisensory stimuli in the rabbit using the well-established eyeblink classical conditioning (EBCC) paradigm.

Due to its extensive application in both human and animal behavioral learning, EBCC is an especially useful paradigm for studying perceptual and cognitive processes (Lonsbury-Martin *et al.*, 1976; Martin *et al.*, 1977; Woodruff-Pak and Steinmetz, 2000a, 2000b). Use of this paradigm has greatly advanced our knowledge of the processes underlying learning and memory, even suggesting an essential engram for simple associative learning (Thompson, 1986). Classical conditioning involves presenting a neutral stimulus (i.e., tone) just prior to the presentation of a biologically significant stimulus (i.e., paraorbital shock). The learner reflexively responds to the biologically significant stimulus (i.e., eyeblink), but after several pairings will begin responding to the neutral stimulus in anticipation of the shock, thereby demonstrating learning. Comparisons between human and animal (typically rabbit) EBCC studies revealed parallel cognitive and neurological phenomena (Woodruff-Pak and Steinmetz, 2000a, 2000b).

An important methodology for assessing dimensional interaction in humans includes the set of speeded classification tasks initially proposed by Garner (1974b, 1976) and further studied by Ashby and Maddox (1994), Kadlec and Townsend (1992a, 1992b), and Thomas (1996, 2001). Although three tasks comprised the set originally (Garner, 1974b), two have been found to be useful in distinguishing separable from integral processing (Maddox, 1992), the *baseline* and *filtering* conditions. Often, four objects are constructed from factorial combination of two dimensions sampled at two levels each. For example, the stimuli could be sinusoidal tones whose amplitudes and frequencies each vary across two levels. On a given trial, the observer classifies the tone according to its level on "relevant" dimension. In the baseline task, the irrelevant dimension is held constant, providing a measure of discriminative performance on the relevant dimension. In the filtering task, both the relevant and irrelevant dimensions vary but the observer continues to classify along the relevant dimension only. If filtering performance does not differ from baseline, then the dimensions are deemed separable. However, if filtering performance declines (either by slower response times or higher error rates) relative to baseline, Garner interference is thought to have occurred and the dimensions are deemed integral. In human sensory perception, examples of integral dimensions include auditory frequency and amplitude (Fletcher and Munson, 1933; Melara and Marks, 1990) and color hue and saturation (Garner and Felfoldy, 1970). We adapted the Garner methodology to study separability of auditory dimensions [starting frequency (SF) and frequency sweep (FS)] in rabbits, by modifying the EBCC paradigm.

Most studies of EBCC use auditory stimuli, but very few have used auditory stimuli that vary orthogonally on two dimensions, and none have explored the potential for a dimensional interaction. Therefore, the current study aims to answer the question of whether rabbits can attend separably to one dimension of an auditory tone (i.e., frequency starting value) when another dimension (i.e., FS) varies orthogonally. The answer to this question will provide insights into the rabbit's perceptual abilities and document that the eyeblink conditioning paradigm can be extended to discrimination tasks of greater complexity than is typically employed. The ability to construct larger stimulus sets of clearly discriminable auditory stimuli would enable neurobiological studies of more complex cognitive processes, such as categorization, using the rabbit EBCC paradigm.

A. The current study

In the current study, rabbits in the experimental conditions were trained on a discrimination task with two phases using tones that varied in two dimensions, starting or initial frequency versus magnitude 1 or 2 octaves of FS (the difference in starting and stopping frequency values for a smoothly linear changing tone frequency). Phase 1 served as the baseline condition in which only the dimension relevant to the discrimination task varied. This was phase 2 in the counterbalance group. In the second phase, both dimensions varied but eyeblink responses were reinforced only with respect to the relevant dimension value. This was phase 1 in the counterbalance group. In the control task, the dimension deemed relevant differed across the two phases. So control rabbits were required to learn a new set of conditioned stimuli in their second phase. If the rabbits in the experimental condition were unable to filter out variation in the irrelevant dimension, the stimuli in the second phase would be perceived as a new set as with the control animals. In that case, the performance in the second phase should be similar to that of performance in the second phase of the control condition. However, if rabbits can filter out variations of the irrelevant dimension, performance on the second phase should be similar to performance at the end of the first phase and be much better than performance in the second phase of the control condition. Therefore, comparisons between control and experimental performance in the second phase will demonstrate whether rabbits integrate SF with FS or whether they can categorize tones independently on these two dimensions.

II. METHODS

A. Subjects

Subjects were six New Zealand White rabbits (*Oryctolagus cuniculus)* obtained from Myrtles Rabbitry, Inc. (Thompson Station, TN). Animals were maintained on a 12:12 light/dark schedule, with training taking place during the light phase. Food was given ad lib, and water was continuously available. Handling and treatment of animals were in accordance with National Institutes of Health guidelines and Miami University's Institutional Animal Care and Use Committee.

B. Apparatus

During adaptation and training sessions, each animal was placed into a restrainer inside a Faraday cage designed to attenuate extraneous sounds and reduce electrical interference. The restrainer was a Plexiglas box with a sloping front plate with a U-shaped slot through which the animal's head could protrude, and an adjustable stock that slipped over the slot to keep head in place (Gormezano, 1966).

After securing the head, a rear plate was adjusted up to the animal's hindquarters to keep the body firmly restricted. A sliding Plexiglas plate was inserted into grooves to cover the top of the box, thus preventing back injuries (Patterson

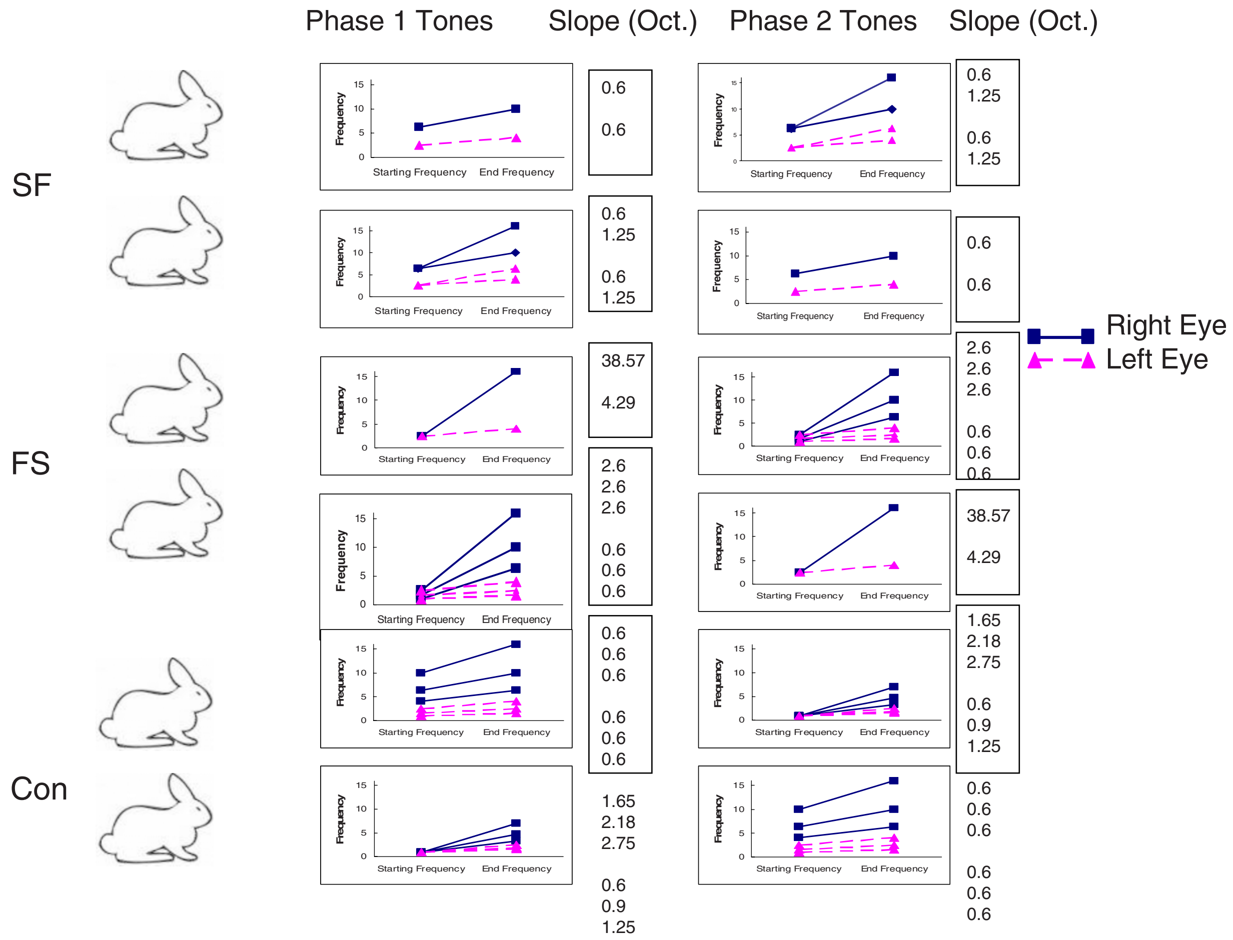


FIG. 1. (Color online) A depiction of the experimental paradigm. The frequencies of the tones given in phases 1 and 2 are shown for all six rabbits. Note that the *y*-axis is not in octaves so slopes in the higher frequency range convert to smaller slopes when converted into octaves.

and Romano, 1987). Animals were typically in the restrainer for no more than 1.5 h at a time. 2 days of adaptation to restraint was provided.

C. Stimuli

The conditioned stimuli were a set of tones that ranged between 1 and 16 kHz. Rabbits have relatively small variation in sensitivity threshold across this range (Martin *et al.*, 1980), making it useful for auditory experiments. Using a go-no go paradigm, the tone frequencies were pretested in other rabbits to make sure that they could discriminate them (pilot studies in our laboratory). These tones were calibrated using a sound level meter (Quest Electronics Model OB-300) and adjusted to 83 dB. Each tone was ramped up from 0 to 83 dB at its onset. The tone was produced by an auditory signal control interface (Tucker Davis Technologies, Alachua, FL) through a small speaker set 10 cm in front of the rabbits' ears. Each tone was 350-ms long and varied in SF and FS. Figure 1 and Table II provide a description of the stimuli.

The unconditioned stimuli (US) consisted of periorbital shock. Four stainless steel autoclip wound clips were placed transcutaneously 10 mm caudal to the rabbit's left and right eyes with the lower clip 10 mm ventral to the first clip. Two constant-current, 60-Hz square wave stimulators (Model SD9 Grass, Quincy, MA) delivered between 0.5 and 1 mA, 100-ms ac to the clip on the eye paired with the category of the CS given on that trial. Amplitude of the current was adjusted to elicit a blink from the rabbit resulting in electromyography (EMG) activity that exceeded 1 standard deviation above the mean of the baseline noise.

D. Behavioral training

The current methods are an adaptation of the Garner (1976) interference task to the rabbit EBCC paradigm. The basic behavioral training was the same for every animal across groups. On each trial, the rabbit was presented with a 350-ms tone (CS) that coterminated with a 100-ms shock (US) adjacent to the appropriate eye (see below). The time between CS onset and US onset (interstimulus interval) was 250 ms and the intertrial interval was 30 s. To simulate the typical two response/two category structure of the Garner interference task, the shock was administered to the eye appropriate for the experimenter-defined category of the tone given. Tones from one category resulted in stimulation to the right eye; tones from the other resulted in stimulation to the left eye. A response was considered correctly conditioned when the rabbits blinked the appropriate eye in the 200-ms period prior to the stimulation. Each training day consisted of seven 20 trial blocks for a total of 140 trials. Stimuli were

TABLE I. Experimental design depicting phases 1 and 2 of the experimental (left) and control (right) conditions. SF=starting frequency in kHz. FS =frequency sweep in octaves.

Experimental design					
Experimental conditions			Control conditions		
Task: Starting frequency relevant					
	Phase 1	Phase 2		Phase 1	Phase 2
Test	Baseline: Change in starting frequency. No change in sweep	Orthogonal: Change in starting frequency. Change in sweep	SF to FS	Classify by starting frequency.	Classify by sweep.
				Change in starting frequency. No change in sweep.	Change in sweep. No change in starting frequency.
Counterbalance	Orthogonal: Change in starting frequency. Change in sweep.	Baseline: Change in starting frequency. No change in sweep	FS to SF	Classify by sweep.	Classify by starting frequency.
				Change in sweep. No change in starting frequency.	Change in starting frequency. No change in sweep.
Task: Frequency sweep relevant					
	Phase 1	Phase 2			
Test	Baseline: Change in sweep. No change in starting frequency.	Orthogonal: Change in sweep. Change in starting frequency.			
Counterbalance	Orthogonal: Change in sweep. Change in starting frequency.	Baseline: Change in sweep. No change in starting frequency.			

randomized at the beginning of each session so that every stimulus and, therefore, each category had an equal probability of being presented on a given trial.

The overall design required three conditions (two experimental and one control), each with two phases. The experimental conditions were defined by the dimension that was relevant to the discrimination task, either SF or FS. In both phases, variations in the relevant dimension could be used to predict the eye that was going to be stimulated. For instance, in the FS relevant experimental task the rule was defined as “If the FS is large, blink the right eye. If the FS is small, blink the left eye.” Accordingly, if the FS was 0.6 octaves the left eye was stimulated, if the FS was 2.5 octaves the right eye was stimulated. The irrelevant dimension was stationary in one phase and varied in a manner orthogonal to the other dimension in the other phase. Figure 1 and Table I diagram the experimental design. All of the phases within conditions were counterbalanced. Two rabbits were assigned to each condition for a total of six rabbits (four experimental and two control). The stimuli were tones that varied in their frequency at onset (SF) and increase in frequency from onset to offset (FS). The time from tone onset to offset was held constant across all tones.

To control for dimension integrality the irrelevant dimension was constant in both phases. However, the relevant and irrelevant dimensions switched from phase 1 to phase 2 so that the relevant dimension in phase 1 became irrelevant in phase 2 and the irrelevant dimension in phase 1 became relevant in phase 2. For instance, in one phase, variations in FS (right eye: 1.65, 2.18, and 2.75 octaves; left eye: 0.6, 0.9, and 1.25 octaves) were relevant to the discrimination task while the level of SF was the same for all stimuli (1 kHz). In the other phase variations SF was relevant to the discrimination (right eye: 4, 6.3, and 10 kHz; left eye: 1, 1.6, and 2.5 kHz) while the level of FS was the same for all stimuli (0.6 octaves). An example of two trial types for the SF condition is given in Fig. 2. The control condition tested speed of acquisition when learning a new set of stimuli in phase 2.

E. EMG

EMG activity was recorded from blunt tailor’s hooks placed on the rabbits’ upper eyelids. Each time the rabbit blinked, the EMG activity was digitized by DataWave and scored using data analysis software in MATLAB. The learning criterion was defined as the block in which 60% of the trials had (a) a maximum EMG amplitude in the conditioned response (CR) period that was greater than 1 standard deviation above the mean amplitude of the baseline EMG and (b) the area under the curve in the CR period of the appropriate eye was significantly greater than 50% of the total area under the curve for both eyes (i.e., a significant difference score). The area under the curve was calculated by summing the amplitudes of the EMG signal for 100 ms before the onset of

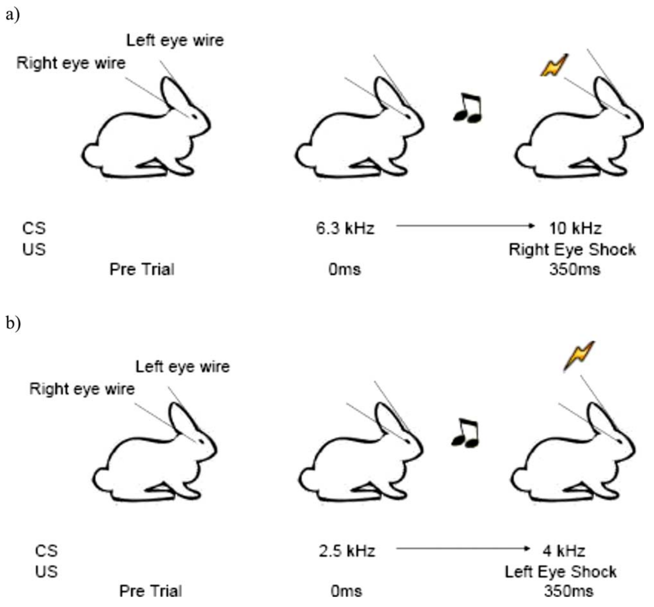


FIG. 2. (Color online) An example of a right-eye enforced trial (a) and a left-eye enforced trial (b). These examples are taken from the SF experimental condition.

the shock. A one-way t-test was used to address whether the area under the curve for the appropriate eye was significantly ($p < 0.05$) greater than 50% of the sum of the area under the curve for both eyes. Part 2 of the criterion ensured that the response of the appropriate eye was significantly greater than the response of the inappropriate eye, thus demonstrating discriminative performance. Refer to Fig. 3 for examples of discriminative and nondiscriminative responses.

III. RESULTS

Comparison of the rabbits' performance in phases 1 and 2 reveal that the experimental animals performed significantly better on phase 2 than on phase 1 [phase 1: SF (starting frequency) $\mu=27$, $\sigma=15.56$, FS (frequency sweep) $\mu=61.5$, $\sigma=0.71$; phase 2: SF $\mu=1$, $\sigma=0$; FS $\mu=2.5$, $\sigma=2.12$] indicating that change in the irrelevant dimension did not interfere with transfer of learning from phase 1 to phase 2. As was expected, the controls did not show this advantage (phase 1:$\mu=40.5$, $\sigma=16.26$; phase 2: $\mu=48$, $\sigma=7.07$).

Figure 4 displays performance on both phases for all conditions. Each bar pair (gray and black bars presented side-by-side) represents the number of blocks to criterion in either phase 1 (gray bars) or phase 2 (black bars) for a single rabbit. As can be seen from this figure, transfer of training (i.e., learning of the categorization rule) was obtained for the experimental conditions but not for the control conditions.

While both the SF group and the FS group began emitting CRs at nearly the same time, it took longer for the FS group to emit discriminative responses. This led to a greater number of blocks to criteria on phase 1. However, it is important to note that their number of blocks to criterion on phase 2 did not differ significantly.

Anecdotally, some rabbits tended to prefer to blink their right eye initially while other rabbits preferred to blink their left eye. Once the rabbits began blinking both eyes this bias was observed by a more intense blink of the preferred eye. However, once the rabbit began to discriminate the stimuli correctly these biases were no longer observed.

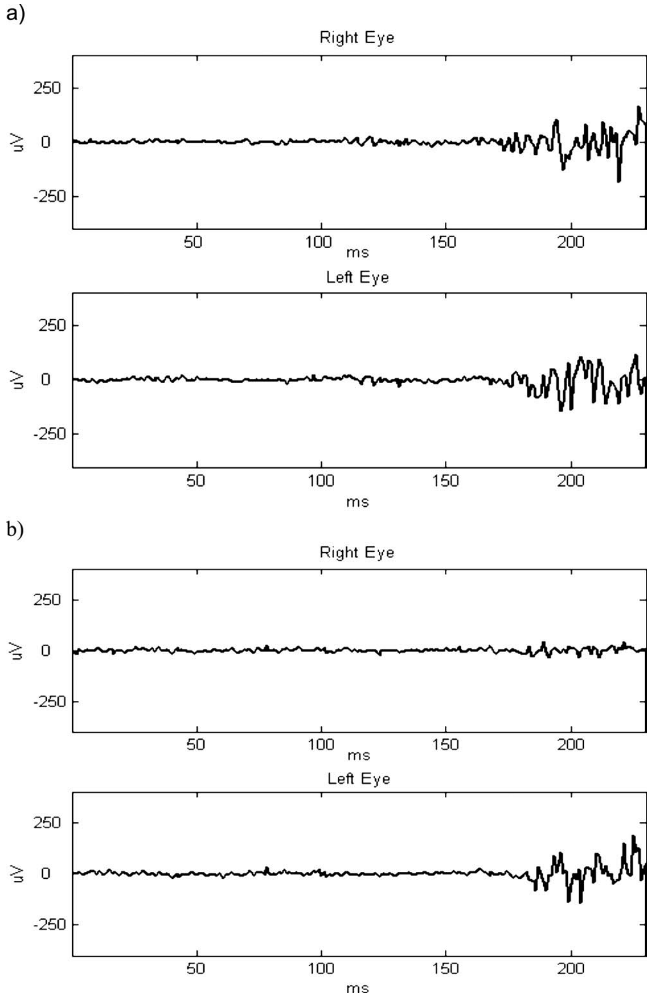


FIG. 3. Examples of the EMG responses from both eyes during (a) bilateral CRs and (b) a left-eye unilateral CR. μV=microvolts; ms=milliseconds.

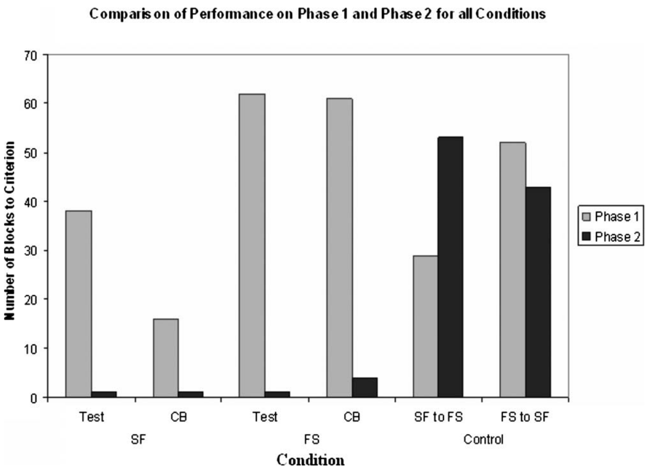


FIG. 4. Comparison of number of blocks to criterion between phases 1 and 2 for all three conditions. CB=counterbalance. Each column pair (side-by-side gray and black column) represents the performance of one rabbit, for a total of six rabbits.

TABLE II. Stimulus setups for all conditions. SF=starting frequency in kHz. FS=frequency sweep in octaves. CR=conditioned response.

	Experimental conditions										Control condition			
	Phase 1		Phase 2			Phase 1		Phase 2			Phase 1		Phase 2	
	SF	FS	SF	FS		SF	FS	SF	FS		SF	FS	SF	FS
SF condition					FS condition					SF to FS				
CR=Rt. eye-blink	6.3	0.6	6.3	0.6	CR=Rt. eye-blink	2.5	2.5	1	2.5	CR=Rt. eye-blink	4	0.6	1	1.65
			6.3	1.25				1.6	2.5		6.3	0.6	1	2.18
								2.5	2.5		10	0.6	1	2.75
CR=Lft. eye-blink	2.5	0.6	2.5	0.6	CR=Lft. eye-blink	2.5	0.6	1	0.6	CR=Lft. eye-blink	1	0.6	1	0.6
			2.5	1.25				1.6	0.6		1.6	0.6	1	0.9
											2.5	0.6	1	1.25
FS condition					SF condition					FS to SF				
CR=Rt. eye-blink	6.3	0.6	6.3	0.6	CR=Rt. eye-blink	1	2.5	2.5	2.5	CR=Rt. eye-blink	1	1.65	4	0.6
	6.3	1.25				1.6	2.5				1	2.18	6.3	0.6
						2.5	2.5				1	2.75	10	0.6
CR=Lft. eye-blink	2.5	0.6	2.5	0.6	CR=Lft. eye-blink	1	0.6	2.5	0.6	CR=Lft. eye-blink	1	0.6	1	0.6
	2.5	1.25				1.6	0.6				1	0.9	1.6	0.6
						2.5	0.6				1	1.25	2.5	0.6

In order to factor out individual performance differences, the number of blocks to criterion on phase 2 was divided by the number of blocks to criterion for phase 1. Percentages close to 0 indicate minimal interruption while percentages near 1 or greater (similar to the control condition's percentages) indicate substantial interference. The results were as follows: In the experimental condition there is little to no interruption from the varying, irrelevant dimension. SF test condition: 2.6%, SF counterbalance: 6.3%, FS test condition: 1.6%, FS counterbalance: 6.6%. In the control condition there was no clear advantage of practice. SF to FS control: 183%, and FS to SF control: 83%. The data were transformed logarithmically in order to compensate for the high variability and small sample size that can produce volatility in ratio data (Howell, 2006). Upon comparison of the transformed data, significant differences were found between the control and the experimental conditions [$t(4)=6.2, p=0.003$]. Learning curves of the percentage correct on each 20 trial block showed an incremental increase in learning as opposed to a step-function indicating a gradual learning process.

IV. CONCLUSIONS

Rabbits in the experimental conditions continued performance on phase 2 almost as if nothing had changed. In fact, three out of the four experimental rabbits reached criterion within the first block of phase 2. Rabbits in the control conditions, however, took nearly as long or longer to complete phase 2 compared to phase 1. Unaltered performance across phases indicates that perception of the relevant dimension did not change; hence, SF and FS were perceived as separable.

The experiment presented here required rabbits to rely on discrimination of either SF or FS in order to reach criterion. It was not possible that the rabbits used SF to reach criterion in the FS condition or used the FS to reach criteria in the SF condition (Tables II and III and Fig. 1). It might have been possible, however, that they were able to discriminate by focusing on the end frequency (EF) in all conditions (Table III). Importantly, if that were the case, one would expect the control rabbits to have performed just as well as the experimental rabbits in phase 2. This was clearly not the case, leading us to conclude that the rabbits had used the SF and FS to perform the task.

One might be tempted to claim that this task closely resembles a simple frequency discrimination task. While more than one dimension varies, the animals perform as if only the relevant dimension varies. This is exactly the point we are trying to make. The fact that the animals perform as if only the relevant dimension is varying is evidence that the animals are unaffected by the varying irrelevant dimension. In other words, variation of the irrelevant dimension does not alter their performance, indicating that it does not alter their perception of the relevant dimension. It is this evidence that leads us to conclude that the rabbits are able to separably attend to these dimensions. In other words, these dimensions can be perceptually represented as independent of each other.

The distinction between selective attention, as it is used in the animal learning literature, and perceived separability, as it is used here may be unclear. While selective attention is related to, and affected by, the perception of dimensional interactions, these terms have been applied to somewhat different constructs (Garner, 1976; Kehoe and Gormezano, 1980; Zentall, 2005). Studies on separability examine the effects of change in one dimension on the perception of another dimension. Selective attention, on the other hand, has been defined as the way attentional resources are allocated. This can be altered by many factors including saliency and reinforcement of a dimension. The two constructs are related in the following manner: If two dimensions are perceived as separable, and only one is relevant to the task, little to no resources will be wasted on the irrelevant dimension. However, if the two dimensions are perceived as integral, an orthogonal change in the irrelevant dimension will be a drain

TABLE III. The SF, slope (in octaves), and EF for all rabbits in all conditions and phases. In addition, the rules that could have been used to reach criteria are given for each condition and phase.

			Phase 1						Phase 2			
			SF	Slope	EF	Possible rules			SF	Slope	EF	Possible rules
FS	Expt.	RT	6.3	0.6	10	If the SF/EF frequency is high, blink the right eye. If it is low, blink the left eye	Expt.	RT	6.3	0.6	10	
		LT	2.5	0.6	4				6.3	1.25	16	
	CB	RT	6.3	0.6	10			LT	2.5	0.6	4	If the SF/EF frequency is high, blink the right eye. If it is low, blink the left eye.
			6.3	1.25	16				2.5	1.25	6.3	
		LT	2.5	0.6	4	If the SF/EF frequency is high, blink the right eye. If it is low, blink the left eye.	CB	RT	6.3	0.6	10	If the SF/EF frequency is high, blink the right eye. If it is low, blink the left eye.
			2.5	1.25	6.3			LT	2.5	0.6	4	
SF	Expt.	RT	1	2.6	6.3		Expt.	RT	1	2	6.3	
			1.6	2.6	10				1	2.6	10	
			2.5	2.6	16				1	3.25	16	
		LT	1	0.6	1.6			LT	1	0	1.6	
			1.6	0.6	2.5	If the slope or EF frequency is high, blink the right eye. If it they are low, blink the left eye.			1	0.6	2.5	If the slope or EF frequency is high, blink the right eye. If it they are low, blink the left eye.
			2.5	0.6	4				1	1.25	4	
	CB	RT	1	2	6.3		CB	RT	1	2.6	6.3	
			1	2.6	10				1.6	2.6	10	
			1	3.25	16				2.5	2.6	16	
		LT	1	0	1.6			LT	1	0.6	1.6	
			1	0.6	2.5	If the slope or EF frequency is high, blink the right eye. If it they are low, blink the left eye.			1.6	0.6	2.5	If the slope or EF frequency is high, blink the right eye. If it they are low, blink the left eye.
			1	1.25	4				2.5	0.6	4	
Control	Con1	RT	1	1.65	3.3		Con1	RT	4	0.6	6.3	
			1	2.18	4.7				6.3	0.6	10	
			1	2.75	7				10	0.6	16	
		LT	1	0.6	1.6			LT	1	0.6	1.6	
			1	0.9	1.9	If the slope or EF frequency is high, blink the right eye. If it they are low, blink the left eye.			1.6	0.6	2.5	If the SF/EF frequency is high, blink the right eye. If it is low, blink the left eye.
			1	1.25	2.5				2.5	0.6	4	
	Con2	RT	4	0.6	6.3		Con2	RT	1	1.65	3.3	If the slope or EF frequency is high, blink the right eye. If it they are low, blink the left eye.
			6.3	0.6	10				1	2.18	4.7	
			10	0.6	16				1	2.75	7	
		LT	1	0.6	1.6			LT	1	0.6	1.6	
			1.6	0.6	2.5	If the SF/EF frequency is high, blink the right eye. If it is low, blink the left eye.			1	0.9	1.9	
			2.5	0.6	4				1	1.25	2.5	

on the attentional capacity. The processes underlying these two constructs are thought to rely on different neural substrates; the parietal, thalamic, and midbrain structures underlying selective attention of the task and the anterior cingulate underlying attention to the relevant dimension (Shalev and Algom, 2000). Thus, our finding of perceived separability between SF and FS indicates that, when only one of these dimensions is relevant, changes in the irrelevant dimension will not tax attentional resources. In other words, the animal was able to selectively attend to the relevant dimension and ignore the irrelevant dimension.

The criterion used in this study required all of the rabbits to produce a substantially larger blink by the eye paired with the CS than by the contralateral eye. All of the rabbits in this study reached this criterion. This may seem unexpected, given the body of evidence demonstrating that bilateral conditioning often occurs in unilateral paradigms (Christian and Thompson, 2003). However, as the current results others have shown (Brandon *et al.*, 1994), rabbits are quite capable of producing discrete unilateral responses when required by a bilateral eyeblink discrimination task [Fig. 3(b)]. Some eyelid movement can be seen in the contralateral eye, but it is significantly smaller than that of the ipsilateral eye once the animal has learned the task.

Our demonstration of separable dimensions in the rabbit adds to an important and growing knowledge of the rabbit auditory system (i.e., Heffner and Masterton, 1980; Howard *et al.*, 2002; Kettner and Thompson, 1978, 1985; Lenhardt, 1979; Lonsbury-Martin *et al.*, 1976; Martin *et al.*, 1977; Kraus and Disterhoft, 1981, 1982) and offers the potential for further work on neurobiological mechanisms of learning and categorization. In addition, the successful use of bilateral EBCC in a multiple stimulus discrimination paradigm provides a foundation that may now be studied using rabbits. Dimensions and stimulus compounds shown to be independent here (i.e., SF and FS) can now be used to study cognitive processes and behavioral phenomena such as overtraining, expertise effects, context effects, and categorization. Examination of the neural substrates that underlie these processes will contribute to our understanding of the relationship between brain systems and mnemonic processes in humans and animals (Cook and Smith, 2006). This would add another level of sophistication to our knowledge of the neurobiological substrates of associative learning using the rabbit EB paradigm and may change the way those results are interpreted.

Appeltants, D., Gentner, T. Q., Hulse, S. H., Balthazart, J., and Ball, G. F. (**2005**). “The effect of auditory distractors on song discrimination in male canaries (*Serinus canaria*),” Behav. Processes **69**, 331–341.

Ashby, F. G., and Maddox, W. T. (**1994**). “A response time theory of separability and integrality in speeded classification,” J. Math. Psychol. **38**, 423–466.

Ashby, F. G., and Townsend, J. T. (**1986**). “Varieties of perceptual independence,” Psychol. Rev. **93**, 154–179.

Brandon, S. E., Betts, S. L., and Wagner, A. R. (**1994**). “Discriminated lateralized eyeblink conditioning in the rabbit: An experimental context for separating specific and general associative influences,” J. Exp. Psychol. Anim. Behav. Process. **20**, 292–307.

Christian, K. M., and Thompson, R. F. (**2003**). “Neural substrates of eyeblink conditioning: Acquisition and retention,” Learn. Memory **10**, 427–455.

Cook, R. G., and Smith, J. D. (**2006**). “Stages of abstraction and exemplar memorization in pigeon category learning,” Psychol. Sci. **17**, 1059–1067.

(**2000a**). *Eyeblink Classical Conditioning: Volume I: Applications in Humans*, edited by D. S. Woodruff-Pak and J. E. Steinmetz (Kluwer Academic, Boston, MA).

(**2000b**). *Eyeblink Classical Conditioning: Volume II: Animal Models*, edited by D. S. Woodruff-Pak and J. E. Steinmetz (Kluwer Academic, Boston, MA).

Fletcher, H., and Munson, W. A. (**1933**). “Loudness, its definition, measurement and calculation,” J. Acoust. Soc. Am. **5**, 82–108.

Garner, W. R. (**1974a**). “Attention: The processing of multiple sources of information,” in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. **2**, pp. 23–59.

Garner, W. R. (**1974b**). *The Processing of Information and Structure* (Erlbaum, Potomac, MD).

Garner, W. R. (**1976**). “Interaction of stimulus dimensions in concept and choice processes,” Cognit Psychol. **8**, 98–123.

Garner, W. R., and Felfoldy, G. L. (**1970**). “Integrality of stimulus dimensions in various types of information processing,” Cognit Psychol. **1**, 225–241.

Gibson, J. J. (**1979**). *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston, MA).

Gormezano, I. (**1966**). “Classical conditioning,” in *Experimental Methods and Instrumentation in Psychology*, edited by J. B. Sidowski (McGraw-Hill, New York).

Heffner, H., and Masterton, B. (**1980**). “Hearing in glires: Domestic rabbit, cotton rat, feral house mouse, and kangaroo rat,” J. Acoust. Soc. Am. **68**, 1584–1599.

Howard, M. A., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2002**). “Effects of reversible noise exposure on the suppression tuning of rabbit distortion-product otoacoustic emissions,” J. Acoust. Soc. Am. **111**, 285–296.

Howell, D. C. (**2006**). *Statistical Methods for Psychology* (Wadsworth, Belmont, CA).

Hulse, S. H., MacDougall-Shackleton, S. A., and Wisniewski, A. B. (**1997**). “Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (Sturnus vulgaris),” J. Comp. Psychol. **111**, 3–13.

Kadlec, H., and Townsend, J. T. (**1992a**). “Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimension,” J. Math. Psychol. **36**, 325–374.

Kadlec, H., and Townsend, J. T. (**1992b**). “Signal detection analyses of dimensional interactions,” in *Multidimensional Probabilistic Models of Perception and Cognition*, edited by F. G. Ashby (Erlbaum, Hillsdale, NJ,), pp. 181–227.

Kehoe, E. J., and Gormezano, I. (**1980**). “Configuration and combination laws in conditioning with compound stimuli,” Psychol. Bull. **87**, 351–378.

Kettner, R. E. (**1978**). “Neural correlates of a signal detection task in the rabbit,” J. Acoust. Soc. Am. **64**, S137–S137.

Kettner, R. E., and Thompson, R. F. (**1985**). “Cochlear nucleus, inferior colliculus, and medial geniculate responses during the behavioral detection of threshold-level auditory stimuli in the rabbit,” J. Acoust. Soc. Am. **77**, 2111–2127.

Kraus, N., and Disterhoft, J. F. (**1981**). “Location of rabbit auditory cortex and description of single unit activity,” Brain Res. **214**, 275–286.

Kraus, N., and Disterhoft, J. F. (**1982**). “Response plasticity of single neurons in rabbit auditory cortex during tone-signalled learning,” Brain Res. **246**, 205–215.

Lenhardt (**1979**). “Pneumatic whistle for animal hearing tests,” Lab. Anim. Sci. **29**, 812–813.

Lonsbury-Martin, B. L., Martin, G. K., Schwartz, S., and Thompson, R. F. (**1976**). “Neural correlates of auditory plasticity during classical conditioning in the rabbit,” J. Acoust. Soc. Am. **60**, S82–S82.

MacMillan, N. A., and Creelman, C. D. (**2004**). *Detection Theory: A User’s Guide* (Lawrence Erlbaum Associates, Philadelphia, PA).

Maddox, W. T. (**1992**). “Perceptual and decisional separability,” in *Multidimensional Models of Perception and Cognition*, edited by F. G. Ashby (Erlbaum, Hillsdale, NJ), pp. 147–180.

Marks, L. E. (**2004**). “Cross-modal interactions in speeded classification,” in *The Handbook of Multisensory Processes*, edited by G. A. Calvert, C. Spence, and B. E. Stein (MIT, Cambridge, MA), pp. 85–105.

Martin, G. K., Lonsbury-Martin, B. L., and Kimm, J. (**1977**). “Auditory sensitivity in the rabbit determined by a conditional nictitating of membrane response,” J. Acoust. Soc. Am. **62**, S88–S88.

Martin, G. K., Lonsbury-Martin, B. L., and Kimm, J. (**1980**). “A rabbit preparation for neuro-behavioral auditory research," Hear. Res. **2**, 65–78.

Melara, R. D. (**1989**). "Dimensional interaction between color and pitch," J. Exp. Psychol. Hum. Percept. Perform. **15**, 69–79.

Melara, R. D., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness," Percept. Psychophys. **48**, 169–178.

Melara, R. D., and O'Brien, T. P. (**1987**). "Interaction between synesthetically corresponding dimensions," J. Exp. Psychol. Gen. **116**, 323–336.

Palmer, S. E. (**1978**). "Structural aspects of visual similarity," Mem. Cognit. **6**, 91–97.

Patterson, M. M., and Romano, A. G. (**1987**). "The rabbit in Pavlovian conditioning," in *Classical Conditioning*, 3rd ed., edited by A. H. Black and W. F. Prokasy (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 1–36.

Pruett, J. R., Jr., Sinclair, R. J., and Burton, H. (**2001**). "Neural correlates for roughness choice in monkey second somatosensory cortex (SII)," J. Neurophysiol. **86**, 2069–2080.

Shalev, L., and Algom, D. (**2000**). "Stroop and Garner effects in and out of Posner's beam: Reconciling two conceptions of selective attention," J. Exp. Psychol. Hum. Percept. Perform. **26**, 997–1017.

Thomas, R. D. (**1996**). "Separability and independence of dimensions in the same-different judgment task," J. Math. Psychol. **40**, 318–341.

Thomas, R. D. (**2001**). "Perceptual interactions of facial dimensions in speeded classification and identification," Percept. Psychophys. **63**, 625–650.

Thompson, R. F. (**1986**). "The neurobiology of learning and memory," Science **233**, 941–947.

Treisman, A. M. (**1969**). "Strategies and models of selective attention," Psychol. Rev. **76**, 282–299.

Wasserman, E. A., and Miller, R. R. (**1997**). "What's elementary about associative learning?," Annu. Rev. Psychol. **48**, 573–607.

Wisniewski, A. B., and Hulse, S. H. (**1997**). "Auditory scene analysis in European starlings (Sturnus vulgaris): Discrimination of song segments, their segregation from multiple and reversed conspecific songs, and evidence for conspecific song categorization," J. Comp. Psychol. **111**, 337–350.

Yokoyama, K., Dailey, D., and Chase, S. (**2006**). "Processing of conflicting and redundant stimulus information by pigeons," Learning & Behavior **34**, 241–247.

Zentall, T. R. (**2005**). "Selective and divided attention in animals," Behav. Processes **69**, 1–15.

Discrimination of complex tones with unresolved components using temporal fine structure information

Brian C. J. Moore,[a] Kathryn Hopkins, and Stuart Cuthbertson
Department of Experimental Psychology, University of Cambridge, Downing Street,
Cambridge CB2 3EB, England

(Received 1 September 2008; revised 12 December 2008; accepted 25 February 2009)

The information used to discriminate complex tones with (largely) unresolved components was assessed. In experiment 1, subjects discriminated a harmonic complex tone, H, with fundamental frequency F0, from an inharmonic tone, I, in which all components were shifted upwards by the same amount in hertz. Tones H and I had the same envelope repetition rate but different temporal fine structure (TFS). The tones were passed through a fixed bandpass filter centered on harmonic N, to reduce excitation pattern cues. For all F0s (35–400 Hz), performance decreased as N was increased from 11 to 15, but, except for F0=35 Hz, remained above chance for $N=15$, where all harmonics should be unresolved. This suggests that discrimination can be based on TFS rather than on partially resolved components. In experiment 2, subjects discriminated the F0 of complex tones filtered as in experiment 1. Here, both envelope rate and TFS cues were available. Except for F0 =35 Hz, discrimination thresholds, expressed as the Weber fraction for a change in time interval, were similar to those measured in experiment 1, suggesting that performance in experiment 2 was dominated by the use of TFS rather than envelope cues.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106135]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Nm [MW] Pages: 3214–3222

I. INTRODUCTION

Complex tones containing only high, unresolved harmonics (roughly the ninth or above) are capable of evoking a pitch, even though that pitch is usually somewhat less clear than when lower harmonics are present (Moore and Rosen, 1979; Houtsma and Smurzynski, 1990). The mechanisms underlying the perception of the pitch of such tones remain somewhat controversial. Some authors have argued that the pitch is largely based on envelope cues (Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994; Bernstein and Oxenham, 2003). Others have proposed that the pitch is extracted using temporal fine structure (TFS) information (de Boer, 1956; Schouten *et al.*, 1962; Moore and Moore, 2003; Moore *et al.*, 2006a; Hopkins and Moore, 2007), as represented in the patterns of phase locking in the auditory nerve. According to the latter view, pitch is determined using the time intervals between peaks in the TFS close to adjacent envelope maxima.

To assess the role of TFS, de Boer (1956) and Schouten *et al.* (1962) obtained pitch matches to "frequency-shifted" inharmonic complex tones (I), derived from harmonic tones (H) by shifting each component upwards (or downwards) by the same amount in hertz. The envelope repetition rate is the same for the I and H tones, but the time intervals between peaks in the TFS are smaller (or larger) for the I tones. The matched pitch was found to shift with the shift of the component frequencies, suggesting that TFS plays an important role in the pitch perception of complex tones. However, Moore and Moore (2003) argued that these results might have been influenced by shifts in the spectrum or excitation pattern produced by the frequency shift of the components. To eliminate this effect, Moore and Moore (2003) used similar stimuli to those of de Boer (1956) and Schouten *et al.* (1962), but both the test and matching tones were passed through a fixed bandpass filter, centered on harmonic N. When no resolved components were present, the test and matching tones evoked very similar excitation patterns, as calculated using the method described by Moore *et al.* (1997). Moore and Moore (2003) included complex tones with components that were clearly unresolved ($N=16$) or mostly unresolved ($N=11$) (Moore *et al.*, 2006b). Pitch shifts were found for $N=11$, suggesting that information from TFS influences the pitch of complex tones with intermediate harmonic numbers. However, for $N=16$, no pitch shifts were found. This suggests that only envelope cues were used to determine the pitch in this case. Moore and Moore (2003) proposed that TFS information contributes to pitch provided that some harmonics with numbers of 14 or below are present and audible.

This work was extended by Hopkins and Moore (2007). They used stimuli similar to those of Moore and Moore (2003), but the task was to discriminate the H and I tones, in a three-alternative forced-choice task. In one of their experiments, the components of the tones had random starting phases, to prevent the use of possible cues relating to differences in envelope shape for the H and I tones. For $N=11$, discrimination performance was good for F0s of 100, 200, and 400 Hz. However, performance did not differ significantly from chance for $N=18$. This pattern of results was consistent with the conclusion of Moore and Moore (2003) that TFS information contributes to pitch when some har-

[a] Author to whom correspondence should be addressed. Electronic mail: bcjm@cam.ac.uk

monics with numbers of 14 or below are present, but that TFS cues cannot be used when only harmonics above the 14th are present.

An alternative explanation for the pitch shifts observed by Moore and Moore (2003) when $N=11$ is that subjects matched partially resolved harmonics in the test and matching stimuli rather than using TFS cues. Similarly, the good performance found by Hopkins and Moore (2007) in their discrimination task when $N=11$ might have been based on detecting changes in the frequencies of partially resolved harmonics. Several researchers have explored the limits of resolvability for components in complex tones with equal-amplitude components, and have concluded that, for harmonic complexes, only the lowest 5–8 harmonics can be "heard out" (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Moore *et al.*, 2006b). This is consistent with the "rule" that components need to be separated by about 1.25ERB$_N$ to be heard out with 75% accuracy (Moore, 2003), where ERB$_N$ stands for the equivalent rectangular bandwidth of the auditory filter for normally hearing listeners as measured at low sound levels (Glasberg and Moore, 1990). However, Bernstein and Oxenham (2003) used a method in which the "target" harmonic was pulsed on and off and found that harmonics up to the 10th or 11th may be heard out. Although the validity of this method has been questioned (Hartmann and Goupell, 2006; Moore *et al.*, 2009), the harmonic number of the highest resolvable harmonic remains in some doubt.

In the present experiment, we used the task of Hopkins and Moore (2007), as described earlier, to clarify whether the ability to discriminate H and I tones depends on the use of TFS or on the discrimination of partially resolved harmonics. While Hopkins and Moore (2007) used values of N of 7, 11, and 18, we used values of 11, 13, and 15. With $N=15$, the lowest component within the passband of the bandpass filter was the 13th, and the lowest audible component was the 11th (given the background noise that was used: See later for details). Thus all audible components should have been unresolved, even using the limit of resolvability proposed by Bernstein and Oxenham (2003). Hence, if above-chance performance for discrimination of the H and I tones was found with $N=15$, this would support a role for TFS in the discrimination of such tones. On the other hand, if performance fell to chance for $N=15$, this would suggest that performance for lower values of N might have been based on the discrimination of partially resolved harmonics.

The lowest F0 used by Hopkins and Moore (2007) was 100 Hz. We were interested in the possible role of TFS for lower F0s. Since the value of ERB$_N$ increases at low center frequencies when expressed as a proportion of the center frequency (Glasberg and Moore, 1990), fewer harmonics should be resolved for very low F0s. Probably, only the lowest 2–4 harmonics are resolved. Some data suggest that the dominant region for pitch shifts upwards, when expressed in terms of harmonic number, when the F0 becomes very low (Moore and Glasberg, 1988; Miyazono *et al.*, 2009). This means that the harmonics in the dominant region are probably unresolved, implying a greater reliance on temporal information from unresolved harmonics. Hence, we included F0s of 50 and 35 Hz. The latter value is close to the lower limit for pitch (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). For the F0s of 35 and 50 Hz, we also used $N=7$ and 9, for which all components should have been unresolved.

A second issue addressed by this study was the extent to which envelope cues contribute to F0 discrimination when both envelope and TFS cues are available. To address this issue, a second experiment was conducted in which the task was to discriminate the F0s of two harmonic complex tones, which were bandpass filtered in a similar way to the H and I tones used in experiment 1. Comparison of the results for the two experiments allowed us to determine whether the extra envelope cues in experiment 2 led to better performance than when only TFS cues were available, as in experiment 1.

II. EXPERIMENT 1: DISCRIMINATION OF HARMONIC AND INHARMONIC TONES

A. Subjects

Four subjects with normal hearing were tested, aged 19–25 years. Two had some musical training. The test ear was selected as the one with the lowest audiometric thresholds, averaged over 500, 1000, and 2000 Hz. All subjects had absolute thresholds less than 15 dB hearing loss (HL) in their test ears at all of the standard audiometric frequencies. All subjects were given practice until their performance appeared to be stable, which took about 3 h.

B. Stimuli

Nominal F0s of 35, 50, 100, 200, and 400 Hz were used. Discrimination was measured with $N=11$, 13, and 15 for all F0s. For the two lowest F0s, N was also set to 7 and 9. A trial consisted of three intervals, two containing a harmonic tone (H) and one containing an inharmonic frequency-shifted tone (I), where all components were shifted upwards by ΔF Hz. The tones H and I were initially generated with many components. The tones were passed though a fixed bandpass filter with a central flat region with a width of 5F0 and skirts that decreased in level at a rate of 30 dB/octave. The lowest component within the passband had a harmonic number equal to $N-2$. The use of a relatively shallow slope ensured minimal changes in the excitation pattern as components moved in and out of the passband (Hopkins and Moore, 2007). The starting phases of the components were chosen randomly for each tone. This meant that the shape of the envelope differed randomly from one tone to the next. However, the envelope repetition rate was always equal to F0, regardless of whether the tone was harmonic or inharmonic. The H and I tones were most different when ΔF=0.5F0. Hence, the value of ΔF was limited to this value. The overall level of the bandpass-filtered complex was set to 65 dB sound pressure level (SPL).

Figure 1 shows samples of waveforms of the stimuli after passing through a simulated auditory filter centered at 1000 Hz. The value of F0 was 100 Hz and the bandpass filter was centered at 1100 Hz ($N=11$). Note that the envelope shape and modulation depth differ across samples, depending on the specific selection of random starting phases. However, on average, the modulation depth is the same for the H and I

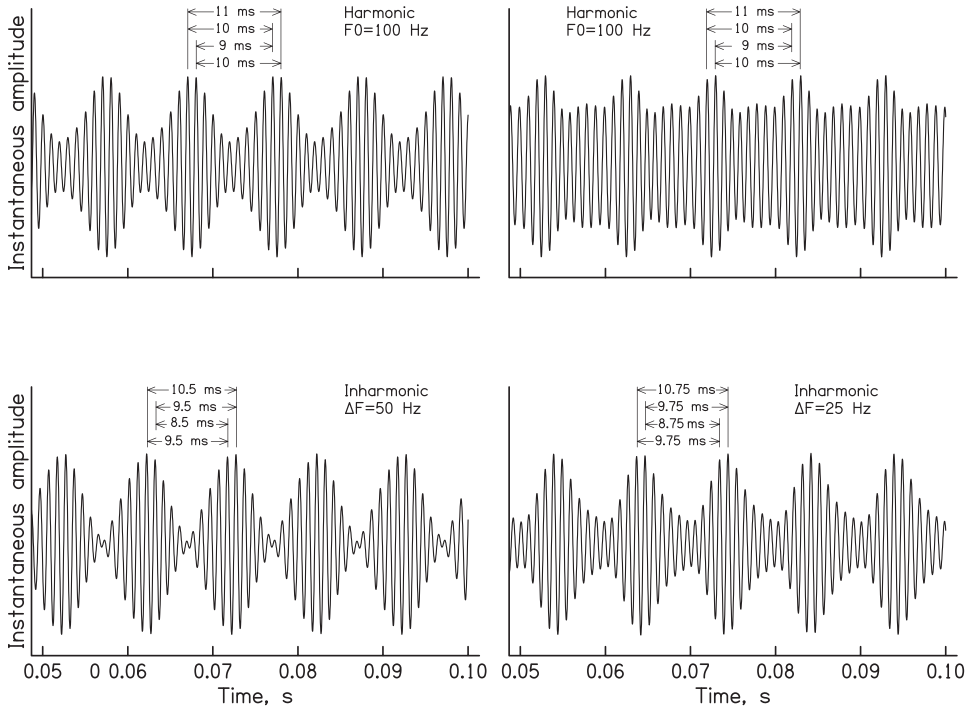


FIG. 1. Samples of waveforms of the harmonic (H) and inharmonic (I) stimuli after passing through a simulated auditory filter centered at 1000 Hz. The value of F0 was 100 Hz and the bandpass filter was centered at 1100 Hz ($N=11$). In the two top panels, the samples are from H tones. The lower-left and lower-right panels show samples from I tones with ΔF=50 and 25 Hz, respectively. The times between peaks in the TFS close to adjacent envelope maxima are indicated. Note that the exact envelope shape and modulation depth depend on the random selection of starting phases, and vary from one tone to the next.

tones. In the two top panels, the samples are from H tones and the times between peaks in the TFS close to adjacent envelope maxima are 9, 10, and 11 ms. Of all the possible inter-peak times in the TFS, the interval 10 ms occurs most often. The TFS repeats every 10 ms, and the position of TFS peaks relative to the envelope maximum remains constant from one envelope period to the next. In the lower-left panel, the sample is from an I tone with ΔF=50 Hz. The times between peaks in the TFS close to adjacent envelope maxima are 8.5, 9.5, and 10.5 ms, the most common interval being 9.5 ms. For the I tone, the position of TFS peaks relative to the envelope maximum changes from one envelope period to the next. In the lower-right panel, the sample is from an I tone with ΔF=25 Hz. The times between peaks in the TFS close to adjacent envelope maxima are 8.75, 9.75, and 10.75 ms, the most common interval being 9.75 ms. These differences in the inter-peak intervals in the TFS for the H and I tones provide a possible basis for the discrimination of the tones and for the pitch difference between them.

Threshold equalizing noise (TEN) (Moore *et al.*, 2000) was used to mask combination tones and to mask components of the H and I tones that fell well outside the passband of the bandpass filter. The TEN level at 1 kHz, expressed in dB/ERB$_N$, was set 20 dB below the level of the most intense component in the complex tones. We calculated that, at this level, the lowest component that would be (just) audible in the background noise would have a harmonic number equal to $N-4$.

C. Signal generation

Stimuli were generated using a Tucker-Davies Technologies (TDT) system II, via a 16-bit digital-to-analog converter (TDT DA4) with a sampling rate of 50 kHz. The levels of the tones and the TEN were controlled independently using two TDT PA4 attenuators and stimuli were low-pass filtered at 20 kHz using Kemo (VBF8) filters with a slope of 96 dB/octave. Stimuli were presented via one earpiece of a Sennheiser HD580 headset. Subjects were seated in a double-walled sound-attenuating chamber.

D. Procedure

A trial consisted of three successive stimuli, indicated by lights on a response box. Each stimulus was 540 ms long including 20-ms raised-cosine onset and offset ramps. The inter-stimulus interval was 200 ms. Two intervals contained the H tone and the third, chosen at random, contained the I tone. Subjects were instructed to press the button corresponding to the interval that sounded different. Feedback was given after every trial via lights on the response box.

For most conditions, an adaptive procedure was used to estimate the value of ΔF required for threshold. However, for some conditions, this was not possible, since the procedure called for a value of ΔF larger than 0.5F0. Table I lists the subjects and conditions for which the adaptive procedure could not be used. For conditions where the adaptive procedure could not be used, performance was measured with ΔF fixed at 0.5F0. For the adaptive procedure, the value of ΔF was increased by a factor K after one incorrect response and

TABLE I. Subject identifiers (1, 2, 3, and 4) of subjects who were unable to complete the adaptive procedure for each F0 and each N.

F0	N				
(Hz)	7	9	11	13	15
35	134	1234	1234	1234	1234
50		4	14	134	134
100			1		134
200				2	123
400				1	1234

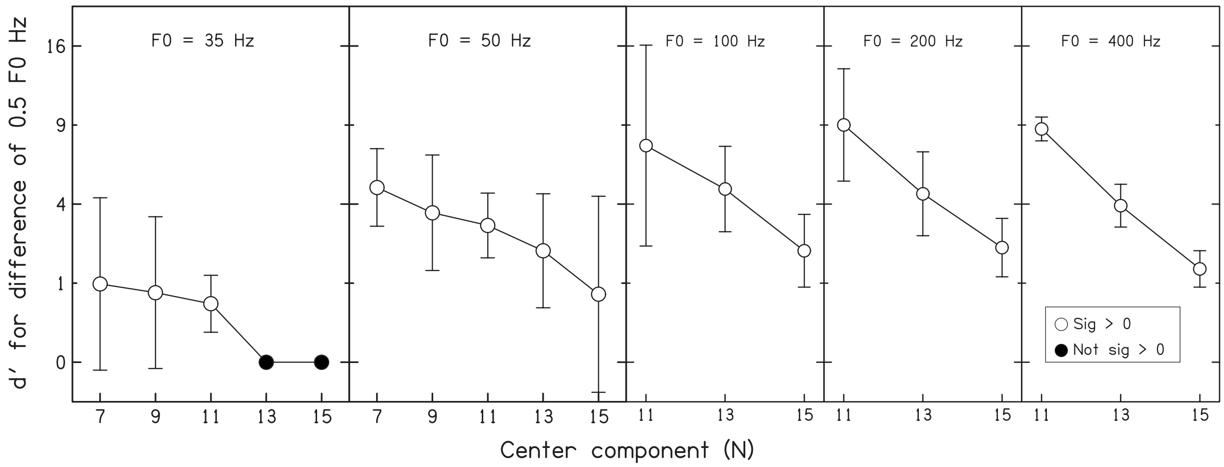


FIG. 2. Results of experiment 1, showing the value of the detectability index d' required to discriminate a frequency shift, ΔF, equal to 0.5F0, plotted on a square-root axis, as a function of the harmonic number, N, on which the bandpass filter was centered. Each panel shows results for one F0. Open symbols indicate d' values significantly above zero, while filled symbols indicate d' values not significantly above zero. Error bars indicate standard deviations across subjects.

was decreased by the same factor after three consecutive correct responses. For the first four turnpoints, K equaled 1.414 and for the last eight turnpoints, it was reduced to 1.189. Twelve turnpoints were obtained and the geometric mean of the values of ΔF at the last eight was taken to be the "threshold" corresponding to a d' value (Green and Swets, 1974) of 1.63. The standard deviation of the logarithms of the turnpoint values was also calculated. If this standard deviation was greater than 0.2, the results of the run were discarded and the condition was repeated. Each condition was tested three times and the geometric mean of the threshold estimates was calculated.

For the procedure with ΔF fixed at 0.5F0, a run consisted of 55 trials, and the last 50 trials were used to calculate a percent-correct value. Subjects completed three non-adaptive runs. The mean of the percent-correct values for each run was used to estimate the final percent-correct value.

E. Statistics

Percent-correct values from the non-adaptive procedure were converted to d' values (Green and Swets, 1974). To allow comparison of these values with ΔF values from the adaptive procedure, the ΔF values were used to calculate the d' value that would be obtained if ΔF were set to 0.5F0, assuming that d' was proportional to ΔF. This was done by dividing 1.63 (the d' value tracked by the adaptive procedure) by the threshold measured in the adaptive procedure, and multiplying this value by 0.5F0. A similar comparison method was used by Hopkins and Moore (2007). This method often yielded values of d' that were much larger than would normally be encountered as, for most conditions, a frequency difference of 0.5F0 Hz was much larger than the threshold value. Such large values of d' would be difficult or impossible to measure in practice, and should not be taken too literally. The main point here is that the calculated d' values are inversely proportional to the estimated threshold values, so that large d' values indicate low thresholds.

Statistical tests were performed on the square root of the absolute d' values, as this transformation gave roughly uniform variability across conditions. Values that were negative before the transformation (which happened only for one subject for F0=35 Hz) were multiplied by -1 after the transformation to restore their sign.

F. Results

To average d' values across subjects, the square root of the individual values was determined, the resulting values were averaged, and the outcome was squared. The results of this averaging process (with error bars indicating standard deviations across subjects) are shown in Fig. 2. Values of d' are plotted on a square-root scale. For all values of F0, the d' values decreased with increasing N, but for F0=100, 200, and 400 Hz, the mean d' values remained well above 1 even for the highest value of N (N=15). The variability across subjects, as indicated by the error bars, was greater for the lower F0s than for the higher F0s. This is consistent with the finding of Moore and Sek (2009), using a similar test of sensitivity to TFS, that all subjects could perform the task for F0s of 100 and 200 Hz, but that for F0=50 Hz only about one-half of their subjects could perform the task. Moore and Sek (2009) also found less individual variability for F0 =400 Hz than for F0=50, 100, or 200 Hz, except for one subject who could not perform the task for F0=400 Hz.

A within-subjects analysis of variance (ANOVA) was performed with factors F0 and N (values of 11, 13, and 15 only). The effect of F0 was significant, $F(4,12)=14.66$, $p<0.001$. *Post hoc* tests, based on Fisher's least-significant-difference test, showed that the mean d' value was significantly lower for F0=35 Hz than for all other F0s (all $p<0.01$). Also, the mean d' value was significantly lower for F0=50 Hz than for F0s of 100, 200, and 400 Hz. The d' values did not differ significantly for F0s of 100, 200, and 400 Hz. The effect of N was also significant, $F(2,6)=33.88$, $p<0.001$. *Post hoc* tests showed that performance worsened progressively with increasing N; all pairwise comparisons were significant at $p<0.01$. The interaction between F0 and N was not significant, $F(8,24)=1.82$, $p=0.123$. A second ANOVA was performed using the data for

F0s of 35 and 50 Hz only, but including data for $N=7$, 9, 11, 13, and 15. Again the effect of F0 was significant, $F(1,3)=17.45$, $p=0.025$, as was the effect of N, $F(4,12)=11.02$, $p<0.001$. The interaction was not significant, $F(4,12)=0.33$, $p=0.855$.

To assess whether specific d' values were significantly greater than 0, corresponding to chance performance, the procedure described by Hopkins and Moore (2007) was used. The d' values were expressed on a square-root scale and the mean and standard error (SE) of these converted values across repeated runs were calculated. If the mean was more than 2 SEs, the mean was taken as being significantly greater than 0. Only two values were not significantly greater than 0, those for F0=35 Hz with $N=13$ and 15. These are indicated by filled circles without error bars in Fig. 1. For F0=100, 200, and 400 Hz, the d' values remained above 1.4 even for $N=15$. It seems highly unlikely that the complex tones contained any resolved components with $N=15$, since the lowest audible component would have been the 11th. Thus, the results support the idea that the H and I tones can be discriminated on the basis of differences in their TFS, when information from partially resolved components is minimal.

Low-numbered harmonics usually give rise to distinct peaks in the excitation pattern evoked by a complex tone. As harmonic number increases, the peaks and dips (ripples) in the excitation pattern become less distinct (Moore, 2003), and this presumably corresponds to the reduced ability to resolve or hear out the higher harmonics. The ripple depth required for detection of ripples in a limited frequency region is 2–3 dB (Eddins and Bero, 2007). For medium F0s, the 2-dB limit is reached at about the sixth harmonic, which is consistent with the idea that only the lowest 5–8 harmonics can be heard out (Plomp, 1964; Plomp and Mimpen, 1968). However, it remains possible that, for values of N below 15, the H and I tones were discriminated on the basis of local differences in excitation level.

To assess this possibility, excitation patterns (Moore *et al.*, 1997) were calculated for the H and I tones for all values of F0 and N, with ΔF for the I tones set to the mean value required for threshold ($d'=1.63$). For conditions where the threshold could not be determined for some subjects with the adaptive procedure, the results for the subjects who *could* complete the adaptive procedure were expressed as the d' value that would be obtained for ΔF=0.5F0. The d' values were then averaged across all subjects. Where the resulting mean was greater than 1.63, the value of ΔF required for $d'=1.63$ was calculated and used in determining the excitation patterns. Where the resulting mean was less than 1.63, the value of ΔF was set to 0.5F0, which gives the largest possible difference between the H and I tones. The patterns were calculated including the effects of the background TEN that was used in the experiment. The largest differences between the excitation patterns for the H and I tones are given in Table II. Note that performance for discriminating the H and I tones was above chance for all conditions except for F0=35 Hz with $N=13$ and 15.

The differences in excitation level are generally small, most of the values being below 1 dB. The differences are nearly all markedly smaller than the smallest detectable change in excitation level over a restricted region of the excitation pattern, which is typically 2–5 dB (Moore *et al.*, 1989; Moore and Sek, 1994; Buus and Florentine, 1995). This suggests that the H and I tones were not discriminated on the basis of differences in their excitation patterns. It is noteworthy that, for F0=35 Hz and $N=15$, a difference in excitation level of 0.5 dB was associated with chance performance. In this case, the maximum difference in excitation level occurred at 451 Hz, in the region of the 13th harmonic (the lowest harmonic within the passband of the bandpass filter). For F0=50 Hz and $N=11$, the maximum difference in excitation level was only slightly larger (0.6 dB) and occurred in a similar frequency region (444 Hz, corresponding roughly to the frequency of the ninth harmonic, again the lowest within the passband of the bandpass filter). Yet, in this case, discrimination of the H and I tones was well above chance. This comparison suggests that differences in the excitation patterns of the H and I tones did not provide a basis for discrimination.

TABLE II. Maximum differences in excitation level in decibels between the excitation patterns evoked by the H and I tones when the value of ΔF for the I tone was equal to the mean value required for threshold ($d'=1.63$) or to 0.5F0 (see text for details). The effect of the TEN background was included in the calculations.

F0	*N*				
(Hz)	7	9	11	13	15
35	2.0	1.3	0.9	0.5	0.5
50	1.2	0.8	0.6	0.5	0.5
100			0.4	0.4	0.6
200			0.6	0.7	0.8
400			0.7	1.0	1.2

III. EXPERIMENT 2: DISCRIMINATION OF THE F0 OF BANDPASS-FILTERED HARMONIC COMPLEX TONES

This experiment used similar stimuli and procedures to experiment 1, except that all tones were harmonic, and the task was to discriminate the F0 of the tones. Thus, the major difference from experiment 1 was that both envelope and TFS cues could be used to perform the task.

A. Subjects, stimuli, and procedure

Four normal-hearing subjects were tested, aged 20–25 years, one of whom also took part in experiment 1 (the other subjects were no longer available). Three of the subjects were musically trained. The test ear was selected as the one with the lowest audiometric thresholds, averaged over 500, 1000, and 2000 Hz. All subjects had absolute thresholds less than 10 dB HL in their test ears at all of the standard audiometric frequencies. Each subject was trained for 3 h on a subset of conditions, after which performance appeared to be stable.

Harmonic tones were used, with nominal F0=35, 50, and 100 Hz. The tones were generated and bandpass filtered in the same way as for experiment 1. The starting phases of

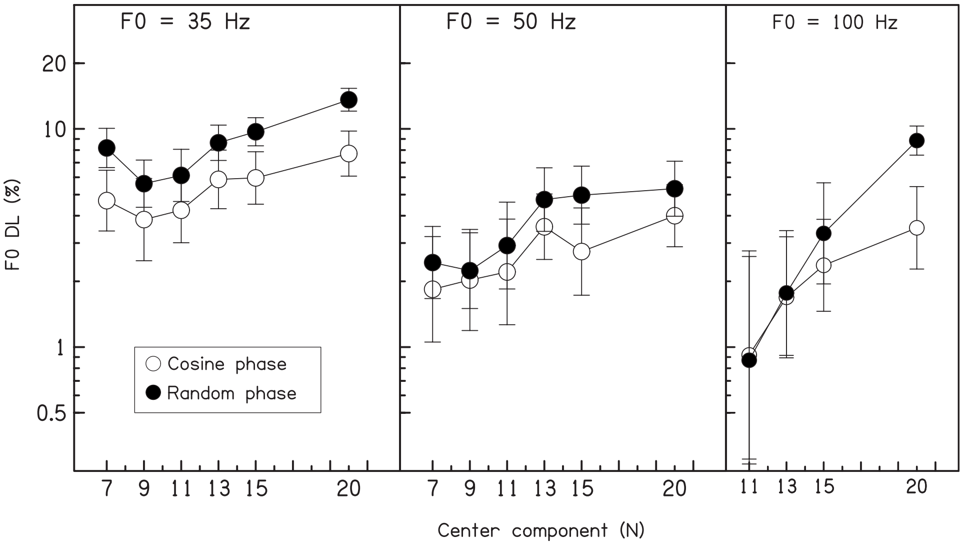


FIG. 3. Results of experiment 2, showing F0 DLs as a function of the harmonic number, N, on which the bandpass filter was centered. Each panel shows results for one F0. Open and filled symbols show F0 DLs for cosine- and random-phase stimuli, respectively. Error bars indicate standard deviations across subjects.

the components were either chosen randomly (as in experiment 1) or were set to 90 deg. The latter phase was used since it leads to a waveform on the basilar membrane with a high peak factor, which would be expected to promote the use of envelope cues. Also, when F0 discrimination is influenced by component phase, this is usually taken as supporting the idea that the components determining performance are unresolved (Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994; Bernstein and Oxenham, 2005; Moore *et al.*, 2006a). Values of N of 11, 13, 15, and 20 were used for all F0s. In addition, values of N of 7 and 9 were used for F0=35 and 50 Hz. The timing of the stimuli and the three-alternative forced-choice procedure were the same as for experiment 1. Two intervals contained a tone with F0 equal to the nominal value, and the other interval contained a tone whose F0 was increased by ΔF0. Subjects were asked to indicate the interval that sounded different from the other two, and were told that usually the tone in that interval would sound higher in pitch than the tones in the other intervals. The value of ΔF0 required for threshold (79% correct, $d'=1.63$) was estimated using the same adaptive procedure as described for experiment 1. A background TEN was used, and levels of the tests tones and TEN were the same as for experiment 1.

B. Results

The geometric mean thresholds, expressed as ΔF0/F0 in percent, are shown in Fig. 3. We refer to these thresholds as F0 difference limens (DLs). For F0=100 Hz, the F0 DLs worsened progressively with increasing N, from 11 to 20. There was no effect of component phase for $N=11$ or 13, but random phase led to larger F0 DLs than cosine phase for $N=15$ and 20. A within-subjects ANOVA based on the logarithms of the F0 DLs for F0=100 Hz alone, with factors N and component phase, showed a significant effect of N, $F(3,9)=40.24$, $p<0.001$; a significant effect of phase, $F(1,3)=63.59$, $p=0.004$; and a significant interaction, $F(3,9)=7.39$, $p=0.008$.

For F0=50 Hz, the F0 DLs were approximately the same for $N=7$ and 9, increased as N was increased further up to 13, and then flattened off. Random phase led to higher F0 DLs for cosine than for random phase for all N. A within-subjects ANOVA showed a significant effect of N, $F(5,15)=4.02$, $p=0.016$; a significant effect of phase, $F(1,3)=12.1$, $p=0.04$; but no significant interaction, $F(5,15)=0.88$, $p=0.517$.

For F0=35 Hz, F0 DLs initially decreased slightly as N was increased from 7 to 9, and then increased as N was increased to 20. Random phase led to higher F0 DLs than cosine phase for all N. A within-subjects ANOVA showed an effect of N that just failed to reach significance, $F(5,15)=2.78$, $p=0.057$; a significant effect of phase, $F(1,3)=80.08$, $p=0.003$; and no significant interaction, $F(5,15)=0.34$, $p=0.879$.

The fact that there was a phase effect for all values of N for F0=35 and 50 Hz is consistent with the argument made earlier, that even for $N=7$, the components that determined the F0 DLs were unresolved. However, the phase effect was absent for $N=11$ and 13 when F0=100 Hz. While the presence of a phase effect implies that the components determining performance were unresolved, the absence of a phase effect does not necessarily mean that the components were resolved. The use of TFS to detect a difference between the H and I tones requires that the envelope has some degree of modulation, since the pitch of the tones is assumed to be based on the time intervals between peaks in the TFS close to adjacent envelope maxima. If the peaks in the envelope are not distinct, the auditory system will not "know" which intervals between TFS peaks to use. However, even random phase can lead to a waveform on the basilar membrane with a distinct envelope periodicity, as illustrated in Fig. 1. This envelope structure may be sufficient to allow the effective use of TFS information. Thus, increasing the envelope modulation depth (via the use of cosine or sine starting phase) may not always result in an increase in the effectiveness of TFS cues.

IV. COMPARISON OF RESULTS FOR EXPERIMENTS 1 AND 2

To assess whether the F0 DLs measured in experiment 2 were determined by the use of TFS cues, envelope cues, or a

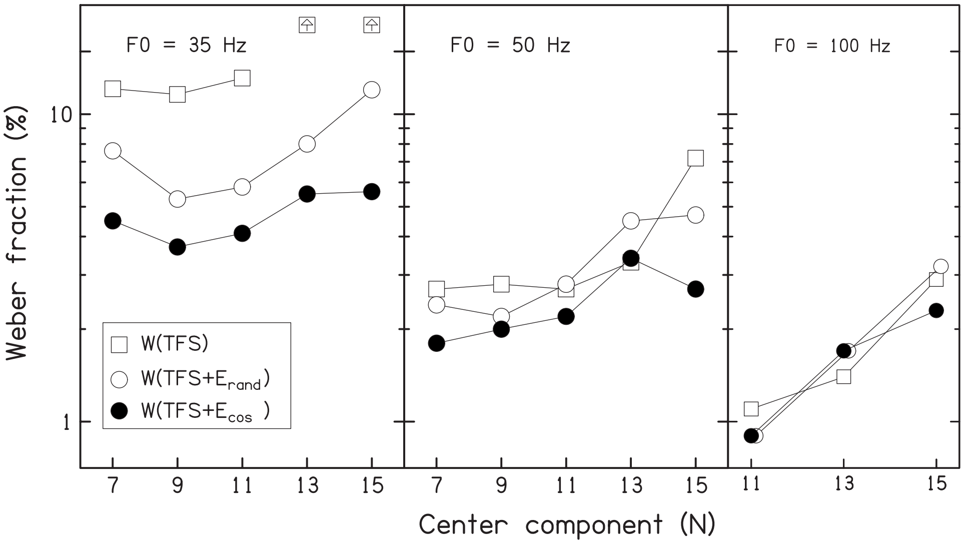


FIG. 4. Comparison of results for experiments 1 and 2. All scores were converted to Weber fractions for time-interval discrimination (see text for details). Weber fractions when only TFS cues were available (experiment 1) are shown by open squares. Weber fractions when both TFS and envelope cues were available (experiment 2) are shown by open circles (random-phase stimuli) and filled circles (cosine-phase stimuli). Weber fractions are plotted as a function of the harmonic number, N, on which the bandpass filter was centered. Each panel shows results for one F0. Up-pointing arrows in the panel for F0=35 Hz indicate cases where performance was too poor for the Weber fraction to be determined.

combination of the two, we compared the results for experiments 1 and 2. To express the results in a comparable way for the two experiments, all thresholds were expressed in terms of Weber fractions for time-interval discrimination. We assumed that, in all cases, performance was determined using the output of an auditory filter centered a little below the center of the passband of the stimuli, at the frequency corresponding to $N-1$. For example, for F0=100 Hz and $N=11$, we assumed that the filter was centered on the tenth harmonic, i.e., at 1000 Hz. For experiment 1, the relevant time intervals can be illustrated using Fig. 1. We take as a "reference" the time interval between corresponding peaks in the TFS for the H tone. This interval is the same as the envelope period, and for F0=100 Hz is equal to 10 ms (top panels of Fig. 1). The closest interval in the TFS of the just-discriminable I tone was taken as the comparison interval. For example, if a frequency shift of 25 Hz could just be detected ($d'=1.63$), then the closest interval in the TFS was taken as 9.75 ms (see the bottom-right panel of Fig. 1). The threshold was expressed as a Weber fraction in percent: 100× (reference interval–comparison interval)/(reference interval). Thus, for the example given above, the threshold would be $100(0.25/10)=2.5\%$. We denote the Weber fraction for stimuli where TFS but not envelope cues were available (experiment 1) as $W(\text{TFS})$.

The thresholds for experiment 2 were expressed in a similar way. Say, for example, that the F0 DL was 5% for a specific N for F0=100 Hz. This would mean that a period of 10 ms (corresponding to 1/100 Hz) could just be discriminated from a period of 9.52 ms (corresponding to 1/105 Hz). Thus, the Weber fraction is $100(10-9.52)/10=4.8\%$. We denote the Weber fraction for stimuli where both TFS and envelope cues were available (experiment 2) as $W(\text{TFS}+E_{\text{cos}})$ for the case where components were added in cosine phase and $W(\text{TFS}+E_{\text{rand}})$ for the case where components were added in random phase.

Figure 4 compares the results for experiments 1 and 2, with all thresholds expressed as Weber fractions, as described above. For F0=100 Hz, $W(\text{TFS})$ and $W(\text{TFS}+E_{\text{rand}})$ were similar for all N. $W(\text{TFS})$ and $W(\text{TFS}+E_{\text{cos}})$ were also similar, except perhaps for the slightly lower value of $W(\text{TFS}+E_{\text{cos}})$ for $N=15$. This pattern of results suggests that discrimination performance when both TFS and E cues were available (experiment 2) was dominated by the use of TFS cues, at least for $N=11$ and 13. For F0=50 Hz, $W(\text{TFS})$ and $W(\text{TFS}+E_{\text{rand}})$ were similar for $N=7$, 9, 11, and 13. However, $W(\text{TFS})$ was somewhat larger than $W(\text{TFS}+E_{\text{rand}})$ for $N=15$. The values of $W(\text{TFS}+E_{\text{cos}})$ tended to be below the values of $W(\text{TFS})$, except for $N=13$. This might indicate that envelope cues contributed to performance when both E_{cos} and TFS cues were available. However, the greater envelope modulation depth for the cosine-phase stimuli might also have made TFS cues more effective. Finally, for F0=35 Hz, the values of $W(\text{TFS})$ were clearly larger than the values of $W(\text{TFS}+E_{\text{rand}})$ and $W(\text{TFS}+E_{\text{cos}})$ for all N. This suggests that, for this very low F0, envelope cues contributed significantly to performance when both TFS and envelope cues were available.

In summary, for F0s of 100 and 50 Hz, the Weber fractions were generally similar when only TFS cues were available and when both TFS and envelope cues were available, except perhaps for $N=15$. This suggests that, performance in F0 discrimination (experiment 2) was dominated by the use of TFS cues for these F0s. However, for F0=35 Hz, the Weber fractions were larger when only TFS cues were available than when both TFS and envelope cues were available, suggesting a significant contribution of envelope cues to performance in the F0-discrimination task.

V. GENERAL DISCUSSION

The results of experiment 2 showed an effect of component phase on F0 DLs for all values of N for F0=35 and 50 Hz, and for $N=15$ and 20 for F0=100 Hz. The phase effect found for F0=100 Hz is comparable to that found in several earlier studies, although some studies have shown phase effects for somewhat lower values of N (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2005; Moore *et al.*, 2006a). The phase effects for F0s of 35 and 50 Hz can be compared to those found by Miyazono *et al.* (2009), although in their experiment the harmonics whose F0 was to be discriminated were flanked by harmonics whose frequency remained fixed across the two intervals of a forced-choice trial. Miyazono *et al.* (2009) found an effect of phase for all N (including $N=1$) when F0 was 35 Hz, and an effect of phase for $N>5$ for F0=50 Hz. The phase effect found here can be interpreted as indicating that the components in the tones were unresolved for these conditions. For tones filtered in a similar way, the results showed that subjects could discriminate the H and I tones for N up to 11 for F0 =35 Hz and for N up to 15 for F0=50 and 100 Hz. Taken together, these results indicate that the discrimination of the H and I tones in experiment 1 was largely based on information from the TFS of unresolved components, rather than information from partially resolved components. This conclusion is supported by the calculation of excitation patterns, which showed that the maximum difference in excitation level between the H and I tones when ΔF was set to the threshold value was typically smaller than the smallest detectable change in excitation level over a restricted frequency region.

Comparison of Weber fractions for time-interval discrimination from experiments 1 and 2 suggested that, under conditions where both TFS and envelope cues were available (F0 discrimination, experiment 2), discrimination was primarily based on TFS cues for F0=50 and 100 Hz and for values of N up to 13 or 15. The earlier results of Hopkins and Moore (2007) showed that discrimination of the H and I tones was close to chance for $N=18$ when the starting phases of the components were random (hence that condition was not tested in the current experiments), while the results of experiment 2 showed that F0 discrimination was still possible for random-phase stimuli when $N=20$. Thus, when N is very high (above 15–18) it seems likely that F0 discrimination is based on envelope cues alone, and not on TFS. This is consistent with the suggestion of Moore and Moore (2003) that TFS can only be used to discriminate complex tones when harmonics below about the 14th are present and audible.

Performance for F0=35 Hz was generally poorer than for higher F0s, and in experiment 1 d' values were at chance for $N=13$ and 15. One subject in that experiment performed at close to chance levels for all N with F0=35 Hz. The generally poor performance may occur because the auditory system has difficulty in accurately estimating long interspike intervals, associated with low F0s. For F0=35 Hz, the time interval between adjacent envelope maxima is about 29 ms. de Cheveigné and Pressnitzer (2006) proposed a mechanism by which the delays required to measure long intervals are synthesized from cross-channel phase interaction. The maximum duration of the synthetic delays is limited by the duration of the impulse responses of the cochlear filters. For $N=15$, the relevant filters would be centered close to 525 Hz. The value of $\mathrm{ERB_N}$ at this center frequency is about 81 Hz. It is unlikely that a filter with this bandwidth would ring for as long as 29 ms, since the duration of the ringing is roughly the reciprocal of the bandwidth. This could account for the relatively poor performance with $N=13$ and 15 for F0=35 Hz.

With F0=35 Hz, the Weber fractions when both TFS and envelope cues were available (experiment 2) were consistently lower than the Weber fractions when TFS cues alone were available (experiment 1). This suggests that envelope cues contribute substantially to F0 discrimination at this low F0. Despite this, performance based on TFS cues alone with F0=35 Hz was above chance for $N=7$, 9, and 11. For $N=7$, the lowest audible component was probably the fourth (140 Hz), and this is separated from the next highest component by only $0.84\mathrm{ERB_N}$. Thus it is unlikely that any audible components were resolved for $N=7$, and even less likely that any components were resolved for $N=9$ and 11. The argument that the components were unresolved is consistent with the phase effect that was observed for all N in experiment 2. Thus, the above-chance performance for N =7, 9, and 11 in experiment 1 is most readily explained in terms of a sensitivity to TFS.

For the F0 of 400 Hz, the passband with $N=15$ was centered at 6000 Hz, and the lowest component within the passband fell at 5200 Hz. Initially, it may seem surprising that the H and I tones could be discriminated on the basis of their TFS at such high frequencies, since phase locking is known to be weak for frequencies above 5000 Hz, at least in animals (Palmer and Russell, 1986). However, residual phase locking occurs for frequencies up to 10,000–12,000 Hz (Heinz *et al.*, 2001; Recio-Spinoso *et al.*, 2005). Indeed, Heinz *et al.* (2001) suggested that the frequency discrimination of pure tones with frequencies from 5 to 10 kHz might be based on the use of residual phase-locking information rather than place information. Residual phase locking at high frequencies may allow reasonable discrimination based on the TFS of complex tones, although it may allow only relatively poor frequency discrimination of pure tones.

VI. CONCLUSIONS

For F0s of 50, 100, 200, and 400 Hz, the discriminability of harmonic (H) and inharmonic (I) frequency-shifted tones, bandpass filtered with a filter centered on harmonic N and with components added with random starting phases, decreased as N was increased up to 15. However, discrimination performance remained well above chance when N was 15. It is unlikely that any resolved harmonics would be audible for such a high N, suggesting that the tones with this N were discriminated on the basis of their TFS rather than by discriminating the frequencies of partially resolved harmonics. Even for lower values of N, the differences in excitation patterns between the H and I tones when the frequency shift was set to the threshold value were probably too small to be detected, again suggesting that discrimination was based on TFS.

For F0=35 Hz, discrimination of the H and I tones was possible only for N up to 11. However, even in this case, it seems likely that there were no audible resolved harmonics, and performance was based on the use of TFS information.

Measures of the discrimination of the F0 of harmonic complex tones, bandpass filtered in a similar way, showed that performance was affected by the starting phase of the components (cosine versus random) for all N for F0=35 and 50 Hz, and for $N=15$ and 20 for F0=100 Hz, suggesting that the components were unresolved for these conditions.

When performance was expressed as the Weber fraction for time-interval discrimination, performance for F0=50 and 100 Hz was similar when only TFS cues were available (discrimination of H from I tones), and when both envelope and TFS cues were available (F0 discrimination). This suggests that, in the latter case, performance was dominated by the use of TFS cues, except perhaps when *N* was very high.

For F0=35 Hz, performance was worse when only TFS cues were available than when both envelope and TFS cues were available, suggesting that envelope cues contribute markedly to F0 discrimination for this very low F0.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Alexsander Sek and Brian Glasberg for their assistance with this work, and Aleksander Sek and Tom Baer for helpful discussions. We also thank two anonymous reviewers for helpful comments on an earlier version of this manuscript.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Buus, S., and Florentine, M. (**1995**). "Sensitivity to excitation-level differences within a fixed number of channels as a function of level and frequency," in *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oekinghaus (World Scientific, Singapore).

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature (London) **178**, 535–536.

de Cheveigné, A., and Pressnitzer, D. (**2006**). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," J. Acoust. Soc. Am. **119**, 3908–3918.

Eddins, D. A., and Bero, E. M. (**2007**). "Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region," J. Acoust. Soc. Am. **121**, 363–372.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Hartmann, W. M., and Goupell, M. J. (**2006**). "Enhancing and unmasking the harmonics of a complex tone," J. Acoust. Soc. Am. **120**, 2142–2157.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Miyazono, H., Glasberg, B. R., and Moore, B. C. J. (**2009**). "Dominant region for pitch at low fundamental frequencies (F0): The effect of fundamental frequency, phase and temporal structure," Acoust. Sci. & Tech. (in press).

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego, CA).

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments," in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240.

Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006a**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.

Moore, B. C. J., Glasberg, B. R., and Jepsen, M. L. (**2009**). "Effects of pulsing of the target tone on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. (in press).

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006b**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.

Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820–836.

Moore, B. C. J., and Rosen, S. M. (**1979**). "Tune recognition with reduced pitch and interval information," Q. J. Exp. Psychol. **31**, 229–240.

Moore, B. C. J., and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation," J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161–171.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner haircells," Hear. Res. **24**, 1–15.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am. **43**, 764–767.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

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 auditory-nerve fibers," J. Neurophysiol. **93**, 3615–3634.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

Concurrent-vowel and tone recognitions in acoustic and simulated electric hearing

Xin Luo[a)]
Communication and Auditory Neuroscience, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057 and Department of Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, Indiana 47907

Qian-Jie Fu
Communication and Auditory Neuroscience, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 16 July 2008; revised 27 February 2009; accepted 5 March 2009)

Because of the poor spectral resolution in cochlear implants (CIs), fundamental frequency (F0) cues are not well preserved. Chinese-speaking CI users may have great difficulty understanding speech produced by competing talkers, due to conflicting tones. In this study, normal-hearing listeners' concurrent Chinese syllable recognition was measured with unprocessed speech and CI simulations. Concurrent syllables were constructed by summing two vowels from a male talker (with identical mean F0's) or one vowel from each of a male and a female talker (with a relatively large F0 separation). CI signal processing was simulated using four- and eight-channel noise-band vocoders; the degraded spectral resolution may limit listeners' ability to utilize talker and/or tone differences. The results showed that concurrent speech recognition was significantly poorer with the CI simulations than with unprocessed speech. There were significant interactions between the talker and speech-processing conditions, e.g., better tone and syllable recognitions with the male-female condition for unprocessed speech, and with the male-male condition for eight-channel speech. With the CI simulations, competing tones interfered with concurrent-tone and syllable recognitions, but not vowel recognition. Given limited pitch cues, subjects were unable to use F0 differences between talkers or tones for concurrent Chinese syllable recognition.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106534]

PACS number(s): 43.66.Hg, 43.66.Ts, 43.71.Es, 43.71.Ky [RYL] Pages: 3223–3233

I. INTRODUCTION

Auditory sensory inputs are used to identify multiple sound sources within complex listening environments, a phenomenon described by Bregman (1990) as "auditory scene analysis." One of the most challenging listening conditions is simultaneous presentation of competing speech. Segregation and streaming of sound sources allow a target talker to be understood in the presence of competing talkers. Many previous studies have measured identification of two simultaneously presented, synthesized vowel-like sounds, and have shown that the normal auditory system is able to stream and segregate concurrent vowels using acoustic cues such as fundamental frequency (F0) difference, harmonic misalignment, pitch period asynchrony, and formant transitions (e.g., Scheffers, 1983; Assmann and Summerfield, 1990; Summerfield and Assmann, 1991; Assmann, 1995).

For example, Scheffers (1983) and Assmann and Summerfield (1990) showed that normal-hearing (NH) listeners' identification of concurrent vowels was significantly better when the two vowels had different F0's (separated by one to four semitones) rather than the same F0. Based on such perceptual data, computational models have been proposed that involve voice pitch estimation from the output of the auditory periphery, segregation of competing voices according to the estimated voice pitches, and vowel template matching for the segregated spectral patterns (Assmann and Summerfield, 1990; Meddis and Hewitt, 1992). It is possible that when two concurrent vowels have different F0's, listeners may better attend to the components of each vowel by making use of their respective harmonic and periodic structures. For example, Summerfield and Assmann (1991) found that concurrent-vowel recognition was improved by shifting the harmonics of a component vowel by half of its F0 relative to those of the other vowel, as long as the F0 was high enough (e.g., 200 Hz) and the harmonics were well separated in frequency. They also found that shifting the temporal waveforms of a component vowel by half of its period relative to those of the other vowel (i.e., introducing pitch period asynchrony) improved concurrent-vowel recognition, as long as the F0 was very low (e.g., 50 Hz) and the periods were well separated in time.

Frequency modulation (FM) with either linear gliding or sinusoidal functions may be another acoustic cue for auditory grouping and segregation. For example, Culling and Summerfield (1995) found that identification thresholds for a target vowel were lower when the masking vowel was unmodulated rather than modulated in the same way as the target vowel. Chalikia and Bregman (1989) measured recog-

[a)]Author to whom correspondence should be addressed. Electronic mail: luo5@purdue.edu nition of concurrent vowels, whose frequency components were either not modulated, modulated in parallel, linear glides (in log scale), or modulated in opposite, crossing glides. They found that the recognition with crossing glides was significantly better than that with un-modulated stimuli or with parallel linear glides. These results suggest that coherent gliding of frequency components within vowels, as well as incoherent gliding of components across vowels, might benefit segregation and identification of concurrent vowels.

In these above-cited studies, concurrent-vowel recognition has been measured using English vowels. While F0 differences generally benefit concurrent English vowel recognition, identifying F0 variation patterns is not an essential part of English syllable recognition. In contrast, F0 cues are lexically meaningful for tonal languages such as Mandarin Chinese. Mandarin Chinese has four lexical tones: tone 1 (high, flat pitch contour), tone 2 (rising contour), tone 3 (falling-rising contour), and tone 4 (falling contour). Concurrent Chinese syllable recognition provides a unique opportunity to investigate how competing tonal patterns (i.e., pitch contours) may influence concurrent-vowel recognition. Similarly, concurrent-tone recognition can be measured within the same mixture of Chinese syllables. Previous studies (e.g., Chalikia and Bregman, 1989) have largely focused on the contribution of FM cues to concurrent-vowel recognition, and have paid less attention to identification of modulation properties (e.g., frequency glide direction). Because F0 cues are lexically meaningful, it is also important to understand how Chinese listeners' concurrent-syllable recognition is affected by different vowel and tone pairs. Quite possibly, F0 differences and pitch contours may affect concurrent Chinese syllable recognition differently than concurrent English syllable recognition.

Unlike NH listeners, cochlear implant (CI) users have limited access to F0 cues. Contemporary implant systems typically consist of multiple electrodes (16–22) and use speech-processing strategies based on waveform representation (e.g., Wilson *et al.*, 1991). Because of the limited number of electrodes and/or limited channel selectivity, the spectral resolution is too poor to resolve F0 and harmonic information. F0 is well represented in the temporal envelopes within individual frequency channels, but CI users are able to extract only some of the temporal pitch information, and only for relatively low F0's (Green *et al.*, 2004). While other co-varying cues are available (e.g., overall duration and amplitude envelope), CI users' limited pitch perception capabilities result in only moderate levels of performance for Chinese tone recognition (e.g., Fu *et al.*, 2004; Luo *et al.*, 2008). Given the limited Chinese tone recognition, it is unclear how competing tonal patterns may influence CI users' concurrent Chinese syllable recognition.

CI users' poor speech perception in the presence of competing talkers may also be due to limited F0 coding. Qin and Oxenham (2005) found that NH listeners' concurrent-vowel recognition performance with acoustic simulations of CI processing (even with 24 channels) was significantly poorer than with unprocessed speech. In these acoustic CI simulations, NH listeners were unable to use F0 differences between concurrent vowels to achieve better recognition performance. Similar results were reported by Stickney *et al.* (2007), who showed that, for unprocessed speech, increasing the F0 separation between the target and masker sentences gradually improved NH listeners' recognition of target sentences; for CI users or NH listeners listening to acoustic CI simulations, increasing the F0 separation provided no benefit.

In the present study, the effects of CI speech processing on concurrent-vowel and tone recognitions were acoustically simulated in NH listeners. While it might be of more practical interest to investigate target speech recognition in the presence of competing speech, concurrent-vowel and tone recognitions enable detailed analyses of performance and confusion patterns across different vowel and tone pairs, which might provide some insights into the strategies and cues used by NH listeners for sound source segregation in acoustic and simulated electric hearing. In the real CI case, patient-related factors (e.g., etiology of hearing loss, proximity of electrodes to healthy neural populations, and implant device differences) can result in significant inter-subject variability, making it sometimes difficult to measure the effect of a processing parameter change. Acoustic CI simulations allow better control of subject variables within the (presumably) more homogenous group of NH listeners, at least in terms of hearing health. The amount of spectral and temporal cues can be manipulated independently in NH listeners by varying the number of frequency channels and the temporal envelope cutoff frequency, respectively. These manipulations have great relevance for CI users, who must fuse the spectral and temporal cues delivered by electrical stimulation patterns. By using CI simulations in NH listeners, we can more cleanly measure the effect of a processing parameter change that might be important to the real CI case. CI processing was simulated here using a four- or eight-channel noise-band vocoder and a 500-Hz temporal envelope filter in each band. The number of frequency channels and temporal envelope cutoff frequency were chosen to produce overall performance similar to that of CI users with clinically assigned speech processors (e.g., Friesen *et al.*, 2001), and in turn provide relevant implications for real CI speech perception. We hypothesized that increased spectral resolution would improve concurrent-vowel recognition, consistent with previous studies with single syllables (e.g., Xu *et al.*, 2002). However, increased spectral resolution has been shown to have a much less effect on tone recognition than on vowel recognition with single syllables (e.g., Xu *et al.*, 2002). Given the limited F0 cues in CI processing, it is unclear whether increased spectral resolution would improve concurrent-tone recognition. In this study, we explored the contribution of spectral cues (with unprocessed speech, eight- and four-channel CI simulations) to concurrent-vowel and tone recognitions. To investigate the effects of mean F0 difference on concurrent-vowel and tone recognitions, vowels were combined from a male and a female talker to produce a relatively large difference in mean F0, or within the same male talker to produce a nearly identical mean F0.

TABLE I. Ranges for F0, F1, and F2 values for the Chinese single-vowel stimuli, for the male and female talkers.

		Male talker			Female talker		
		F0 range (Hz)	F1 range (Hz)	F2 range (Hz)	F0 range (Hz)	F1 range (Hz)	F2 range (Hz)
/a/	Tone 1	147–155	930–960	1170–1290	277–290	1111–1261	1594–1775
	Tone 2	87–160	960–1050	1200–1322	178–280	1050–1231	1412–1654
	Tone 3	82–137	870–1080	1111–1231	130–202	1200–1260	1654–1805
	Tone 4	82–164	869–1111	1141–1292	129–293	1111–1382	1443–1684
/e/	Tone 1	157–170	356–597	1141–1352	277–296	446–870	1473–1594
	Tone 2	95–175	356–597	1231–1412	194–294	476–839	1352–1563
	Tone 3	82–121	446–658	1141–1292	148–202	507–809	1443–1533
	Tone 4	78–188	416–658	1141–1443	140–333	386–748	1503–1594
/u/	Tone 1	150–161	356–385	718–778	277–304	386–446	627–960
	Tone 2	92–183	356–476	567–809	186–284	386–537	627–960
	Tone 3	78–133	356–385	446–597	138–206	416–446	567–900
	Tone 4	83–206	356–446	446–688	110–320	386–537	658–1020
/i/	Tone 1	157–185	325–386	2348–2530	265–290	295–325	2681–3073
	Tone 2	96–157	235–295	2379–2439	186–266	265–326	3013–3254
	Tone 3	78–127	265–325	2228–2469	156–218	295–416	3194–3254
	Tone 4	78–195	265–356	2379–2439	150–332	356–356	3073–3284

II. METHODS

A. Subjects

Six adult native Chinese-speaking NH subjects (three males and three females) participated in the present study. All subjects had pure-tone thresholds better than 20 dB hearing level (HL) at octave frequencies from 125 to 8000 Hz in both ears. All subjects were very experienced with the acoustic CI simulations from previous experiments.

B. Stimuli and speech processing

While synthesized vowels used in previous studies (e.g., Scheffers, 1983) allow precise control over vowel parameters (e.g., F0 and harmonics, formant frequencies and bandwidths, duration, amplitude, etc.), these parameters interact dynamically in natural speech. In addition, Chinese syllable synthesis faces a special challenge, namely, how to accurately synthesize tones. Most of the current tone models (data-driven or rule-based) work for continuous speech. To the best of our knowledge, there are no standard F0 contour models for generating isolated tones. In the present study, concurrent Chinese syllable recognition was measured using naturally produced vowels. Single-vowel stimuli were drawn from the Chinese Standard Database, recorded by Wang (1993). One male and one female talker each produced four Mandarin Chinese single-vowels (/a/, /e/, /u/, and /i/ in Pinyin) according to four lexical tones—tone 1 (high, flat), tone 2 (rising), tone 3 (falling-rising), and tone 4 (falling)—resulting in a total of 32 single-vowel syllables (4 vowels × 4 tones × 2 talkers). These single-vowel stimuli were digitized using a 16-bit analog/digital converter at a 16-kHz sampling rate, without high-frequency pre-emphasis. Table I shows the ranges for F0, first formant (F1), and second formant (F2) frequencies for the Chinese single-vowel stimuli.

The concurrent Chinese syllables were constructed by summing either one single-vowel syllable from each of the male and female talkers (male-female condition) or two single-vowel syllables from the male talker (male-male condition). The male-female condition provided a relatively large difference in mean F0, while the male-male condition provided a relatively small difference in mean F0. To align the onsets and offsets of the two component vowels, all single-vowel syllables were normalized to have the same time duration (405 ms for the vowel segments). The duration-normalization was performed by time-stretching or -compressing the input vowel duration to 405 ms, without changing the input pitch and formant frequencies, using an algorithm in Adobe Audition. The time scaling factor for individual single-vowel syllables ranged from 0.6 to 1.3. The duration-normalization made duration cues unavailable for tone recognition, forcing subjects to attend to other cues such as pitch and amplitude envelope. After duration-normalization, the single-vowel syllables were normalized to have the same long-term root-mean-square (rms) amplitude (65 dB). Therefore, the single-vowel components in the concurrent syllables were equated in terms of overall duration and amplitude. After summation, the long-term rms amplitudes of the concurrent syllables were normalized to 65 dB. In both the male-female and male-male conditions, there were a total of 256 concurrent syllables (16 single-vowel syllables from the male talker × 16 single-vowel syllables from the competing talker). Note that in the male-male condition, each single-vowel syllable was paired with itself once, and each pair of different single-vowel syllables was presented twice.

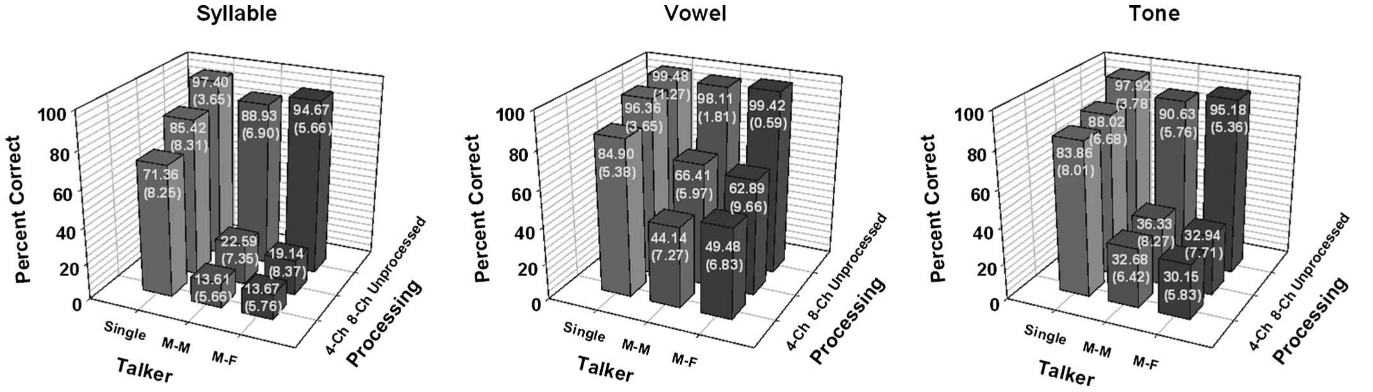


FIG. 1. Mean Chinese syllable (left panel), vowel (middle panel), and tone recognition scores (right panel) for NH subjects listening to unprocessed speech, eight- or four-channel CI simulation, as a function of talker condition. The numbers on each vertical bar represent the corresponding mean score and standard deviation (shown in the bracket).

Noise-band vocoders with either eight or four frequency channels were used to simulate CI speech processing (Shannon *et al.*, 1995). After pre-emphasis (first-order Butterworth high-pass filter at 1200 Hz), the input speech signal was band-pass filtered into eight or four frequency channels (fourth-order Butterworth filters). The overall input acoustic frequency range was 100–6000 Hz. The analysis bands were evenly distributed in terms of cochlear location according to Greenwood's (1990) formula. The corner frequencies (−3 dB points) were 100, 222, 404, 676, 1083, 1692, 2602, 3964, and 6000 Hz for the eight-channel processor and 100, 404, 1083, 2602, and 6000 Hz for the four-channel processor. The temporal envelope from each band was extracted by half-wave rectification and low-pass filtering (fourth-order Butterworth filter at 500 Hz), and was used to modulate wide-band noise. The amplitude-modulated noise carriers were filtered using the same pass-bands used for the analysis filters. The band-limited, amplitude-modulated noise carriers from all frequency channels were summed to produce the CI simulation speech, which was then normalized to have the same long-term rms amplitude as the input speech signal.

C. Procedures

Both single-syllable and concurrent-syllable recognitions were measured using the original, unprocessed speech, as well as speech processed by acoustic CI simulations. Thus, there were a total of nine experimental conditions [3 talker conditions (single, male-male, and male-female talkers) ×3 signal-processing conditions (unprocessed speech, eight- and four-channel CI simulations)]. To minimize potential learning effects, the test order of the experimental conditions was randomized across subjects; no learning trends were observed in terms of test order.

Subjects were seated in a double-walled sound-treated booth and listened to stimuli presented in sound field over a single loudspeaker (Tannoy Reveal) at 65 dBA. A closed-set identification task (16 choices) was used to measure both single-syllable and concurrent-syllable recognitions. In each trial, a stimulus was randomly selected from the stimulus set (without replacement) and presented to the subject. In the single-syllable recognition tasks, subjects were instructed to identify the Chinese syllable by clicking on one of the response choices shown on the screen: a1, a2, a3, a4, e1, e2, e3, e4, u1, u2, u3, u4, i1, i2, i3, and i4; note that the numbers in the response labels refer to the Chinese tones (1—high, flat, 2—rising, 3—falling-rising, and 4—falling). Responses were collected and scored as the percentage that the Chinese syllable, vowel, or tone was correctly identified. In the concurrent-syllable recognition tasks, subjects were instructed to identify the two Chinese syllables by making two consecutive choices; the order of choices was not important for scoring. Responses were collected and scored as the percentage that both syllables, both vowels, or both tones were correctly identified. No preview, feedback, or training was provided. Note that the subjects were very experienced with the CI simulations from their participation in previous studies, meaning that no training was required to familiarize subjects with the test procedure or the signal processing.

III. RESULTS

Figure 1 shows Chinese syllable, vowel, and tone recognition scores for the six NH subjects listening to single-talker, male-male, or male-female syllables, for processed and unprocessed speech. Note that the different recognition tasks had different chance performance levels. Chance level for syllable recognition with single syllables was 6.25% correct (1/16), while chance level for syllable recognition with concurrent syllables was 0.76% correct ($16/256\times 1/256 + 240/256\times 2/256$). Similarly, chance level for vowel or tone recognition with single syllables was 25% correct (1/4), while chance level for vowel or tone recognition with concurrent syllables was 10.94% correct ($4/16\times 1/16+12/16 \times 2/16$). When listening to unprocessed speech, subjects achieved nearly perfect recognition performance with both single and concurrent syllables. When listening to the eight- or four-channel CI simulation, single-talker speech recognition worsened, but remained at relatively high levels (>70% correct). With the CI simulations, recognition performance with concurrent-talker speech ranged from ~15% (e.g., syllables) to ~65% correct (e.g., vowels).

TABLE II. Results from one-way RM ANOVAs performed on single-talker data, with speech processing as the factor. Significant effects are shown in bold. Significant differences ($p < 0.05$) from *post-hoc* Bonferroni t-tests are also shown in bold.

	Speech processing (unprocessed, 8ch, 4ch)			
	dF, res	F-ratio	*p*-value	Post-hoc $p < 0.05$
Vowels	2,10	29.7	**<0.001**	**Unprocessed > 4ch; 8ch > 4ch**
Tones	2,10	14.2	**0.001**	**Unprocessed > 8ch, 4ch**
Syllables	2,10	41.4	**<0.001**	**Unprocessed > 8ch, 4ch; 8ch > 4ch**

A. Effects of speech-processing and talker conditions

Vowel, tone, and syllable recognition performances were analyzed independently. Single-talker performance was analyzed using one-way repeated measures analyses of variance (RM ANOVAs) with speech processing as the factor; unprocessed and processed speech were treated as different levels within the speech-processing factor. Concurrent-talker performance was analyzed using two-way RM ANOVAs with speech-processing and talker conditions as factors.

Table II shows the results from one-way RM ANOVAs performed on single-talker speech performance. Vowel, tone, and syllable recognitions were all significantly affected by speech processing. For vowels, there was no significant difference between unprocessed speech and the eight-channel simulation, but performance with the eight-channel simulation was significantly better than that with the four-channel simulation. For tones, performance with unprocessed speech was significantly better than that with either of the CI simulations, but there was no significant difference between the eight- and four-channel simulations. For syllables, performance was significantly different between any two of the three processing conditions (unprocessed speech, eight- and four-channel simulations).

Table III shows the results from two-way RM ANOVAs performed on concurrent-talker speech performance. Vowel, tone, and syllable recognitions were all significantly affected by speech processing. For vowels, tones, and syllables, performance with unprocessed speech was significantly better than that with either of the CI simulations. For vowels and syllables, performance with the eight-channel simulation was significantly better than that with the four-channel simulation. There was no main effect for talker conditions. There were significant interactions between speech-processing and talker conditions for tones and syllables. Tone and syllable recognitions were significantly better with the male-female condition for unprocessed speech, and with the male-male condition for eight-channel speech. However, the difference in mean performance between the talker conditions was quite small (<5%) within any of the speech-processing conditions.

B. Effects of vowel pairs

Figure 2 shows concurrent Chinese vowel and tone recognition scores with the CI simulations, as a function of vowel pairs in the concurrent syllables. Because there was no main effect for talker conditions and the detailed performance patterns were similar between the talker conditions, the male-male and male-female performance data were combined. Because syllable recognition performance was largely predictable from the vowel and tone scores, only vowel and tone recognition scores are shown. Also, because performance was nearly perfect with unprocessed speech, performance is shown only for the four- and eight-channel CI simulations.

Table IV shows the results from two-way RM ANOVAs performed on the data shown in Fig. 2. Concurrent-vowel recognition was significantly affected by the speech processing and vowel pairs, and there was a significant interaction between the speech processing and vowel pairs. Concurrent-tone recognition was also significantly affected by the vowel pairs; there was no significant interaction between the speech processing and vowel pairs.

Vowel response patterns were generated for the different vowel pairs used in the concurrent-syllable recognition tasks. Figure 3 shows the distribution of vowel responses for different vowel pairs. Although these vowel response patterns do not provide a totally unambiguous representation of confusions between vowel pairs, they may still provide useful information. The response patterns with unprocessed speech (data not shown) corresponded to nearly perfect recognition performance. When the concurrent syllables had the same vowel (i.e., /a/-/a/, /e/-/e/, /u/-/u/, and /i/-/i/), subjects chose

TABLE III. Results from two-way RM ANOVAs performed on concurrent-talker data, with speech-processing and talker conditions as factors. Significant effects are shown in bold. Significant differences ($p < 0.05$) from *post-hoc* Bonferroni t-tests are also shown in bold.

	Speech processing (Unprocessed, 8ch, 4ch)			Talker condition (M-M, M-F)			Speech processing ×talker condition		
	dF, res	F-ratio	*p*-value	dF, res	F-ratio	*p*-value	dF, res	F-ratio	*p*-value
Vowels	2,10	158.7	**<0.001**	1,10	1.2	0.33	2,10	2.7	0.11
Post-hoc $p < 0.05$	**Unprocessed > 8ch > 4ch**								
Tones	2,10	602.4	**<0.001**	1,10	0.2	0.65	2,10	11.5	**0.003**
Post-hoc $p < 0.05$	**Unprocessed>8ch, 4ch**						**Unprocessed: M-F > M-M; 8ch: M-M > M-F**		
Syllables	2,10	812.6	**<0.001**	1,10	0.9	0.40	2,10	19.6	**<0.001**
Post-hoc $p < 0.05$	**Unprocessed>8ch > 4ch**						**Unprocessed: M-F > M-M; 8ch: M-M > M-F**		

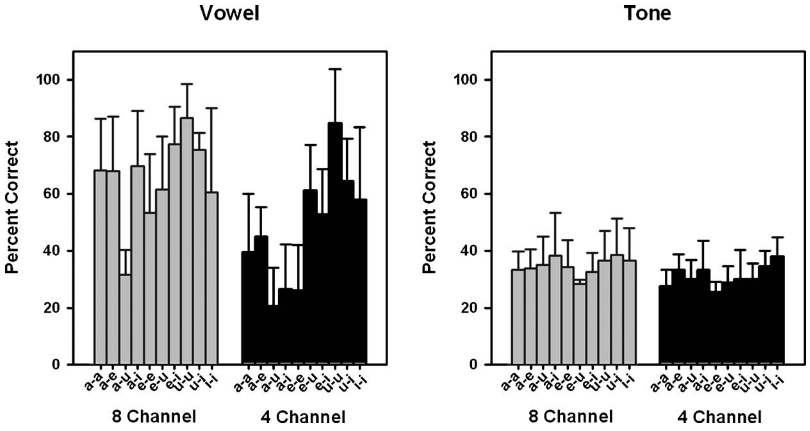


FIG. 2. Mean concurrent Chinese vowel (left panel) and tone recognition scores (right panel) for NH subjects listening to the eight- or four-channel CI simulation, as a function of vowel pairs in the concurrent syllables. The error bars represent one standard deviation of the mean.

the target vowel nearly 100%. When the concurrent syllables had two different vowels (e.g., /a/-/e/, /a/-/u/, /a/-/i/, etc.), subjects' responses were evenly split between the two target vowels (~50%). There was a broad distribution of vowel responses with the CI simulations, with subjects choosing vowels that were not present in the concurrent syllables and/or favoring one of the component vowels. For example, for vowel pair /a/-/u/ with the eight-channel CI simulation, subjects most often heard /a/ (55%), seldom heard /u/ (17%), and sometimes heard /e/ (19%), which was not present. Similarly, with four channels, subjects most often heard /a/ (53%), seldom heard /u/ (15%), and sometimes heard /e/ (27%).

To analyze whether subjects were biased toward responding with the same vowel for individual syllable pairs, the percentage of responses with the same vowel within a pair was calculated. Averaged across both talker and CI simulation conditions, subjects made such responses only 30% of the time, i.e., close to the percentage of concurrent syllables actually consisting of the same vowel (25%). Therefore, there was no strong bias toward responding with the same vowel within a pair.

C. Effects of tone pairs

Figure 4 shows concurrent Chinese vowel and tone recognition scores with the CI simulations, as a function of tone pairs in the concurrent syllables. Similar to Fig. 2, the male-male and male-female performance data were combined. Compared with the vowel pairs (left panel of Fig. 2), the different tone pairs had a smaller effect on vowel recognition (left panel of Fig. 4). Conversely, the different tone pairs had a stronger effect on tone recognition (right panel of Fig. 4) than did the different vowel pairs (right panel of Fig. 2).

Table V shows the results from two-way RM ANOVAs performed on the data shown in Fig. 4. Concurrent-vowel recognition was significantly affected by the speech processing, but not by the tone pairs; there was a significant interaction between the speech processing and tone pairs. Except for tone pair 1-1, concurrent-tone recognition was significantly better when tone pairs consisted of the same tone rather than different tones. For tone pairs consisting of the same tone, performance was significantly better for tone pairs 3-3 and 4-4 than for tone pair 1-1.

Tone response patterns were generated for the different tone pairs used in the concurrent-syllable recognition tasks. Figure 5 shows the distribution of tone responses for different tone pairs. Similar to the vowel response patterns, the tone response patterns do not provide a totally unambiguous representation of tone pair confusions, but they may still provide useful information. Again, the response patterns with unprocessed speech are not shown as subjects achieved nearly perfect recognition performance. There was a broad and/or uneven distribution of tone responses with the CI simulations. For example, for tone pair 1-4, subjects more often responded with tone 4 (56%) than with tone 1 (32%). For tone pair 2-4, subjects responded with tone 2 (33%), tone 4 (37%), and tone 1 (23%).

Different from concurrent-vowel recognition, subjects

TABLE IV. Results from two-way RM ANOVAs performed on the concurrent-talker data shown in Fig. 2, with speech processing and vowel pairs as factors. Significant effects are shown in bold. Significant differences ($p < 0.05$) from *post-hoc* Bonferroni t-tests are also shown in bold.

	Speech processing (8ch, 4ch)			Vowel pairs (a-a, a-e, a-u, a-i, e-e, e-u, e-i, u-u, u-i, i-i)			Speech processing × vowel pairs		
	dF, res	F-ratio	*p*-value	dF, res	F-ratio	*p*-value	dF, res	F-ratio	*p*-value
Vowels	1,45	28.2	**0.003**	9,45	6.7	**<0.001**	9,45	6.1	**<0.001**
Post-hoc $p < 0.05$		**8ch > 4ch**			**e-u, e-i, u-u, u-i, i-i > a-u; u-u > a-a, a-i, e-e**			**8ch: a-a, a-e, a-i, e-i, u-u, u-i > a-u 4ch: e-u, u-u, u-i > a-u, a-i, e-e; i-i > a-u; u-u > a-a, a-e**	
Tones	1,45	5.9	0.06	9,45	3.0	**0.008**	9,45	1.5	0.19
Post-hoc $p < 0.05$					**i-i > e-u**				

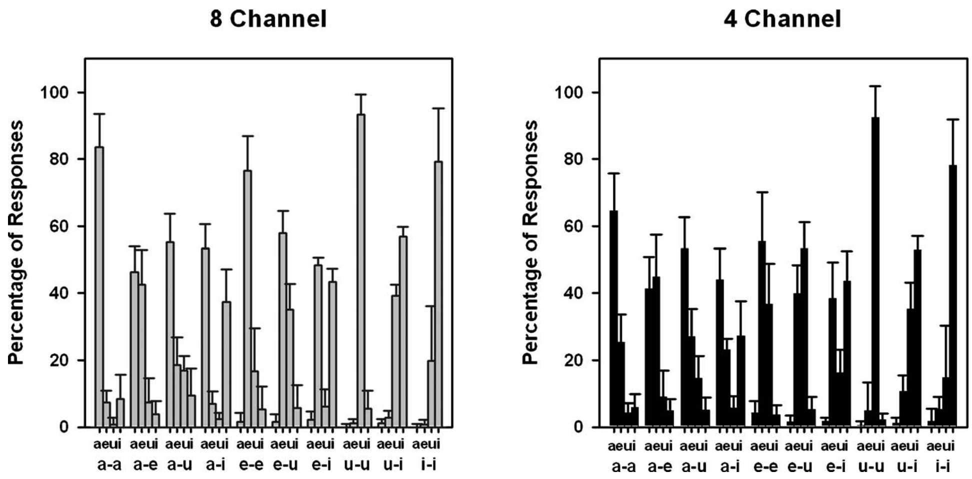


FIG. 3. Mean distribution of vowel responses (in percentage of responses) for the different vowel pairs in the concurrent syllables, with the eight-channel (left panel) or four-channel CI simulation (right panel). The error bars represent one standard deviation of the mean.

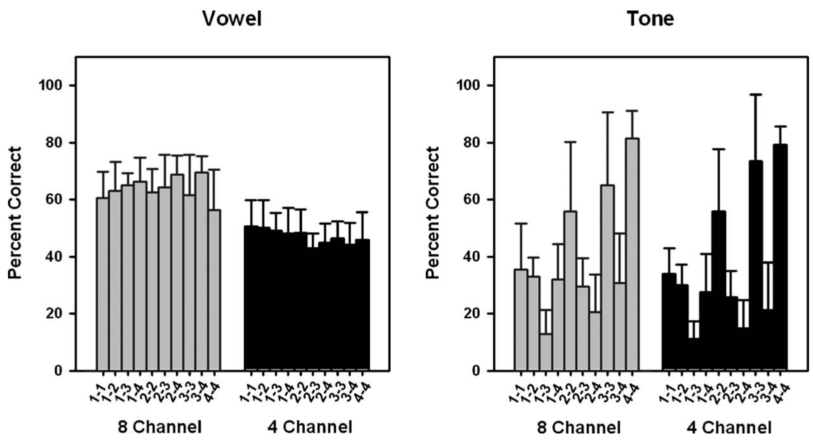


FIG. 4. Mean concurrent Chinese vowel (left panel) and tone recognition scores (right panel) for NH subjects listening to the eight- or four-channel CI simulation, as a function of tone pairs in the concurrent syllables. The error bars represent one standard deviation of the mean.

TABLE V. Results from two-way RM ANOVAs performed on the concurrent-talker data shown in Fig. 4, with speech processing and tone pairs as factors. Significant effects are shown in bold. Significant differences ($p<0.05$) from *post-hoc* Bonferroni t-tests are also shown in bold.

	Speech processing (8ch, 4ch)			Tone pairs (1-1, 1-2, 1-3, 1-4, 2-2, 2-3, 2-4, 3-3, 3-4, 4-4)			Speech processing × tone pairs		
	dF, res	F-ratio	*p*-value	dF, res	F-ratio	*p*-value	dF, res	F-ratio	*p*-value
Vowels	1,45	22.6	**0.005**	9,45	1.2	0.31	9,45	2.4	**0.02**
Post-hoc $p<0.05$		**8ch > 4ch**						**8ch: 2-4, 3-4 > 4-4**	
Tones	1,45	2.8	0.15	9,45	18.2	**<0.001**	9,45	1.5	0.16
Post-hoc $p<0.05$					**2-2, 3-3, 4-4 > 1-3, 1-4, 2-3, 2-4, 3-4; 3-3, 4-4 > 1-1, 1-2**				

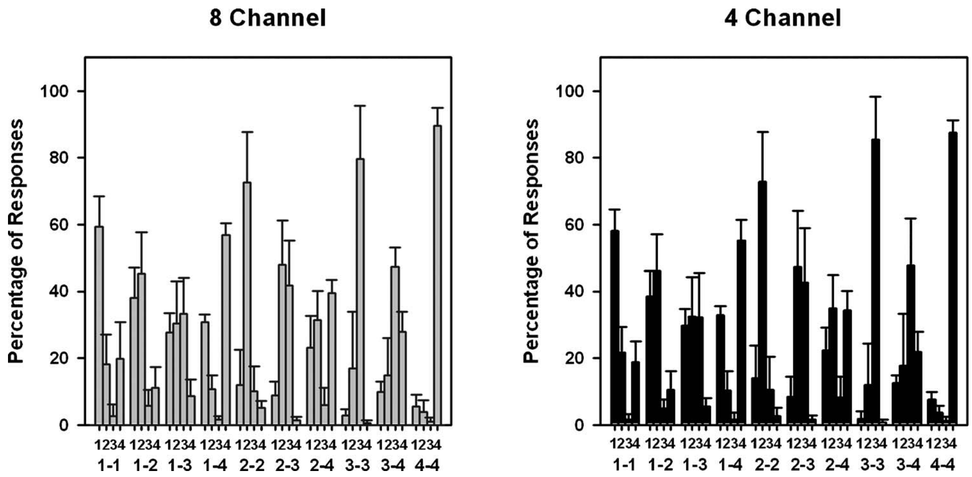


FIG. 5. Mean distribution of tone responses (in percentage of responses) for the different tone pairs in the concurrent syllables, with the eight-channel (left panel) or four-channel CI simulation (right panel). The error bars represent one standard deviation of the mean.

had a strong bias toward responding with the same tone for individual syllable pairs. The percentage of such responses averaged across both talker and CI simulation conditions was 50%, which was two times the percentage of concurrent syllables actually consisting of the same tone (25%).

IV. DISCUSSION

Concurrent Chinese syllable, vowel, and tone recognitions were nearly perfect with unprocessed speech. With unprocessed speech, concurrent-syllable and tone recognition scores were slightly (~5%) but significantly better for the male-female condition, showing some benefit of the larger F0 separation. For the Chinese vowel stimuli used in the present study, different vowels produced by the same male talker had different instantaneous F0's because of the different tonal patterns and variations in production. With unprocessed speech, these F0 differences (along with pitch period asynchrony and formant transitions) may have been sufficient to produce nearly perfect vowel recognition performance in the male-male condition, similar to that in the male-female condition. Also, with unprocessed speech, there was little variability in concurrent recognition performance across the different vowel and tone pairs, due to ceiling effects.

The reduced spectral resolution in the acoustic CI simulations had a more detrimental effect on concurrent-syllable recognition than on single-syllable recognition (Fig. 1). This confirms that while gross spectral and temporal representations may support good speech understanding in quiet, they do not provide sufficient acoustic cues for sound source segregation. The single-talker vowel and tone recognition scores with the CI simulations in the present study were slightly higher than previously reported with real CI listeners (Luo *et al.*, 2008), possibly because the present stimulus set was a subset of the stimuli used in Luo *et al.*, 2008 and because of inherent differences between simulated and real CI listening. The present concurrent-vowel recognition results with the 8-channel CI simulation were comparable to those reported by Qin and Oxenham (2005) with 24-channel processing, possibly due to the smaller number of vowel choices used in the present study [four, as compared to five in the Qin and Oxenham (2005) study].

In the single- and concurrent-talker conditions, when the number of frequency channels was increased from four to eight, vowel and syllable recognitions significantly improved, while tone recognition was unchanged. These findings extend those for single-syllable recognition with 1–4 frequency channels (Fu *et al.*, 1998) or up to 12 frequency channels (Xu *et al.*, 2002). In the study by Fu *et al.* (1998), the effects of spectral resolution may have been obscured by the unequal distribution of each tone type in different conditions. In the present study, doubling the number of frequency channels from four to eight may have improved spectral envelope representations, but not to the point of resolving F0 and harmonics. Thus, single- or concurrent-vowel recognition improved with the number of frequency channels while tone recognition remained unchanged. As suggested by Fu *et al.* (1998), when spectral resolution is severely limited, either single- or concurrent-tone recognition in electric hearing strongly relies on temporal envelope cues.

When listening to the CI simulations, NH listeners' vowel recognition was much better than tone recognition for concurrent syllables. In contrast, vowel and tone recognition performances were similar for single syllables, possibly due to ceiling effects. When two Chinese syllables are presented simultaneously, interference is created in both the spectral and temporal domains. In the CI simulations, the degraded spectral envelopes (i.e., formant structures) of the two vowels were mixed together and further smeared, resulting in poor segregation and recognition of individual vowels. In each frequency channel, the temporal waveforms of the two vowels were also mixed, creating some degree of modulation detection interference (e.g., Richardson *et al.*, 1998). It is plausible that such modulation detection interference may have adversely affected tracking of amplitude envelopes and periodicity fluctuations, resulting in poor segregation and recognition of individual tones. Because concurrent-vowel recognition was so much better than concurrent-tone recognition with the CI simulations, listeners may have been more susceptible to interference in the temporal domain than in the spectral domain. These results indicate that compared with English syllable recognition, which does not require tone recognition, concurrent Chinese syllable recognition with CI may be more challenging, due to the strong interference between concurrent tones.

In general, there was little difference in performance between the male-male and male-female conditions with the CI simulations. Note that with the eight-channel simulation, concurrent-syllable and tone recognition scores were slightly but significantly better for the male-male condition. In the male-male condition, each single-vowel was paired with itself once. Such pairs of two exactly same vowels may not be informative for concurrent-syllable recognition, but they introduced minimum temporal waveform interference and may have produced better tone recognition. The small number of frequency channels in the CI simulations may have limited listeners' ability to take advantage of the larger F0 separation in the male-female condition. Previous studies have also shown that increasing the F0 separation between concurrent talkers provides no benefit for CI users' masking release from competing talkers (Stickney *et al.*, 2007), or for recognition of concurrent, synthesized vowels in CI simulations (Qin and Oxenham, 2005). Recently, Carlyon *et al.* (2007) showed that CI users could not exploit pulse asynchrony or rate differences between concurrent channels to segregate sounds, suggesting that F0 differences may help segregation only when harmonics are resolved by the peripheral auditory system. However, in a study using sequential (rather than concurrent) presentation of vowels, Gaudrain *et al.* (2008) observed F0-based auditory segregation at a much faster-than-normal presentation rate (7.5 vowels/s) in NH subjects listening to a 12-channel (rather than 8-channel) CI simulation. Such F0-based auditory segregation has been attributed to spectral envelope cues instead of temporal periodicity cues.

With the CI simulations, the variability in performance across different vowel pairs was much larger for vowel recognition than for tone recognition (Fig. 2). The performance patterns (across vowel pairs) were quite different between concurrent-vowel and tone recognitions, suggesting that the percepts may have not been strongly related. In other words, better concurrent-vowel recognition with some vowel pairs was not necessarily associated with better concurrent-tone recognition, or vice-versa. Conversely, with the CI simulations, concurrent-tone (rather than vowel) recognition greatly varied across the tone pairs (Fig. 4). Specifically, concurrent-tone recognition was much better for tone pairs consisting of the same tone (except for tone pair 1-1), while concurrent-vowel recognition was quite similar across the tone pairs. These results are different from those of Chalikia and Bregman (1989), who found that NH listeners' concurrent-vowel recognition was better with crossing pitch contours than with parallel pitch contours. The CI simulations did not preserve low-order, resolved harmonics, which may have made listeners unable to benefit from the different tonal patterns within concurrent syllables (Qin and Oxenham, 2005; Carlyon *et al.*, 2007). Although there are perceptual trade-offs between spectral and temporal cues (e.g., Xu and Pfingst, 2008), vowel recognition in electric hearing strongly relies on spectral envelope cues, while tone recognition depends more strongly on temporal envelope cues. The limited spectro-temporal fine structure cues available in CI speech processing may not have provided sufficiently salient temporal envelope cues to segregate spectral envelope cues (and vice-versa). Therefore, concurrent-vowel and tone recognitions with the present CI simulations were independent of each other.

The distribution of vowel responses (Fig. 3) indicates that performance was poorer for some vowel pairs than for others. With the CI simulations, subjects' perception was typically dominated by one of the component vowels, especially when the other vowel was /u/. For example, subjects more often heard /a/ and /i/ for vowel pairs /a/-/u/ and /i/-/u/, respectively; /u/ was seldom heard. The first two formant frequencies of /u/ are relatively low and closely spaced (see Table I), and may have been masked by the competing vowel /a/ or /i/. McKeown (1992) found that for NH subjects listening to concurrent vowels, /u/ was also perceptually dominated by the other component vowel (e.g., /a/ or /i/). McKeown (1992) suggested that such vowel masking may have occurred at the auditory periphery in the spectral domain, and at a more central level during cognition and attention. Another common error was subjects' perception of a vowel that was not present in the component vowel pair. For example, with the four-channel CI simulation, subjects often responded with /e/ when presented with vowel pair /a/-/i/. The perception of the vowel /e/ may have been due to the combination of the first formant of /i/ with the second formant of /a/ (see Table I). Interestingly, in both the male-male and male-female conditions, vowel pairs consisting of the same vowel (except for vowel pair /u/-/u/) were not always better recognized than those consisting of different vowels. For vowel pairs /e/-/e/ and /i/-/i/, subjects sometimes responded with /u/ in addition to the target vowel (/e/ or /i/). The noise-band carriers used in the present CI simulations may have produced the perceptual illusion of the vowel /u/. Because of the closely spaced F1 and F2 values, CI simulations of /u/ have most energy in only one or two adjacent frequency channels (e.g., the second lowest channel of the four-channel CI simulation), and thus may sound similar to narrow-band noise.

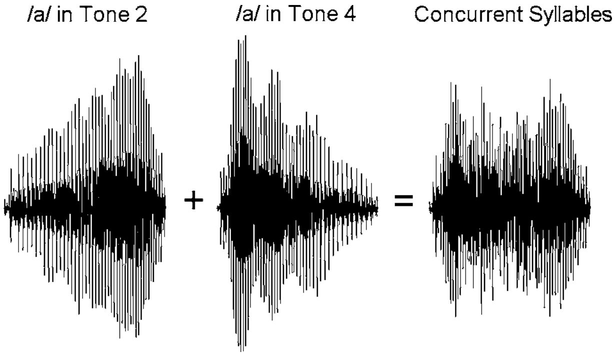


FIG. 6. Example of temporal envelope interference between concurrent syllables. A Chinese vowel /a/ in tone 2 (left panel) is combined with another Chinese vowel /a/ in tone 4 (middle panel). The combined temporal waveforms show a flat amplitude envelope (right panel).

Different from vowel recognition, concurrent-tone recognition with the CI simulations was significantly better when the two syllables had the same tone (except for tone pair 1-1), due to subjects' tendency to respond with the same tone for concurrent syllables within a pair. Different pitch contours between concurrent syllables did not aid in sound source segregation, but rather produced poorer tone recognition. The distribution of tone responses (Fig. 5) revealed that, similar to vowel recognition, concurrent-tone recognition with the CI simulations also exhibited two types of errors, both of which may be explained by temporal envelope interference. Note that for Chinese single-vowel syllables, the amplitude envelope and F0 contour have similar shapes, which is an important temporal cue for Chinese tone recognition with limited spectral resolution (Luo and Fu, 2004). When tone 1 (flat) was presented simultaneously with another tone (e.g., tone 4, falling), the combined temporal envelope largely followed tone 4 and may have resulted in the perceptual dominance of tone 4. When tone 2 (rising) was presented simultaneously with tone 4 (falling), the combined temporal waveforms may have resulted in a flat amplitude envelope, indicating a flat tone (i.e., tone 1). Not surprisingly, listeners often responded with tone 1 for tone pair 2-4. An example of temporal envelope interference is shown in Fig. 6.

The present study used four and eight frequency channels to simulate the typical spectral resolution available to CI users. The simulation results with NH listeners suggest that CI users might not utilize F0 differences across talkers and/or tones to recognize concurrent syllables. However, the strength of temporal envelope pitch may have been reduced in the simulations as compared to the real implant case, due to the noise-band carrier (which might reduce temporal envelope saliency) and the employment of synthesis band-pass filters (Laneau *et al.*, 2006). Also, different from the NH subjects, CI performance can be greatly affected by patient-related factors. While simulation studies with NH listeners presumably reduce these patient-related factors, they are an important consideration for future studies with real CI users.

Complex perceptual tasks such as speech recognition in the presence of competing speech require high degrees of spectral resolution. Even with 24 channels, concurrent English vowel recognition performance is much poorer than that with unprocessed speech (Qin and Oxenham, 2005). Although implanted with 16–22 electrodes, CI users can access only approximately eight channels, due to electrode interactions. Advanced speech-processing strategies that restore spectro-temporal fine structure cues to CI users may enhance their sound source segregation. Binaural cues, via bilateral CIs or a hearing aid in the non-implanted ear, may also improve CI users' sound source segregation. Previous studies with NH listeners have shown improved concurrent-vowel recognition when the two vowels were presented to different ears (i.e., dichotic hearing instead of diotic hearing; Zwicker, 1984) or when adding interaural time differences (Shackleton and Meddis, 1992). Recently, Long *et al.* (2006) found that bilateral CI users' signal detection in noise was significantly better when the 500-Hz temporal envelopes delivered to a single electrode in each ear were out of phase rather than in phase. Thus, it is of interest to investigate concurrent-vowel and tone recognitions in bilateral CI users and listeners with bilaterally combined electric and acoustic stimulation.

V. CONCLUSIONS

The present study measured NH listeners' recognition of concurrent Chinese syllables, vowels, and tones produced by one male and one female talker or by the same male talker. Performance was measured with original, unprocessed speech, and with speech processed by four- or eight-channel acoustic CI simulation. Concurrent-syllable, vowel, and tone recognitions were significantly poorer with the CI simulations than with original, unprocessed speech. Syllable and tone recognitions were significantly better with the male-female condition for unprocessed speech, and with the male-male condition for eight-channel speech. With the CI simulations, concurrent-vowel and syllable recognitions were significantly different across different vowel pairs, while tone recognition remained largely unchanged. In contrast, concurrent-vowel recognition was not significantly affected by the different tone pairs. Tone and syllable recognitions were significantly better when the two syllables had the same tone. With the CI simulations, concurrent-vowel and tone recognitions were independent of each other. The weak pitch coding in the CI simulations may preclude enhanced concurrent recognition performance derived from large F0 separations between the talkers and/or different pitch contours between the syllables. Furthermore, with the CI simulations, concurrent Chinese syllable recognition may be more challenging than English syllable recognition, due to the strong interference between tonal envelope cues.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in these experiments. We thank John J. Galvin III for editorial assistance. We would also like to thank three anonymous reviewers for their constructive comments on an earlier version of this paper. Research was supported in part by NIH (Grant Nos. R03-DC-008192 and R01-DC-004993).

Assmann, P. F. (**1995**). "The role of formant transitions in the perception of concurrent vowels," J. Acoust. Soc. Am. **97**, 575–584.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Carlyon, R. P., Long, C. J., Deeks, J. M., and McKay, C. M. (**2007**). "Concurrent sound segregation in electric and acoustic hearing," J. Assoc. Res. Otolaryngol. **8**, 119–133.

Chalikia, M. H., and Bregman, A. S. (**1989**). "The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation," Percept. Psychophys. **46**, 487–496.

Culling, J. F., and Summerfield, Q. (**1995**). "The role of frequency modulation in the perceptual segregation of concurrent vowels," J. Acoust. Soc. Am. **98**, 837–846.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X.-S. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., Hsu, C.-J., and Horng, M.-J. (**2004**). "Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant users," Ear Hear. **25**, 501–508.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J.-C. (**2008**). "Steaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," J. Acoust. Soc. Am. **124**, 3076–3087.

Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species-29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Laneau, J., Moonen, M., and Wouters, J. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.

Long, C. J., Carlyon, R. P., Litovsky, R. Y., and Downs, D. H. (**2006**). "Binaural unmasking with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **7**, 352–360.

Luo, X., and Fu, Q.-J. (**2004**). "Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants," J. Acoust. Soc. Am. **116**, 3659–3667.

Luo, X., Fu, Q.-J., Wei, C.-G., and Cao, K.-L. (**2008**). "Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users," Ear Hear. **29**, 957–970.

McKeown, J. D. (**1992**). "Perception of concurrent vowels: The effect of varying their relative level," Speech Commun. **11**, 1–13.

Meddis, R., and Hewitt, M. J. (**1992**). "Modeling the identification of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. **91**, 233–245.

Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Richardson, L. M., Busby, P. A., and Clark, G. M. (**1998**). "Modulation detection interference in cochlear implant subjects," J. Acoust. Soc. Am. **104**, 442–452.

Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.

Shackleton, T. M., and Meddis, R. (**1992**). "The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs," J. Acoust. Soc. Am. **91**, 3579–3581.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F.-G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Summerfield, Q., and Assmann, P. F. (**1991**). “Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony,” J. Acoust. Soc. Am. **89**, 1364–1377.

Wang, R.-H. (**1993**). “The standard Chinese database,” University of Science and Technology of China, internal materials.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). “Better speech recognition with cochlear implants,” Nature (London) **352**, 236–238.

Xu, L., and Pfingst, B. E. (**2008**). “Spectral and temporal cues for speech recognition: Implications for auditory prostheses,” Hear. Res. **242**, 132–140.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). “Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses,” J. Acoust. Soc. Am. **112**, 247–258.

Zwicker, U. T. (**1984**). “Auditory recognition of diotic and dichotic vowel pairs,” Speech Commun. **3**, 265–277.

How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli

Leslie R. Bernstein and Constantine Trahiotis
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 2 October 2008; revised 19 February 2009; accepted 23 February 2009)

This study addressed how manipulating certain aspects of the envelopes of high-frequency stimuli affects sensitivity to envelope-based interaural temporal disparities (ITDs). Listener's threshold ITDs were measured using an adaptive two-alternative paradigm employing "raised-sine" stimuli [John, M. S., *et al.* (2002). Ear Hear. **23**, 106–117] which permit independent variation in their modulation frequency, modulation depth, and modulation exponent. Threshold ITDs were measured while manipulating modulation exponent for stimuli having modulation frequencies between 32 and 256 Hz. The results indicated that graded *increases* in the exponent led to graded *decreases* in envelope-based threshold ITDs. Threshold ITDs were also measured while parametrically varying modulation exponent and modulation depth. Overall, threshold ITDs decreased with increases in the modulation depth. Unexpectedly, increases in the exponent of the raised-sine led to especially large decreases in threshold ITD when the modulation depth was low. An interaural correlation-based model was generally able to capture changes in threshold ITD stemming from changes in the exponent, depth of modulation, and frequency of modulation of the raised-sine stimuli. The model (and several variations of it), however, could not account for the unexpected interaction between the value of raised-sine exponent and its modulation depth.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3101454]

PACS number(s): 43.66.Pn, 43.66.Ba [RLF] Pages: 3234–3242

I. INTRODUCTION

In 1997, van de Par and Kohlrausch (1997) introduced a new class of high-frequency signals that they termed "transposed stimuli." They created transposed stimuli in an effort to provide high-frequency auditory channels with envelope-based neural timing information that would mimic waveform-based neural timing information naturally available only in low-frequency channels. van de Par and Kohlrausch (1997) reported data showing that transposed stimuli enhanced high-frequency binaural processing in that they yielded NoSπ thresholds of detection that were comparable to the much lower thresholds of binaural detection routinely obtained at low center frequencies. Following that, Bernstein and Trahiotis specifically showed that transposed stimuli enhanced the processing of envelope-based interaural temporal disparities (ITDs) conveyed by high-frequency channels. They reported that transposed stimuli yielded smaller threshold ITDs (Bernstein and Trahiotis, 2002), larger extents of ITD-based laterality (Bernstein and Trahiotis, 2003), and substantial resistance to binaural interference effects produced by the addition of simultaneously presented, diotic low-frequency energy (Bernstein and Trahiotis, 2004, 2005). Furthermore, physiological studies have also revealed "enhanced" processing in that the neural timing information conveyed by the envelopes of high-frequency transposed stimuli can approximate that conveyed by the fine-structure of low-frequency waveforms (Griffin *et al.*, 2005; Dreyer and Delgutte, 2006).

At this time, it remains an open question just which aspect(s) of the envelopes of any high-frequency stimuli, be they transposed or conventional (e.g., sinusoidally amplitude-modulated (SAM) tones, two-tone complexes, and bands of Gaussian noise), lead to efficient processing of ongoing ITDs. To begin to answer this question, we conducted the first of a series of studies employing "raised-sine" stimuli that were recently described by John *et al.* (2002) in their study of steady state auditory evoked potentials. The algorithm used to generate raised-sine stimuli allows one to vary independently the frequency of modulation, the depth of modulation, and the exponent of the raised-sine. (Varying the exponent affects the "peakedness" or "sharpness" of the envelope of a raised-sine waveform.) These are features that cannot be varied independently with conventional stimuli such as SAM tones, repeated Gaussian clicks (e.g., Buell and Hafter, 1988; Stecker and Hafter, 2002), or the transposed tones used in our previous studies. As will be shown in Sec. I A, the use of raised-sine stimuli allows one to generate high-frequency signals having envelopes with temporal features that "fall in between" those of SAM tones and those of transposed stimuli while having spectral content restricted to a relatively narrow range. The purpose of this study was to determine how the discriminability of ongoing ITDs is affected by systematic and graded changes in the temporal features of such stimuli. To that end, two experiments were conducted. One focused on determining how varying the exponent of the raised-sine affects threshold ITDs for stimuli having frequencies of modulation ranging from 32 to 256

0001-4966/2009/125(5)/3234/9/$25.00 © 2009 Acoustical Society of America

Hz. The second focused on determining how parametric changes in both the exponent and its depth of modulation affect threshold ITDs for a raised-sine stimulus. Together, the results of the experiments provide initial insights regarding how particular features of the temporal signatures of the envelopes of high-frequency stimuli and their interaction affect sensitivity to changes in ongoing ITDs.

A. Generating raised-sine stimuli

The generation of raised-sine stimuli entails raising a dc-shifted sine-wave to a power greater than or equal to 1.0 prior to multiplication with a carrier. The equation used to generate such stimuli is

$$y(t) = (\sin(2\pi f_c t))(2m(((1+\sin(2\pi f_m t))/2)^n - 0.5) + 1), \tag{1}$$

where f_c is the frequency of the carrier, f_m is the frequency of the modulator, m is the modulation index, and n is the exponent denoting the power to which the dc-shifted modulator is raised.[1]

The left side of Fig. 1 depicts the time-waveforms for cases in which a 128 Hz modulating tone was raised using exponents of 1, 2, 4, or 8 prior to multiplication with a 4-kHz carrier. In all cases, $m=1.0$. The bottom row of the figure depicts a 128-Hz tone transposed to 4 kHz. Note that an exponent of 1.0 yields a conventional SAM waveform. Examination of the figure reveals that the peakedness or sharpness of the envelope increases directly with the value of the exponent to which the modulator is raised. Simultaneously, for these 100%-modulated signals, the "dead-time" or "off-time" between individual lobes of the envelope also increases with increasing values of the exponent. The right side of the figure displays the long-term spectrum of each stimulus. Note that increasing the value of the exponent also increases the number of "sidebands" and their spectral extent. It is important to note that, for each of the stimuli depicted, the vast majority of its energy falls within the approximately 500-Hz wide auditory filter centered at 4 kHz (see Moore, 1997).

II. EXPERIMENT I

A. Procedure

Threshold ITDs were measured for raised-sine stimuli having exponents of 1.0 (equivalent to a SAM tone), 1.5, 2.0, 4.0, and 8.0 and for transposed tones. All stimuli were centered at 4 kHz. For each of the six types of "targets," thresholds were measured at rates of modulation ranging between 32 and 256 Hz. Targets were generated digitally using a sampling rate of 20 kHz (TDT AP2), were low-pass filtered at 8.5 kHz (TDT FLT2), and were presented via Etymotic ER-2 insert earphones at a level of 70 dB sound pressure level (SPL). The duration of the targets was 300 ms including 20-ms $\cos^2$ rise-decay ramps. A continuous diotic noise, low-passed at 1.3 kHz (spectrum level equivalent to 30 dB SPL) was presented to preclude listeners' use of low-frequency distortion products arising from normal, non-linear peripheral auditory processing (e.g., Nuetzel and Hafter, 1976; Bernstein and Trahiotis, 1994).

Threshold ITDs were measured using a two-cue, two-alternative, forced choice, adaptive task. Each trial consisted of a warning interval (500 ms) and four 300-ms observation intervals separated by 400 ms. Each interval was marked visually by a computer monitor. Feedback was provided for approximately 400 ms after the listener responded. The stimuli in the first and fourth intervals were diotic. The listener's task was to detect the presence of an ongoing ITD (left-ear leading) that was presented with equal *a priori* probability in either the second or the third interval. The remaining interval, like the first and fourth intervals, contained diotic stimuli. Ongoing ITDs were imposed by applying linear phase-shifts to the representation of the signals in the frequency domain and then gating the signals destined for the left and right ears coincidentally, after transformation to the time-domain. The starting phases of the envelopes and carriers of the targets were chosen randomly for each observation interval both within and across trials. The ITD for a particular trial was determined adaptively in order to estimate 70.7% correct (Levitt, 1971). The initial step-size for the adaptive track corresponded to a factor of 1.584 (equivalent to a 2-dB change of ITD) and was reduced to a factor of 1.122 (equivalent to a 0.5-dB change of ITD) after two reversals. A run was terminated after 12 reversals and threshold was defined as the geometric mean of the ITD across the last 10 reversals.

Four normal-hearing adults served as listeners. Particular stimulus combinations were chosen pseudo-randomly and three consecutive estimates of threshold were obtained for each of the 24 stimulus combinations (six types of target × four frequencies of modulation) before moving on to the next one. Then, three more thresholds were obtained by revisiting the same stimulus conditions in reverse order. The entire procedure was repeated, yielding 12 estimates of threshold for each stimulus condition. The final values of threshold for each listener and stimulus condition were obtained by computing the median of the 12 estimates.

B. Results and discussion

Figure 2 displays the mean "normalized" threshold ITDs, calculated across the four listeners as a function of the exponent of the raised-sine stimulus. For purposes of comparison, normalized threshold ITDs obtained with the transposed stimuli are plotted at the far right. The parameter of the plot is the frequency of modulation. Normalized thresholds are shown in order to remove the differences in absolute sensitivity to ITD that are commonly found across listeners with high-frequency, complex stimuli (e.g., Bernstein *et al.*, 1998). The goal was to remove such inter-listener variability in order to reveal more precisely the *changes* in threshold ITDs that occur across conditions. The normalization was accomplished by dividing an individual listener's threshold ITDs by that listener's threshold ITD obtained with a SAM tone (raised-sine exponent equal to 1.0) having a frequency of modulation of 128 Hz. The individual threshold ITDs for

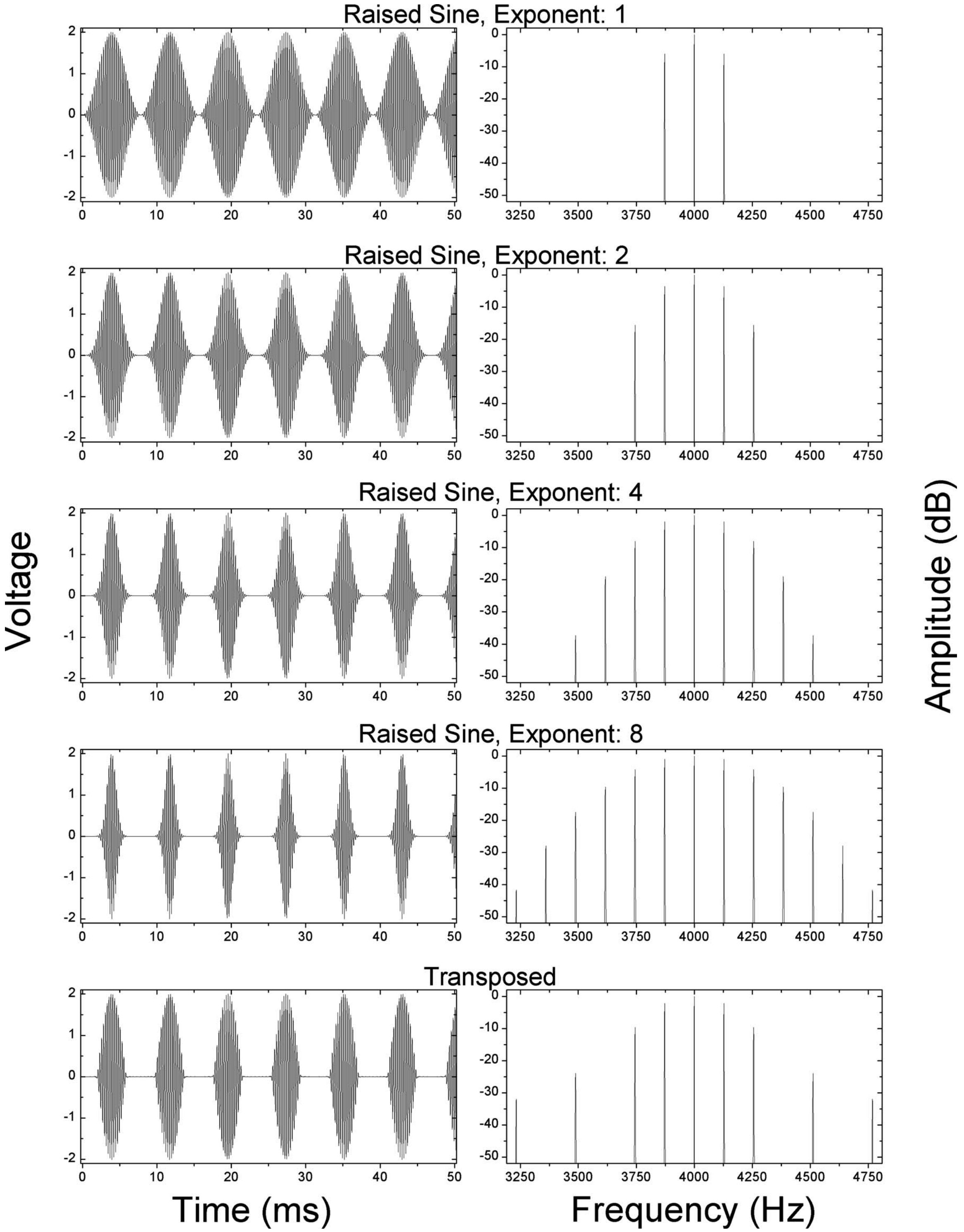


FIG. 1. Left-hand panels: 50 ms epochs of the time-waveforms of 4-kHz-centered raised-sine stimuli modulated at 128 Hz and having exponents of 1, 2, 4, and 8 (rows 1–4, respectively) and of a 4-kHz-centered transposed stimulus modulated at the same rate (bottom row). Right-hand panels: Each row depicts the corresponding long-term power spectrum of the time-waveform shown immediately to its left.

that "reference" stimulus were 128, 271, 113, and 217 μs. The error bars in Fig. 2 represent ±1 standard error of the mean normalized thresholds.

Visual inspection of the patterning of the normalized thresholds reveals three general outcomes. First, for all four rates of modulation, threshold ITDs decreased with increases in the exponent of the raised-sine and approximated threshold ITDs obtained with transposed stimuli when the exponent was 8.0. Intuitively, this outcome appears to be consistent with the notion that, for a given rate of modulation, graded changes in the amounts of peakedness/sharpness of the envelope of 100%-modulated stimuli lead to graded changes in sensitivity to ongoing envelope-based ITDs.[2] Such changes appear to be largest for the rate of modulation of 32 Hz, where threshold ITDs are generally largest, and smallest for the rate of modulation of 128 Hz, where thresh-

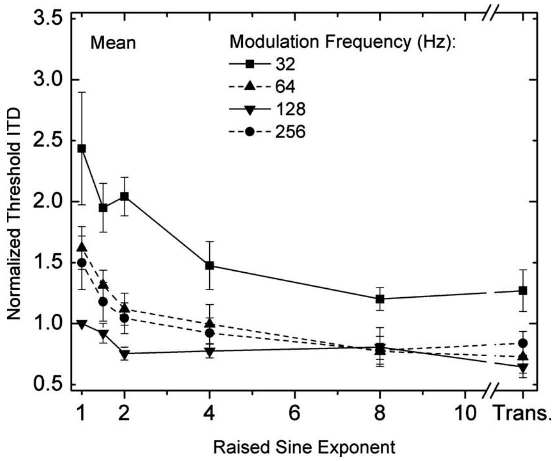


FIG. 2. Mean normalized threshold ITDs, calculated across the four listeners as a function of the exponent of the raised-sine stimulus. Normalized threshold ITDs obtained with the transposed stimuli are plotted at the far right. The normalization was accomplished by dividing an individual listener's threshold ITDs by that listener's threshold ITD obtained with a SAM tone (raised-sine exponent equal to 1.0) having a frequency of modulation of 128 Hz. The individual threshold ITDs for that reference stimulus were 128, 271, 113, and 217 μs. The parameter of the plot is the frequency of modulation. Error bars represent ±1 standard error of the mean normalized thresholds.

old ITDs are generally smallest. Second, threshold ITDs decreased with increases in rate of modulation from 32 to 128 Hz and then increased slightly when the rate of modulation was increased to 256 Hz. The latter trend was found previously with SAM and transposed tones by Bernstein and Trahiotis (2002). Third, the relatively small standard errors about each mean indicate that the relative changes in threshold ITD as a function of changes in either rate of modulation of the exponent of the raised-sine were homogeneous across listeners.

The data obtained with the raised-sine stimuli in Fig. 2 were subjected to a two-factor (four modulation frequencies ×five exponents), within-subjects analysis of variance (ANOVA). The error terms for the main effects and for the interactions were the interaction of the particular main effect (or the particular interaction) with the subject "factor" (Keppel, 1973). In addition to testing for significant effects, the proportions of variance accounted for (ω^2) were determined for each significant main effect and interaction (Hays, 1973).

Consistent with visual inspection of the data, the main effect of frequency of modulation was significant (assuming an α of 0.05) [$F(3,9)=22.2,\ p<0.001$] and accounted for 43% of the variability of the data. This significant main effect reflects the fact that, on average, threshold ITDs were lower for higher modulation frequencies. The main effect of the raised-sine exponent was also significant [$F(4,12)=52.0,\ p<0.001$] and accounted for 22% of the variability in the data. This significant main effect reflects the fact that, on average, threshold ITDs decreased with increases in the value of the exponent. The interaction between frequency of modulation and value of raised-sine exponent was also significant [$F(12,36)=2.5,\ p<0.02$] and accounted for 6% of the variability in the data. This reflects the finding that the magnitudes of the *relative* changes in threshold ITD produced by changes in the raised-sine exponent depended on the frequency of modulation. Overall, the ANOVA reveals that 71% of the variability in the relative magnitudes of the threshold ITDs calculated across the four listeners is accounted for by the stimulus variables.

III. EXPERIMENT 2

The new data obtained in experiment 1 revealed that, in general, increasing the exponent of the raised-sine led to decreases in threshold ITD. It occurred to us that it would be fruitful to measure threshold ITDs while manipulating depth of modulation of raised-sine stimuli. The motivation for doing so follows directly from the fact that sensitivity to ITDs conveyed by the envelopes of high-frequency two-tone complexes and SAM tones has been shown to vary directly with their depths of modulation (e.g., McFadden and Pasanen, 1976; Nuetzel and Hafter, 1981). In addition, by varying, in a parametric fashion, both the exponent and the depth of modulation of the raised-sine stimuli, one could assess not only the separate influences of those variables on ITD-discrimination but also any interactive influences between them.

A. Procedure

Threshold ITDs were obtained for 4-kHz-centered raised-sine stimuli having exponents of 1.0, 1.5, and 8.0 at a rate of modulation of 128 Hz. For each of the three raised-sine exponents, thresholds were measured at indices of modulation (m) of 0.25, 0.5, 0.75, and 1.0. The 128-Hz rate of modulation was chosen because (as shown in Fig. 2) it yielded the smallest values of threshold ITD and, therefore, would provide the largest "dynamic range" for observing the expected increases in threshold ITD that would result from reductions in depth of modulation from 1.0. The general procedures were those described under experiment 1. For this experiment, however, thresholds were collected in pairs (rather than in triplets) and a total of 10 thresholds was collected for each listener and condition.[3]

B. Results and discussion

Figure 3 displays the mean normalized threshold ITDs, calculated across four listeners, three of whom participated in experiment 1. Once again, the normalization was accomplished by dividing an individual listener's threshold ITDs by that listener's threshold ITD obtained with a SAM tone (raised-sine exponent equal to 1.0) having a frequency of modulation of 128 Hz. The error bars represent +1 standard error of the mean. One of the four listeners was unable to consistently perform the task when the raised-sine exponent was 1.0 and the index of modulation was 0.25. When thresholds could be obtained from this listener and condition, they were in the region of 1 ms. For purposes of computing the normalized threshold ITDs, this listener's threshold for that condition was coded as 1 ms. The time-waveforms corresponding to four of the stimuli are depicted atop their corresponding bars.

Each of the four sections of the figure contains data obtained with a single depth of modulation and raised-sine

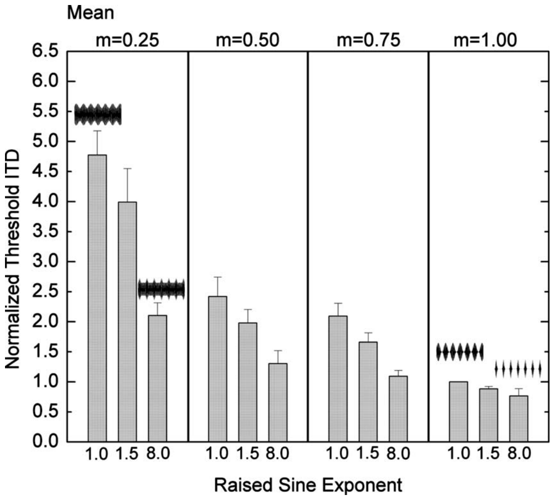


FIG. 3. Mean normalized threshold ITDs. Each of the four sections of the figure contains data obtained with a single depth of modulation and raised-sine exponents having values of 1.0, 1.5, or 8.0. Error bars represent +1 standard error of the mean. The time-waveforms corresponding to four of the stimuli are depicted atop their corresponding bars.

exponents having values of 1.0, 1.5, or 8.0. Consistent with the results of experiment 1, threshold ITDs decreased with increases in the exponent of the raised-sine. In addition, threshold ITDs increased as the index of modulation was decreased from 1.00 to 0.25. Finally, it appears that changes in the exponent of the raised-sine stimuli produced the largest changes in normalized threshold ITD when the index of modulation was 0.25. The data were subjected to the same type of two-factor (four indices of modulation ×three exponents), within-subjects ANOVA described earlier. In accord with visual inspection of Fig. 3, the main effect of raised-sine exponent was significant (again assuming an α of 0.05) $[F(2,6)=74.4,\ p<0.001]$ and accounted for 17% of the variability of the data. The main effect of index of modulation was also significant $[F(3,9)=28.6,\ p<0.001]$ and accounted for 62% of the variability in the data. Finally, the interaction between raised-sine exponent and index of modulation was also significant $[F(6,18)=8.1,\ p<0.001]$ and accounted for 9% of the variability in the data. Overall, the ANOVA reveals that 86% of the variability in the relative magnitudes of the threshold ITDs calculated across the four listeners is accounted for by the stimulus variables.

The two significant main effects were not unexpected. First, the results of experiment 1 showed that threshold ITDs decreased with increases in the exponent of the raised-sine, and the new data in Fig. 3 indicate that general relation held across different depths of modulation. Second, several studies employing SAM tones have demonstrated that threshold ITDs increase with decreases in the index of modulation when ITDs are conveyed by the envelopes of high-frequency stimuli (e.g., McFadden and Pasanen, 1976; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1996a). The data in Fig. 3 indicate that the same general relation also holds for raised-sine stimuli having exponents of 1.5 and 8.0.

The significant interaction between the two main effects reflects the fact that the degree to which increases of the exponent of the raised-sine led to decreases in normalized threshold ITDs depended on the index of modulation. Specifically, when the raised-sine exponent was increased from 1.0 to 8.0, normalized threshold ITDs decreased by 2.7, 1.1, 1.0, and 0.2 for indices of modulation of 0.25, 0.5, 0.75, and 1.00, respectively. This type of outcome, which could not have been discovered without varying parametrically the exponent and depth of modulation of the raised-sine stimuli, was not expected. That is, *a priori*, we had no reason to suspect that increasing the exponent of the raised-sine stimuli would enhance sensitivity to changes in ITD to a greater extent for stimuli having a low index of modulation as compared to stimuli having a high index of modulation.

IV. QUANTITATIVE ACCOUNTS OF THE DATA

Predictions of the threshold ITDs in Fig. 2 were obtained via a cross-correlation-based model that incorporated an initial stage of gammatone-based bandpass filtering at 4 kHz (see Patterson *et al.*, 1995), "envelope compression" (exponent=0.23), square-law rectification, and low-pass filtering at 425 Hz to capture the loss of neural synchrony to the fine-structure of the stimuli that occurs as the center frequency is increased (Weiss and Rose, 1988). As discussed below, the model also includes a second stage of low-pass filtering designed to attenuate spectral components of the envelope above 150 Hz. The reader is referred to Bernstein and Trahiotis (2002) for further details.[4]

In order to account for the data, it was assumed that the listener's threshold ITDs reflect a constant change of the normalized interaural correlation (the value of the cross-correlation at "lag-zero") from 1.0 (the interaural correlation of each diotic reference stimulus). This type of general model and strategy has provided accurate predictions regarding binaural detection and extents of ITD-based laterality for data obtained with a wide variety of complex, high-frequency stimuli (e.g., Bernstein and Trahiotis, 1996b, 2002, 2003; Bernstein *et al.*, 1999).

In order to make the predictions, it was necessary to determine functions relating ITD to normalized interaural correlation. This was done separately for each stimulus used in the experiments (i.e., for each particular combination of frequency of modulation, depth of modulation, and exponent). Numerical measures were obtained by implementing the peripheral stages of the model in MATLAB© and then computing the normalized interaural correlation between the model's "left" and "right" outputs for a wide range of ITDs. Then, using a least-squares criterion, polynomials were fitted to the paired values of normalized correlation and ITD. In order to arrive at predicted mean normalized threshold ITDs, we sought the criterion value of normalized interaural correlation that maximized the amount of variance accounted for between predicted and obtained values.

In order to facilitate visual comparisons between data and predictions, the data in Fig. 2 have been re-plotted in Fig. 4 in four separate panels, one panel for each of the four frequencies of modulation. The squares represent the obtained normalized threshold ITDs. The solid and dashed lines represent two sets of predictions that differ only in terms of

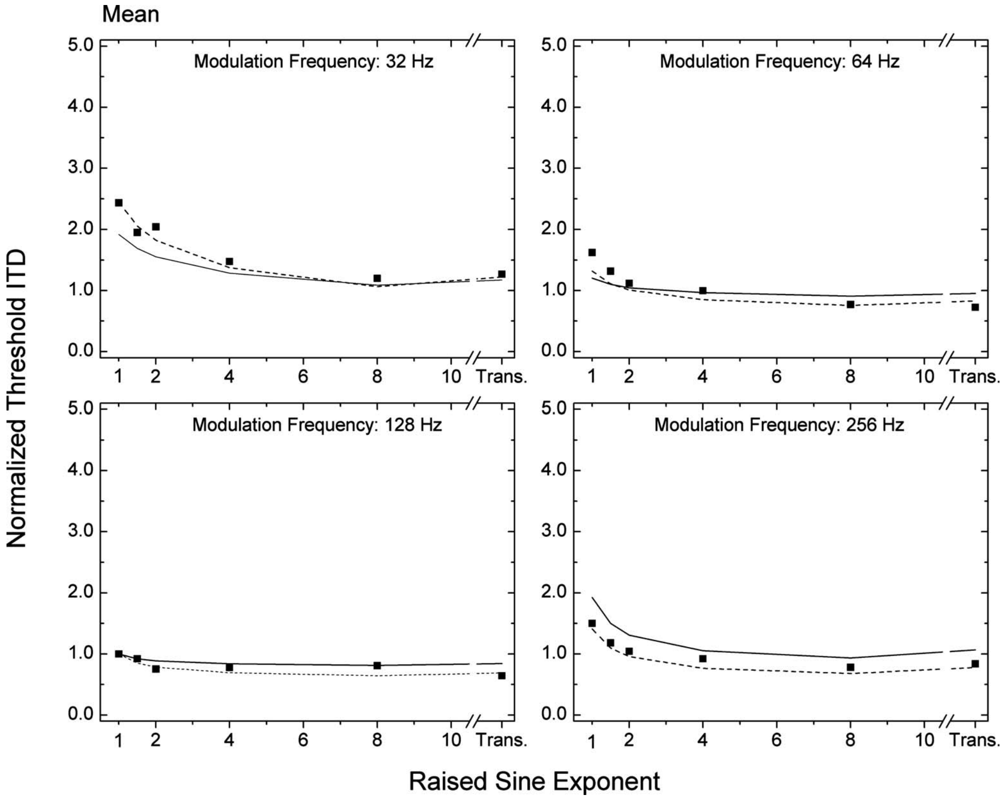


FIG. 4. A re-plotting of the normalized threshold ITDs from Fig. 2 (squares). Each panel contains the data obtained with one of the four frequencies of modulation. The solid lines represent predictions obtained from an interaural correlation-based model incorporating a final stage of second-order (12 dB/octave) low-pass filtering at 150 Hz. The dashed lines represent predictions obtained from the same model but with a final stage of first-order (6 dB/octave) low-pass filtering at 150 Hz.

the order of the 150-Hz low-pass filter applied to the processed stimuli. The predictions represented by the solid lines were generated using the same second-order (12 dB/octave) Butterworth low-pass filter that was employed in our previous studies. Those predictions account for 71% of the variance in the data obtained across the four frequencies of modulation.[5] Note, however, that there are systematic overestimates of normalized threshold ITD for the frequencies of modulation of 128 and 256 Hz. The more accurate predictions represented by the dashed lines were generated using a first-order (6 dB/octave) Butterworth low-pass filter like the one originally used by Kohlrausch *et al.* (2000) and Ewert and Dau (2000) in order to account for temporal modulation transfer functions using sinusoidally amplitude-modulated stimuli. Those predictions account for 93% of the data. This indicates that the interaural correlation-based model captures quite precisely the values of normalized threshold ITDs measured as a function of the value of the exponent of raised-sine stimuli for the range of rates of modulation from 32 to 256 Hz.

The choice of a second-order low-pass filter was originally made by Bernstein and Trahiotis (2002) while attempting to fit threshold ITDs obtained at center frequencies of 4, 6, and 10 kHz with both SAM and transposed tones. Employing a second-order filter provided a better fit than did a first-order filter when the data obtained at all three center frequencies were fitted simultaneously. When the data obtained at the three center frequencies were considered separately, however, predictions made using the first-order filter accounted for 86% of the variance in the data at 4 kHz, while the use of a second-order filter accounted for 83% of the variance. Therefore, it appears that it is neither *ad hoc* nor unreasonable to use a first-order, 150-Hz low-pass filter to account better for threshold ITDs obtained with raised-sine stimuli centered at 4 kHz.

Figure 5 contains the normalized threshold ITDs replotted from Fig. 3 along with two sets of predictions. One set, represented by the closed squares, was calculated via the model using a first-order low-pass filter and the same criterion change in interaural correlation (0.0005) that provided the best-fitting (dashed-line) predictions shown in Fig. 4. Those predictions lead to three important generalizations. First, when the index of modulation is 1.0, the model captures very accurately the changes in normalized threshold ITD that occur when changing the exponent of the raised-sine stimulus from 1.0 to 1.5 to 8.0. The predictions account for 73% of the variance across those three normalized threshold ITDs. It is important to note that the three thresholds shown in Fig. 5 for $m=1.0$ are replications of normalized threshold ITDs included in Fig. 4. The excellent fit to both sets of measures attests to both the consistency of the behavioral thresholds and to the accuracy of the model.

Second, considering only the SAM stimuli (raised-sine exponent of 1.0, left-most bar in each section of Fig. 5), the model captures fairly well the increases in threshold ITD that occur as the index of modulation is reduced from 1.0 to 0.25. The model accounts for 79% of the variance across those four measures. This is logically consistent with the analysis

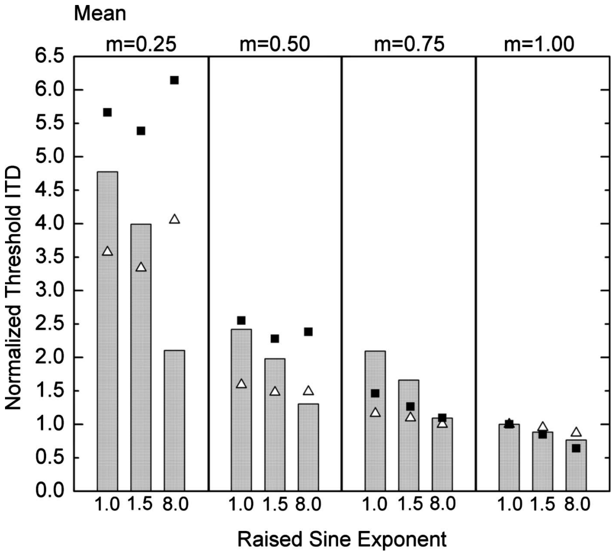


FIG. 5. Normalized threshold ITDs re-plotted from Fig. 3 (bars) along with two sets of predictions (symbols). The predictions were generated via the interaural correlation-based model using a final stage of first-order low-pass filtering at 150 Hz. Predictions represented by the closed squares were calculated using same criterion change in interaural correlation that provided the best-fitting (dashed-line) predictions shown in Fig. 4. Predictions represented by the open triangles were calculated using the criterion correlation that yielded the model's best fit to those thresholds.

of Bernstein and Trahiotis (1996a). They showed that a quantitative account based on normalized interaural correlation could account for the threshold ITDs, taken as a function of the depth of modulation, for high-frequency SAM tones (Nuetzel and Hafter, 1981) and high-frequency two-tone complexes (McFadden and Pasanen, 1976).

Third, and perhaps most important for our purposes, is the fact that the model fails to capture the monotonic decrease in thresholds with increases in the exponent of the raised-sine found at indices of modulation of 0.25 and 0.5. In fact, calculation of the variance accounted for yielded a negative value, indicating that the mean of all of the data provided better predictions than did the model. This failure results mostly because, at those two indices, the model appears consistently to overestimate the normalized threshold ITDs obtained with raised-sine stimuli having an exponent of 8.0.

In order to determine whether this "failure" of the model results from the use of the criterion correlation that best described the data in Fig. 4, new predictions were made for all of the data in Fig. 5 using the criterion correlation that yielded the model's best fit to those thresholds. Those predictions are represented by the open triangles in Fig. 5. Quantitatively, they show an improvement in that the amount of variance in the data accounted for by using the model increased to 54%. This improvement notwithstanding, there are two reasons to evaluate those predictions as problematic. The first is that they also fail to capture the data and their trends at the lowest index of modulation. Accounting for those data was the primary motivation for generating these additional predictions. Second, the criterion change in correlation required to fit the data was an order of magnitude smaller than the criterion changes required to fit the data in Fig. 4 and the threshold ITDs obtained at a center frequency of 4 kHz by Bernstein and Trahiotis (2002).

Several analyses were conducted in attempts to understand why the model failed to predict threshold ITDs obtained with low indices of modulation and a high value of the raised-sine exponent. These included generating predictions after altering the form of the model in the following ways: (1) removing all stages designed to incorporate peripheral auditory processing and then considering only the envelopes calculated via the Hilbert transforms of the stimuli in each channel; (2) increasing the frequency of the low-pass filter designed to attenuate higher spectral components of the envelope of the stimuli; (3) combining the outputs of a series of gammatone filters surrounding the gammatone filter centered at 4 kHz; (4) replacing the gammatone filters by gammachirp filters (e.g., Irino and Patterson, 1997; Unoki *et al.*, 2006; Irino and Patterson, 2006) and conducting analyses using either a single filter centered at 4 kHz or a series of them surrounding 4 kHz [as in (3)]; (5) incorporating values of "modulation gain" reported by Joris and Yin (1992) who measured how changes in the indices of modulation of SAM tones were reflected in indices of modulation of the responses of eighth-nerve units in the cat; (6) changing the type of rectification (linear half-wave vs square-law) and degree of compression (including none).

While none of these manipulations redressed the fundamental shortcomings of the model discussed above, the enterprise proved to be enlightening in one respect. The only way to capture even the trends in the data obtained with the lowest index of modulation was both to increase "operational bandwidth" while simultaneously increasing the cutoff frequency of the low-pass "envelope" filter to at least 200 Hz. This modest success, unfortunately, came at the expense of poorer predictions of normalized threshold ITDs obtained at the higher indices of modulation. Finally, we considered the possibility that the cutoff frequency of the envelope low-pass filter might somehow effectively increase with decreases in the index of modulation. This *ad hoc* notion was rejected because the behavioral data used by Kohlrausch *et al.* (2000) and Ewert and Dau (2000) to place the cutoff frequency of the low-pass filter at 150 Hz were obtained when listeners discriminated between stimuli having no modulation and stimuli having depths of modulations that were just large enough to be detected. That is, the very same 150 Hz low-pass filter that we have shown that enables accurate prediction of threshold ITDs for 100% modulated stimuli was, itself, derived from data obtained with stimuli having very low indices of modulation.

Prompted by a suggestion made by Dr. Wes Grantham, we investigated whether the failure of the model to account for the interaction could be redressed by considering the displacements and patterning of interaural correlation *functions* rather than only its value at lag-zero (i.e., the normalized interaural correlation). To do so, we evaluated changes in the "mean-to-sigma" properties along the delay axis of raised-sine stimuli having exponents of either 1.0 or 8.0. In order to evaluate whether this type of mean-to-sigma approach would account for the observed interaction, we determined the relative increase in the width (variance) of the peak of each function that resulted from decreasing depth of modulation from 1.0 to 0.25. We then compared those relative increases across the two stimuli, one having an exponent of 1.0, the other having an exponent of 8.0 and found them to be, for all practical purposes, identical. This suggests that, within the experiment, in order to overcome the reduction in depth of modulation, the same relative increase in interaural delay would be required to reach threshold for raised-sine stimuli having exponents of 1.0 (SAM) or 8.0. Consequently, the interaural correlation-based model fails to predict the interaction between the value of the exponent and the depth of modulation of raised-sine stimuli independent of whether one considers only activity at lag-zero (the normalized correlation) or mean-to-sigma displacements of the peak of the correlation function along the delay axis. At this time, we can offer no satisfactory way to either change or augment the general interaural correlation-based model in a manner that allows it to capture the interaction in the data between changes in index of modulation and raised-sine exponent.

V. SUMMARY AND CONCLUSIONS

The purpose of this study was to determine how the discriminability of ongoing ITDs is affected by systematic and graded changes in temporal features of such stimuli. To that end, two experiments were conducted. One focused on determining how varying the exponent of raised-sine stimuli affects threshold ITDs. In both experiments, the set of raised-sine stimuli included conventional SAM tones (i.e., raised-sine stimuli having an exponent of 1.0). Overall, the data indicate that graded increases in the exponent led to graded decreases in envelope-based threshold ITDs. The improvements were found to be largest for raised-sine stimuli having a rate of modulation of 32 Hz where thresholds were, overall, the highest. Second, threshold ITDs decreased with increases in rate of modulation from 32 to 128 Hz and then increased slightly when the rate of modulation was increased to 256 Hz. The latter trend was found previously with SAM and transposed tones by Bernstein and Trahiotis (2002).

The second experiment assessed how parametric changes in both the exponent of the raised-sine and changes in its depth of modulation affect threshold ITDs for raised-sine stimuli having a rate of modulation of 128 Hz. The results showed that threshold ITDs decrease with increases in the exponent of the raised-sine for depths of modulation ranging from 0.25 to 1.0. Second, as reported in previous studies concerning discriminability of envelope-based ITDs, threshold ITDs increased with decreases in the index of modulation. One unexpected finding was an interaction between the value of raised-sine exponent and its depth of modulation such that increasing the exponent of the raised-sine stimuli enhanced sensitivity to changes in ITD to a greater extent for stimuli having a low index of modulation than it did for stimuli having a high index of modulation.

Predictions of the data were generated from an interaural correlation-based model. The model was generally able to capture changes in threshold ITD stemming from changes in the exponent, depth of modulation, or frequency of modulation of raised-sine stimuli. The only aspect of the data for which satisfactory predictions of threshold ITD could not be made (even with a variety of major changes in the nature of the model) was the unexpected interaction between the value of raised-sine exponent and its depth of modulation. This failure of the model suggests to us that some, additional, unknown factor or strategy influences how efficiently listeners process envelope-based ITDs for such stimuli. We believe that this finding is potentially important especially because, for only those stimuli, the listeners' sensitivity to envelope-based ITDs is remarkably greater than can be explained either by a generally successful model interaural correlation-based model or several of its variants.

ACKNOWLEDGMENTS

The authors thank Dr. Toshio Irino for his assistance in implementing the gammachirp filterbank. The authors also thank Dr. Richard Freyman, Dr. Wes Grantham, and one anonymous reviewer whose helpful and insightful comments on an earlier version of the manuscript resulted in an improved presentation. This research was supported by research Grant Nos. NIH DC-04147 and DC-04073 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

[1] Equation (1) differs from the one published by John *et al.* (2002) in that we have corrected a typographical error concerning where their parentheses were placed. The corrected equation produces stimuli identical to the ones they used to illustrate the method.

[2] This should not be taken to mean that we are suggesting that it is the peakedness/sharpness of the envelopes of our stimuli, per se, that determines sensitivity to differences in ITD. It should be recognized that the relative peakedness/sharpness of the envelopes of these 100%-modulated stimuli covaries with other characteristics of their temporal signatures, including: 1) the "dead-time" between individual lobes of the envelopes and 2) the slope of the transition from dead-time to the re-emergence of the envelope's positive voltage. Analyses of normalized threshold ITDs plotted against measures of peakedness (defined as the "width" of an individual lobe at 50% or 80% of its peak value) or dead-time revealed that, while either variable could account for variations in those thresholds at a given rate of modulation, neither could account for them across rates of modulation. Said differently, neither similar values of peakedness/sharpness nor similar values of dead-time led to similar threshold ITDs. In any case, one would not expect dead-time to be *generally* useful because that metric would not vary in a systematic fashion where depth of modulation also varied. This is so because decreasing the depth of modulation from 100% would eliminate any straightforwardly defined meaning of dead-time. As will be seen when the data from experiment 2 are discussed, graded decreases in depth of modulation lead to graded increases in threshold. It does not appear that this outcome can be straightforwardly captured by only considering either measures of peakedness/sharpness or dead-time. Part of our ongoing program of research is directed toward discovering useful metrics of the temporal signatures of envelopes of high-frequency stimuli having predictive power that is robust against simultaneous changes in several relevant parameters of the stimuli.

[3] These data were collected in the context of another, larger, set of experimental conditions for which no continuous 1.3-kHz low-pass noise was present. Subsequent to the collection of the data reported here, several "spot-checks" were conducted by repeating the measurements in the presence of continuous 1.3-kHz low-pass noise. The presence or absence of the low-pass noise produced no substantial or systematic affects on the measured thresholds. As will be seen when the data from experiment 2 are discussed, threshold ITDs for conditions that overlap with those measured in experiment 1 were essentially identical.

[4] As discussed in detail by van de Par and Kohlrausch (1998) and by Bernstein *et al.* (1999), the characteristics of the compression observed in the response of the basilar membrane are appropriately modeled by applying

compression to the time-varying magnitude (i.e., the envelope) of the stimulus. In accord with the procedures detailed by Bernstein *et al.* (1999), this was accomplished within the model employed here by compressing the *Hilbert envelope* of the stimulus subsequent to bandpass filtering. Note that after the compressed waveform was passed through the subsequent stages of the model (i.e., square-law rectification and low-pass filtering), the resulting envelope function was *not* equivalent to the compressed Hilbert envelope.

[5] The formula used to compute the percentage of the variance for which our predicted values of threshold accounted was $100\times(1-[\Sigma(O_i-P_i)^2]/[\Sigma(O_i-\bar{O})^2])$, where O_i and P_i represent individual observed and predicted values of threshold, respectively, and $\bar{O}$ represents the mean of the observed values of threshold (e.g., Bernstein and Trahiotis, 1994).

Bernstein, L. R., and Trahiotis, C. (**1994**). “Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise,” J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**1996a**). “The normalized correlation: Accounting for binaural detection across center frequency,” J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., and Trahiotis, C. (**1996b**). “The normalized correlation: Accounting for binaural detection across center frequency,” J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., and Trahiotis, C. (**2002**). “Enhancing sensitivity to interaural delays at high frequencies by using ‘transposed stimuli’,” J. Acoust. Soc. Am. **112**, 1026–1036.

Bernstein, L. R., and Trahiotis, C. (**2003**). “Enhancing interaural-delay-based extents of laterality at high frequencies by using ‘transposed stimuli’,” J. Acoust. Soc. Am. **113**, 3335–3347.

Bernstein, L. R., and Trahiotis, C. (**2004**). “The apparent immunity of high-frequency ‘transposed’ stimuli to low-frequency binaural interference,” J. Acoust. Soc. Am. **116**, 3062–3069.

Bernstein, L. R., and Trahiotis, C. (**2005**). “Measures of extents of laterality for high-frequency ‘transposed’ stimuli under conditions of binaural interference,” J. Acoust. Soc. Am. **118**, 1626–1635.

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). “Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise,” J. Acoust. Soc. Am. **103**, 2069–2078.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). “The normalized correlation: Accounting for NoSπ thresholds obtained with Gaussian and ‘low-noise’ masking noise,” J. Acoust. Soc. Am. **106**, 870–876.

Buell, T. N., and Hafter, E. R. (**1988**). “Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times,” J. Acoust. Soc. Am. **84**, 2063–2066.

Dreyer, A., and Delgutte, B. (**2006**). “Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization,” J. Neurophysiol. **96**, 2327–2341.

Ewert, S. D., and Dau, T. (**2000**). “Characterizing frequency selectivity for envelope fluctuations,” J. Acoust. Soc. Am. **108**, 1181–1196.

Griffin, S. J., Bernstein, L. R., Ingham, N. J., and McAlpine, D. (**2005**). “Neural sensitivity to interaural envelope delays in the inferior colliculus of the guinea pig,” J. Neurophysiol. **93**, 3463–3478.

Hays, W. L. (**1973**). *Statistics for the Social Sciences* (Holt, Rinehart, and Winston, New York).

Irino, T., and Patterson, R. D. (**1997**). “A time-domain, level-dependent auditory filter: The gammachirp,” J. Acoust. Soc. Am. **101**, 412–419.

Irino, T., and Patterson, R. D. (**2006**). “A dynamic compressive gammachirp auditory filterbank,” IEEE Trans. Audio, Speech, Lang. Process. **14**, 2222–2232.

John, M. S., Dimitrijevic, A., and Picton, T. (**2002**). “Auditory steady-state responses to exponential modulation envelopes,” Ear Hear. **23**, 106–117.

Joris, P. X., and Yin, T. C. (**1992**). “Responses to amplitude-modulated tones in the auditory nerve of the cat,” J. Acoust. Soc. Am. **91**, 215–232.

Keppel, G. (**1973**). *Design and Analysis: A Researchers Handbook* (Prentice-Hall, Englewood Cliffs, NJ).

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). “The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,” J. Acoust. Soc. Am. **108**, 723–734.

Levitt, H. (**1971**). “Transformed up-down methods in psychoacoustics,” J. Acoust. Soc. Am. **49**, 467–477.

McFadden, D., and Pasanen, E. G. (**1976**). “Lateralization at high frequencies based on interaural time differences,” J. Acoust. Soc. Am. **59**, 634–639.

Moore, B. C. J. (**1997**). in “Frequency analysis and pitch perception,” *Handbook of Acoustics*, edited by M. Crocker (Wiley, New York), Vol. **III**, pp. 1447–1460.

Nuetzel, J. M., and Hafter, E. R. (**1976**). “Lateralization of complex waveforms: Effects of fine-structure, amplitude, and duration,” J. Acoust. Soc. Am. **60**, 1339–1346.

Nuetzel, J. M., and Hafter, E. R. (**1981**). “Discrimination of interaural delays in complex waveforms: Spectral effects,” J. Acoust. Soc. Am. **69**, 1112–1118.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). “Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,” J. Acoust. Soc. Am. **98**, 1890–1894.

Stecker, G. C., and Hafter, E. R. (**2002**). “Temporal weighting in sound localization,” J. Acoust. Soc. Am. **112**, 1046–1057.

Unoki, M., Irino, T., Glasberg, B., Moore, B. C. J., and Patterson, R. D. (**2006**). “Comparison of the roex and gammachirp filters as representations of the auditory filter,” J. Acoust. Soc. Am. **120**, 1474–1492.

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., and Kohlrausch, A. (**1998**). “Diotic and dichotic detection using multiplied-noise maskers,” J. Acoust. Soc. Am. **103**, 2100–2110.

Weiss, T. F., and Rose, C. (**1988**). “A comparison of synchronization filters in different auditory receptor organs,” Hear. Res. **33**, 175–180.

Release and re-buildup of listeners' models of auditory space

Rachel Keen[a)]
University of Virginia, Charlottesville, Virginia 22901

Richard L. Freyman[b)]
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

(Received 4 August 2008; revised 11 February 2009; accepted 13 February 2009)

When listeners hear sound presented repeatedly in a room with reflections, echo threshold rises. The current experiments tested how long this buildup in echo threshold would last when exposure to a different simulated space (designated as room B) intervened before returning to the original space (designated room A). Stimuli were trains of lead–lag click pairs (room A) and trains of clicks with no reflections (room B) in an ABA sequence. After buildup in room A, echo threshold for click pairs in room A decreased in direct relation to amount of intervening exposure to room B. After 11 click pairs of room B, the effect of exposure to room A was gone. A second buildup in echo threshold in room A was not differentially affected by prior exposure to room A or a different simulated room, room C. Listeners appear to form a model when exposed to sound in a particular space, which is lost quickly upon hearing sound in a different space. Storing previous models is inefficient because the processes of buildup and breakdown occur quickly to sound in a new space.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097472]

PACS number(s): 43.66.Pn, 43.66.Qp [BCJM] Pages: 3243–3252

I. INTRODUCTION

Imagine that you are walking through a house while continuously conversing with an acquaintance. As you progress through the house, the acoustics will vary from room to room, depending on room size, height of ceiling, furniture, rugs, curtains, and so on. In addition to sound waves from the original source, reflections from various surfaces in the room will bounce back to create an overall perception of the room's acoustics. Unless the room is large enough to create long delays between the sound source and its reflections, the reflected sounds will be below echo threshold. That is, they will not be heard as separate sounds localized apart from the original source. Rather, the reflections fuse in location with the original sound, exerting their effect by changing the timbre and loudness of the fused sound (Blauert, 1997; Litovsky *et al.*, 1999). The fusion of the original and reflected sounds into a single image that is perceived at the location of the original sound is referred to as the precedence effect and serves to enable the listener to localize the sound source with reasonable accuracy despite the presence of potentially conflicting directional information from reflections.

In addition to aiding sound localization, we have proposed that the precedence effect plays an important role in informing the listener about a room's acoustics (Clifton *et al.*, 1994; Clifton and Freyman, 1997). Reflected sound, although below echo threshold, is analyzed to form a model of the auditory space. Writing about musical acoustics, Benade (1976) (pp. 208–210) proposed such a process as part of the precedence effect, noting that it informed the listener about objects in the room, distance of walls, ceiling, and so on. Benade (1976) claimed that the listener was very sensitive to frequency and amplitude differences between reflections and the original sound, a claim that has been upheld by later empirical work (Bech, 1995, 1996, 1998). Our research has borne out Benade's (1976) intuitions about this aspect of the precedence effect. Furthermore, we hypothesized that the listener develops expectations about the room's acoustics based on a model formed by ongoing input (Clifton *et al.*, 1994; Clifton and Freyman, 1997). When these expectations are violated by incompatible input, the model collapses, to be replaced by a new model built on the latest input. We refer to this hypothesized process as the "room-acoustics model" in this paper.

This process of buildup and breakdown of the model can be monitored by measuring the listener's echo threshold (for review, see Clifton and Freyman, 1997). It can be raised (called buildup) or lowered (called breakdown), depending on the ongoing sound. A typical trial sequence is to present a train of repeated click pairs followed by a test click pair that is identical to the train (testing buildup) or differs from the train (testing breakdown). The control condition is the test click presented in isolation, without the preceding train. The listener's task is to respond to the test click, indicating whether or not the lagging sound was heard at a location separate from that of the leading sound. While repeating the same click pairs raises echo threshold (Freyman *et al.*, 1991), it can then be lowered by introducing any of several changes. All of the following changes from the repeating train will lower echo threshold: a test click with a different delay between lead and lag from what was heard in the repeating train (Clifton *et al.*, 1994), a lag sound with a different spectrum from what it had in the train while the lead remains unchanged (McCall *et al.*, 1998), and when there is a switch in lead and lag sound locations (Clifton, 1987; Clifton and

[a)]Formerly Rachel K. Clifton.

[b)]Author to whom correspondence should be addressed. Electronic mail: rlf@comdis.umass.edu

Freyman, 1989; Blauert and Col, 1992; Yost and Guzman, 1996). Clifton and Freyman (1997) proposed that only those changes that give information about the room's acoustics would affect echo threshold. For example, a different time delay between source and reflection would signal that the reflecting surface had moved, and a different spectrum for the delayed sound echo would signal that some property of the reflecting surface had changed.

A change between train and test click that does not signal a change in room acoustics does not affect echo threshold. For example, when both lead and lag sounds were changed in frequency or in intensity between train and test, echo threshold was unchanged (Clifton *et al.*, 1994). Congruent changes signal that the source changed, and the "echo" reflected this change. As long as congruent changes in both source and reflections are made, echo threshold is not affected (Clifton, *et al.*, 1994; Yost and Guzman, 1996). This conception of how changes in echo threshold are produced is akin to the "plausibility hypothesis" proposed by Rakerd and Hartmann (1985) and Hartmann and Rakerd (1989). They noted that subjects in time-intensity trade experiments tend to discount values of interaural time delay (ITD) that are outside the range that could be plausible given their head size and location within the room. Rakerd and Hartmann (1985) found that when the ITD cue was implausible, listeners responded on the basis of the interaural intensity difference (IID) cue. In their case and ours, it is proposed that listeners are influenced by their implicit knowledge about how sound behaves in a room when making localization judgments.

By measuring echo threshold under varying conditions, we can examine timing parameters for buildup and breakdown of the precedence effect. The time course of these processes can tell us not only about the functioning of the auditory system but also about the nature of the neural processes themselves. Shifts in echo threshold reflect a decision-making process in the brain. Sanders *et al.* (2008) found larger negativity in the event related potentials at anterior-central and medial sites when listeners reported hearing an echo, than when not hearing an echo, for the same click pair at a delay chosen to be at each listener's echo threshold. The mystery of the precedence effect is why the brain sometimes hears the reflected sound as an independent sound source and at other times suppresses its location information while preserving information about room acoustics. There are many unanswered questions about this process. For example, once buildup in echo suppression has occurred, how long will it last? Does it decay naturally with the passage of time? Using the analogy of walking through rooms while talking, if the talkers fell silent for a few moments while standing in the room, would the listener's model of that room's acoustics collapse? Djelani and Blauert (2000) tested the effect of 1, 4, and 9 s of silence between train and test click. They found that 1 s of silence had little effect, but the longer periods led to monotonic decreases in echo threshold. The time course of the model's decay needs to be systematically examined in relation to echo threshold for a click pair with no preceding train that produces buildup.

The room acoustics hypothesis specifies what type of input would disrupt the model, but not how the process unfolds. Our previous research featured a train of identical click pairs followed by a test click pair that differed from those in the train. This procedure tested whether the listener detected a difference between train and test, evidenced by a lower threshold to the changed test click compared to a test click like those in the train. However, this procedure leaves unanswered the question of whether echo threshold for the stimuli in the ongoing train was disrupted by changing the test click. Djelani and Blauert (2001) found that a brief injection of one aberrant click pair (they reversed the location of the lag) at the end of the train had no effect on echo threshold for a test click pair identical to those in the train. After replicating that basic result, Freyman and Keen (2006) (Exp. 3) increased the number of aberrant sounds to five consecutive presentations before representing the original click pair configuration. Echo threshold for the original configuration decreased by several milliseconds. We concluded that breakdown of a room model does not occur instantly, but like buildup it is a graded process that depends on the amount of input. To understand the time course of a model's decay, the full range of input, from completely ineffective to that sufficient to produce complete breakdown, needs to be investigated.

Another critical question is whether there are lingering effects of buildup such that re-exposure to the same acoustic parameters after breakdown would show savings. To use the room analogy again, if after exposure to other rooms the listener came back to a previously visited room, would buildup be faster the second time around? This issue is related to whether models of familiar rooms are stored in the brain. A study by Robart and Rosenblum (2005) suggests that this is possible. They found that listeners could identify in which of several rooms a sound had been recorded in (e.g., a gym, restroom, classroom), suggesting that models of various spaces could be held simultaneously in memory.

The current experiments attempt to answer several questions about the formation and disruption of a listener's model for how sound is reflected in a room. The purpose of experiment 1 was to determine how varying amounts of exposure to a new space or "room" would affect retention of a model of a previous space. Freyman and Keen (2006) found that five exposures to a new space was sufficient to decrease echo threshold for an existing model, but we do not know what the smallest number for causing a disruption is, or if more than five exposures would reduce echo threshold still further. In experiment 2 listeners experienced a second buildup after the initial model had been broken down. After buildup followed by maximum breakdown for one stimulus configuration was accomplished, we tested whether re-exposure to that room would produce a more rapid buildup than that seen initially. Experiment 2 featured manipulations and controls to test whether savings would occur under a variety of conditions.

II. EXPERIMENT 1

The primary purpose of this experiment was to determine the boundaries for minimum and maximum breakdown

TABLE I. List of stimulus conditions for experiment 1.

Segment 1	Segment 2	Segment 3	Name/description
Train	Train	Test	Basic buildup
	Room A5	Room A1	A5|A1
Train	Train	Test	Buildup then breakdown
Room A5	Room B1	Room A1	A5|B1|A1
Room A5	Room B3	Room A1	A5|B3|A1
Room A5	Room B5	Room A1	A5|B5|A1
Room A5	Room B7	Room A1	A5|B7|A1
Room A5	Room B9	Room A1	A5|B9|A1
Room A5	Room B11	Room A1	A5|B11|A1
Train	Train	Test	Buildup then silence
Room A5	S5	Room A1	A5|S5|A1
Room A5	S11	Room A1	A5|S11|A1
Train	Train	Test	Breakdown only
	Room B1	Room A1	B1|A1
	Room B3	Room A1	B3|A1
	Room B5	Room A1	B5|A1
	Room B11	Room A1	B11|A1
Train	Train	Test	No conditioning train
		Room A1	NC

of a room acoustics model. The procedure was to first allow the model to be built with a click train of five exposures that featured a left-side leading click with a right-side lagging click to simulate a single reflective surface. Without interrupting the click train, there immediately followed clicks from only the left side, ending with the test click, which was a single click pair identical to the pairs used in the first train. The amount of " new space" input varied from a single click on the left side (not expected to have an effect) to 11 consecutive left-only clicks (expected to lead to complete breakdown of the previous model). A necessary control in experiment 1 was used to evaluate whether the built-up model would last if an equivalent length of silence intervened rather than the new input before the test click.

A second control tested listeners' response to the test click without prior buildup to that input; that is, a train of clicks from only the left side preceded the test click pair. This condition simulated sound in an anechoic room, followed by presentation of the same sound in the same location but with a simulated reflection present. By using this anechoic condition as the intervening or "new" room in all experimental conditions described below (referred to as room B in Table I), we sought to present a highly contrasting space with the initial buildup condition. In previous research (Freyman *et al.*, 1991), we found that exposure to a delayed sound after hearing single-source sounds in an anechoic condition lowered echo threshold even below the threshold for the isolated test click. In other words, preceding a lead–lag click pair with a train of lead-side-only clicks caused the echo to "pop out," producing very low thresholds of 6 ms. Interestingly, the pop out does not occur, at least to the same extent, if the preceding click train has a different echo delay or location from that of the test click (as opposed to having no echo) (Clifton *et al.*, 1994, 2002). Thus, the contrast between an anechoic space versus a space with echoes appears to be greater than the contrast between two spaces having different echoes. Using the anechoic condition as the intervening space before the test click was expected to widen the range of echo thresholds among experimental conditions, compared to the introduction of a subtle change in room acoustics. The control condition of the anechoic stimulus preceding the test click without the preceding buildup was a necessary check on the powerful influence of the anechoic clicks on the test click after the preceding buildup.

A. Methods

1. Stimuli and apparatus

Stimuli were pairs of computer generated 150-ms pulses presented from two channels of a 16-bit digital/analog converter (TTES QDA1). The outputs of the two signal channels were low-pass filtered at 8.5 kHz (TTE J1390), attenuated (TTES PAT1), amplified (CROWN D40), and delivered to a pair of loudspeakers (Realistic Minimus 7). The loudspeakers rested on a semicircular arc constructed of foam-covered wood that was housed in an anechoic chamber measuring $4.9\times4.1\times3.12$ m. The floor, ceiling, and walls of the chamber were lined with 0.72 m foam wedges. Subjects sat in a chair in the center of the room with the loudspeakers situated at 45° left (−45°) and 45° right (+45°) at a distance of 1.9 m. The center of the loudspeakers was at ear height for the typical listener seated in the chair. The stimulus level was measured by presenting the click stimuli through the loudspeakers at a rate of 4 clicks/s. A microphone was lowered to the position of the center of the subject's head with the subject absent. The microphone output was fed to a sound level meter (B&K 2204) set on the "fast" meter response on the A-scale. Unattenuated outputs through the system were 61 dBA from either loudspeaker. The experiments were run at 43 dBA (with attenuators set to 18 dB).

2. Procedures and conditions

The primary conditions of the experiment employed a train-test method used previously in a number of studies (e.g., Freyman *et al.*, 1991; Clifton *et al.*, 1994; Grantham, 1996; Yost and Guzman, 1996; Yang and Grantham, 1997; Djelani and Blauert, 2001; Freyman and Keen, 2006). On each trial, repeated pairs of clicks (one to each loudspeaker) were delivered at a rate of 4 clicks/s to form a click "train." Following the train and a pause of 750 ms, a "test click" was presented. In all conditions, the test click from the right loudspeaker was delayed relative to that from the left loudspeaker by 2–14 ms in 2-ms steps. The clicks during the train were either click pairs (one from the left loudspeaker and one from the right), single-source clicks from the left loudspeaker, or a sequential mixture of the two. In all cases in which click pairs were used, the delay used during the train matched that for the test click. The listeners' task was to report whether

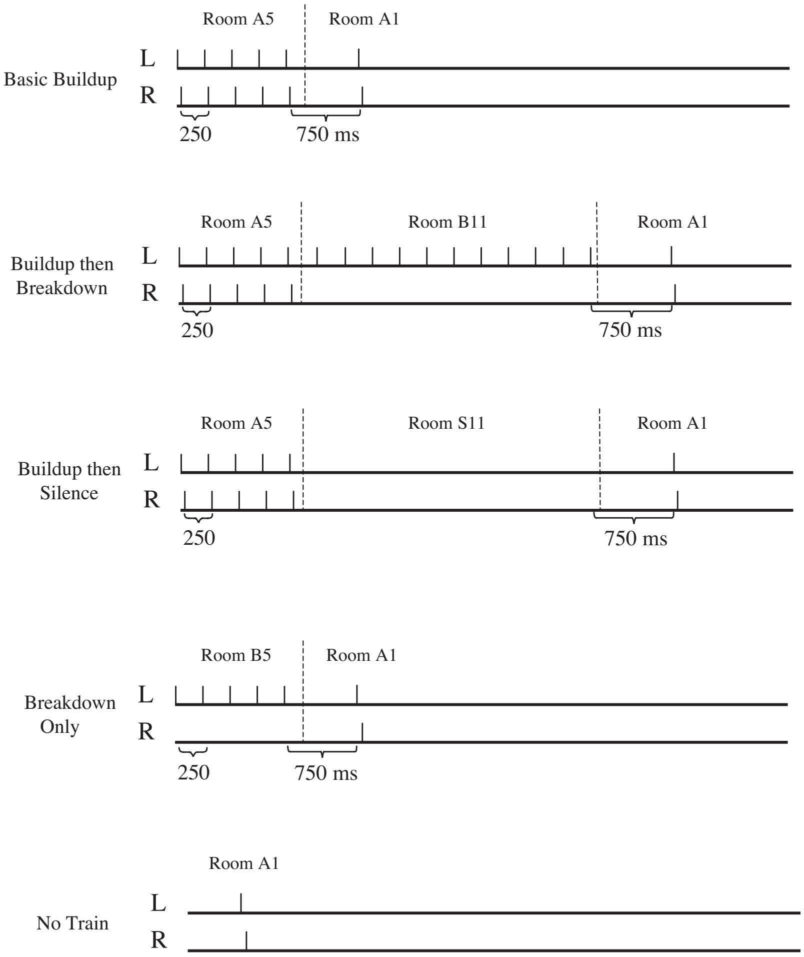


FIG. 1. Schematic illustrations of examples of stimuli used in experiment 1.

they heard a sound in the vicinity of the right (lagging) loudspeaker during the test click. In this design, the test click pair had the same locations (lead left, lag right) in all conditions, so the effects of various conditioning trains on the same test click could be assessed.

Table I lists all conditions tested in this experiment, and Fig. 1 shows example stimuli from the different conditions. The table refers to the lead-lag configurations as representing crude simulations of auditory space referred to as rooms. We call the room with a reflective surface only on the right "room A." The first row of the table shows the basic buildup condition against which other conditions were compared. The train presents five repetitions of room A (called A5), followed by a pause, and then the test click, which is a single presentation of room A (labeled A1). The next six rows illustrate stimulus manipulations intended to lead to breakdown of the listener's acoustic model of room A. Room B presentations are single-source clicks from the left loudspeaker only. The source has not moved relative to its location in room A, but the reflection and, thus, the simulated reflective surface have been eliminated. Room B was presented for 1, 3, 5, 7, 9, or 11 clicks. Thus, these conditions consisted of room A5, room Bn, then room A1 (the test click) at the end of the entire train. The next two rows show the configurations of conditions that control the time lapse in room presentations before the test click. After the initial presentations of room A were completed, the buildup effect of room A might be expected simply to dissipate over time. In order to understand the influence of the room B presentations, it was necessary to determine what the effect would be of presenting no sound during an equivalent period. Two conditions were included, silence for the periods of time equivalent to 5 and 11 clicks (S5 and S11, respectively). Adding the standard pause of 750 ms between train and test, the total silence was 2 s for S5 and 3.5 s for S11. The next four conditions shown in the table were also control conditions. Only the single-source clicks (room B) were presented before the test click. Finally the last condition included no preceding train to determine the response to the test click presented as a solitary sound [no conditioning train (NC)].

The different conditions were run in blocks of 35 trials. The room condition was fixed (e.g., A5|B7|A1) during a block. The seven lag-click delays (2, 4, 6, 8, 10, 12, and 14 ms) were presented five times each within a block in a randomized order. An individual experimental session lasted for approximately 1 h and consisted of a total of 13 blocks, one for each condition. Note that the B11|A1 condition was run later when it was decided that it was necessary for comparison to A5|B11|A1. It was interspersed among the other conditions for experiment 2. The order of blocks was randomized separately for each listener and experimental session. Each listener completed four such sessions so that, across all sessions, a single condition at a specific delay was based on 20 judgments (five repetitions per block × four blocks per condition).

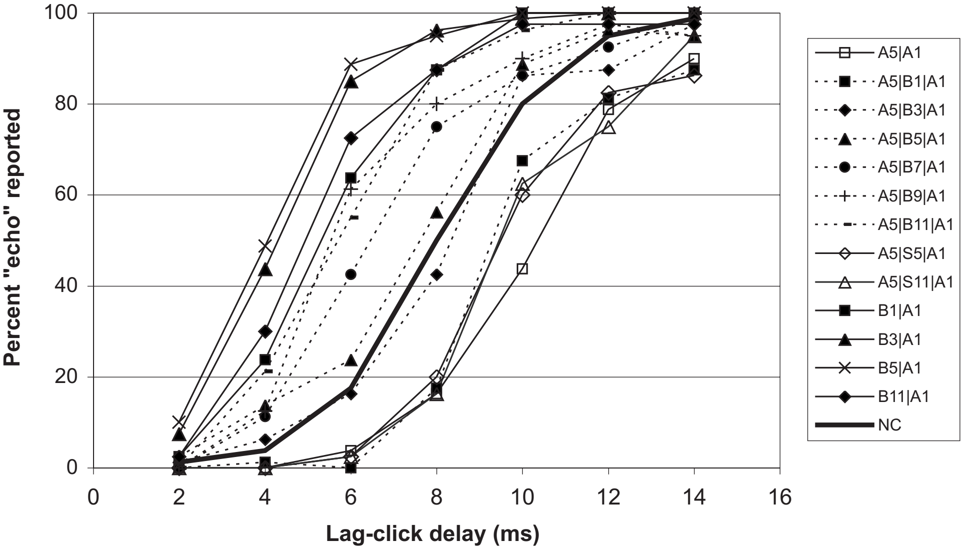


FIG. 2. Mean percentage of trials on which an echo was reported as a function of the delay of the lag click for all conditions of experiment 1. A description of each of the conditions is given in Table I.

3. Listeners

Four graduate students from the University of Massachusetts, all with hearing thresholds ≤20 dB HL (ANSI, 1996) from 500 to 4000 Hz, participated in the study.

B. Results

1. Psychometric functions

To get an overall view of the data, the results for each condition and delay were averaged over subjects to form the group psychometric functions shown in Fig. 2. The percentage of trials on which an echo was reported is plotted as a function of the lag-click delay. As expected, the functions were generally monotonic, showing increasing reporting of echoes as delay was increased; however, these functions also display a large effect of the acoustic stimulation that preceded the test click. It should be noted again that exactly the same test click, i.e., the lead–lag click pair of room A, was the stimulus that subjects responded to in every condition. The thick solid line indicates the function obtained for the NC condition, where this test click was presented in isolation. NC crossed the 50% point around 8 ms. At the delay of 8 ms the percentage of echoes reported ranged from 17% for the basic buildup condition $(A5|A1)$ to 95% for the breakdown condition $B5|A1$, indicating the dramatic influence of context on the listeners' reporting of echoes at the same lead-lag delay. The intermediate values reflect variable amounts of input from the different simulations of acoustic space. Consistent with previous reports (Djelani and Blauert, 2001; Freyman and Keen, 2006), a single instance of room B had little or no effect on buildup, but each additional presentation of room B degraded echo threshold for room A stimuli.

2. Echo thresholds

To quantify the effects observed in Fig. 2, echo thresholds (the delay at which echoes were reported on 50% of the trials) were estimated from the psychometric functions obtained from each individual subject. Each function was fitted with a logistic equation of the form $1/(1+\exp-((t-m)/s))$, where t is the lag-click delay, m is midpoint of the function, and s is the slope. The parameter m represents an estimate of the delay at which 50% echoes were reported, the echo threshold in this case. The fits were generally very good, 85% of the r^2 values being above 0.95. As might be expected, the slopes of the individual functions tended to be slightly steeper than those of the mean functions shown in Fig. 2. No formal comparison of the slopes across conditions was undertaken. However a slight tendency for slopes to be steeper for the functions with lower echo thresholds was apparent in both the mean (e.g., compare B5 and A5 in Fig. 2) and individually fitted functions.

The echo thresholds are displayed in Fig. 3, which plots the group mean thresholds for all conditions shown in Table I. Higher echo thresholds indicate more buildup, that is, more suppression of echoes, and conversely low thresholds indicate that more echoes are heard at shorter delays. The abscissa shows the number of clicks in room B that preceded test click A1, with zero along the axis referring to the basic buildup condition with no room B clicks following buildup

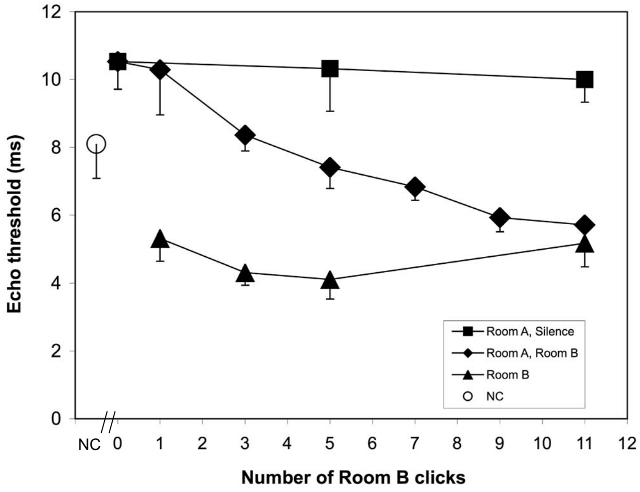


FIG. 3. Mean echo thresholds from experiment 1 as a function of the number of room B clicks presented alone (triangles) or following five room A clicks (diamonds). Squares show results of a control condition using five room A clicks followed by silence equivalent in duration to 0, 5, and 11 room B clicks. Error bars indicate one standard error of the mean.

(A5|A1). Echo threshold was 10.5 ms for buildup in room A with no exposure to room B. The continuation of this line represents the silent control condition (A5|Sn|A1, squares) and shows that, with up to 3.5 s of silence, there was little or no change in built-up threshold. The A5|Bn|A1 condition (diamonds) shows that exposure to room B resulted in a gradual effect, with each additional click reducing the previously built-up threshold. A two-way analysis of variance (ANOVA) comparing the room A conditions (A5, A5|S5, and A5|S11) with the comparable room B conditions (A5|B1|, A5|B5, and A5|B11) yielded a main effect of condition $[F(1,3)=13.21,\ p<0.036]$. There was also a main effect of amount of time since the A5 buildup $[F(2,6)=5.67,\ p<0.042]$, indicating a decrease in echo threshold over time. The effect of increasing the room B input was analyzed with a one-way ANOVA and trend test on the six levels of room B (A5|Bn). A main effect of exposure level was obtained $[F(1,15)=5.94,\ p<0.003]$, with a linear trend that just missed significance $[F(1,3)=8.70,\ p<0.06]$.

Another way to assess the effect of the buildup of the model of room A is to compare the effect of exposure to room B with and without prior exposure to room A (A5|Bn versus Bn). The lower thresholds for the latter condition, shown in Fig. 2, indicate that the initial presentation of the five buildup clicks (A5) still had a strong influence on the reporting of echoes to the room A test click when that room was re-introduced after several presentations of room B. A two-way ANOVA on condition (A5|Bn versus Bn) and amount of room B exposure (B1, B5, and B11 for the two conditions) yielded a main effect of condition $[F(1,3)=22.83,\ p<0.017]$, confirming the lingering effect of the prior room A exposure. There was also a main effect of amount of room B exposure, with echo threshold for A decreasing significantly as room B presentations progressed $[F(2,6)=5.11,\ p<0.05]$.

Reductions in echo threshold began to level off by 9–11 presentations of room B. At this point, the A5|B11|A1 threshold was within 1 ms of the B11|A1 threshold, suggesting that this may be the point where the model of room A was virtually lost. One might assume that during the increasing input of room B the difference between echo thresholds for Bn|A1 and A5|Bn|A1 was due to the influence of prior room A input. This assumption could be wrong because the experience of any other room before the introduction of room B may have had the same elevating effect on echo threshold. To ensure that the response to A1 was affected specifically by prior exposure to room A, a different space (room C) with new locations for lead and lag clicks should replace room A as the initial exposure before room B. In experiment 2, room B was presented in between room C and the test click (A1) to learn how the breakdown created by the single-source clicks progressed under this circumstance. Echo threshold for the test click may be affected by prior exposure to room C because, like room A, room C has a delayed sound, even though it comes from a different direction.

A second way that the room A experience might exert an effect after apparently being replaced by room B is to produce faster buildup during a re-buildup period. The question is whether there would be any savings if room A were reintroduced following 11 repetitions of room B. This gets to the heart of the question of whether we hold on to and store the model of a room so that it can be readily re-activated. In other words, would buildup occur more quickly in room A if the listener had previously experienced buildup in that same space?

III. EXPERIMENT 2

Experiment 2 investigated how the experience in room A continued to exert an effect when followed by presentations of room B. Schematic diagrams of exemplars of all stimulus conditions are illustrated in Fig. 4. In one test we simulated a new room, room C, in which lead clicks were presented from the right and lag clicks from the left. This simulated a reflecting surface on the left side of the room, the opposite of room A. Unlike the presentation of five room A clicks prior to the test click (A5|A1), which built up echo threshold in experiment 1, condition C5|A1 should not raise echo threshold for the test click, A1, above that for NC, the test click presented in isolation. Substituting room C for room A before room B input (CB) provides a control condition for AB from experiment 1 to determine whether the elevated thresholds for the test click were due specifically to the room A buildup.

The second test for a lingering effect of room A was re-exposure to this stimulus after room B experience had apparently obliterated its influence. The model for room A appears to be completely gone after 11 presentations of room B (A5|B11 in experiment 1). We tested whether buildup would take place more quickly the second time by comparing increasing levels of re-exposure to room A after prior buildup/breakdown (ABA) versus exposure to two different rooms (CBA) and prior exposure to room B alone (BA). The latter condition provides a baseline for the effect of no prior exposure to either room A or C. It is possible that prior buildup to room A will have a priming effect such that rebuildup to room A occurs more quickly compared to prior exposure to either room C or room B. This result would suggest that even though echo threshold was very low after B11, some effect of room A in memory served to boost a second buildup quickly. On the other hand if re-exposure to room A follows a similar trajectory regardless of what came before B, this would suggest that the “slate is wiped clean” when there has been sufficient exposure to a new acoustic environment.

A. Methods

The experiment was run with the same four listeners who participated in experiment 1. All equipment and procedures were identical to those used in experiment 1, and echo thresholds were calculated in the same manner. Table II lists all conditions tested in this experiment, and Fig. 4 displays example stimuli from the different conditions. The first four rows of the table show control conditions involving room C. Each condition began with five presentations of room C (the reverse of room A). Results for condition C5|A1 were compared to those for condition A5|A1 (from experiment 1), and

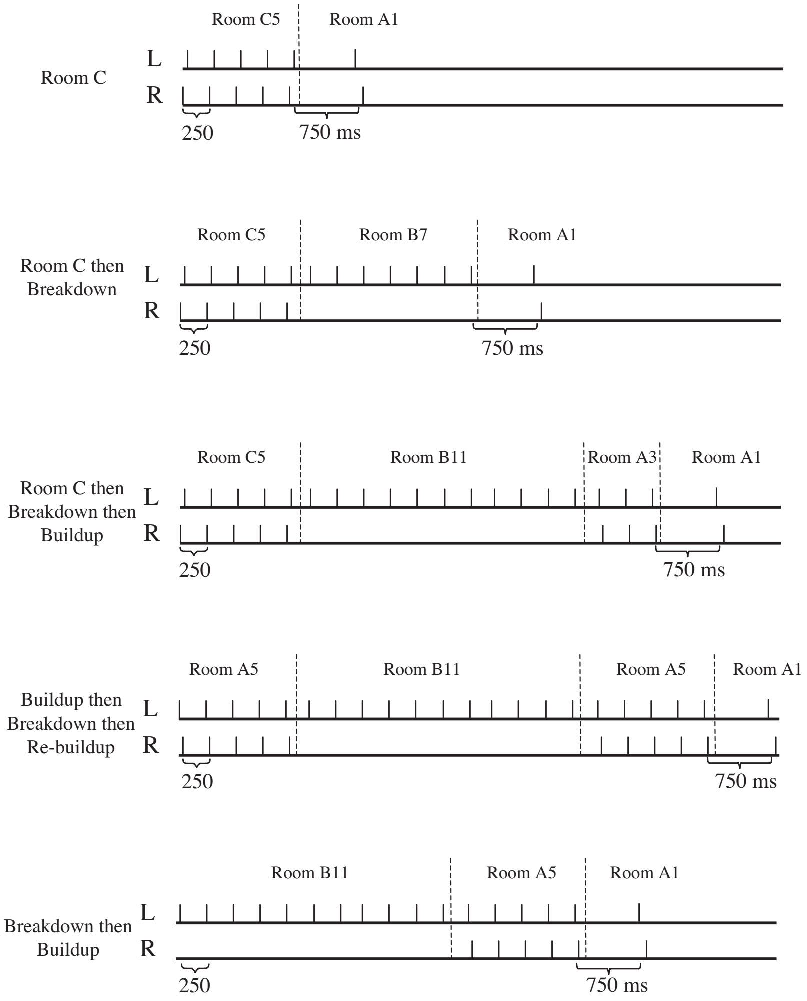


FIG. 4. Schematic illustrations of examples of stimuli used in experiment 2.

three levels of room B exposure (3, 7, and 11) were interspersed between room C input and A1, the test click, for comparison to A5|Bn|A1 from experiment 1. The conditions that tested re-buildup of echo threshold after 11 presentations of room B are shown in the remainder of Table II. The basic condition was five presentations of room A then B11, followed by varying amounts of room A before the test click (A5|B11|An|A1). Values of An were 3, 5, 7, 9, and 11 click pairs. Comparison conditions were C5|B11|An|A1 and B11|An|A1.

Each condition was tested with the same seven lag-click delays (2, 4, 6, 8, 10, 12, and 14 ms) as in experiment 1. All conditions were run in 35-trial blocks as before, with five repetitions of each delay presented in a randomized order within a block. For all conditions there were four blocks per condition across sessions, for a total of 20 trials for each delay and condition combination. The conditions were presented in six experimental sessions. The first four sessions consisted of the conditioning trains A5|B11|An and B11|An, a total of ten conditions. All ten blocks, one for each condition, were presented in a randomized order in each session. The last two sessions consisted of all the conditions involving room C, seven in total. Each of the seven conditions was presented twice in a randomized order during each 14-block session.

B. Results

As expected, hearing five presentations of room C before the test click did not increase echo threshold for room A. As shown in Fig. 5, left panel, the solid circle (mean =8.0 ms for C5|A1) is virtually the same value as the open circle for the isolated test click (mean=8.1 ms for NC). The left panel also displays the effect of the breakdown produced by single-source clicks of room B, and the right panel displays the buildup that occurred with repeated presentations of room A following the breakdown. With the exception of the circles, the data in the left panel are replotted from Fig. 3. We first tested whether the response to the test click differed for conditions A and C. A 2 (condition) ×4 (exposure level to room B) ANOVA tested the decline in echo threshold as room B input increased from zero (A5|A1 and C5|A1) to 11 (A5|B11|A1 and C5|B11|A1). The effect of the number of room B clicks before the test click was significant [$F(3,9)=20.16$, $p<0.0001$], as was the linear trend [$F(1,3)=22.69$, $p<0.018$]. The curves for room C and room A began at different levels, but once input from room B began, they both progressed downward until the means were similar at B11 (mean=5.7 ms for A5|B11|A1 and mean=5.8 ms for C5|B11|A1). There appears to be a general (but rapidly disappearing) effect of having heard five presentations of a delayed sound that elevates echo threshold above those for

TABLE II. List of conditions for experiment 2.

Segment 1	Segment 2	Segment 3	Segment 4	Name/description
Train	Train	Train	Test	Room C
		Room C5	Room A1	$C5 \vert A_1$
Train	Train	Train	Test	Room C, breakdown
	Room C5	Room B3	Room A1	C5|B3|A1
	Room C5	Room B7	Room A1	C5|B7|A1
	Room C5	Room B11	Room A1	C5|B11|A1
Train	Train	Train	Test	Buildup, breakdown, re-buildup
Room A5	Room B11	Room A3	Room A1	A5|B11|A3|A1
Room A5	Room B11	Room A5	Room A1	A5|B11|A5|A1
Room A5	Room B11	Room A7	Room A1	A5|B11|A7|A1
Room A5	Room B11	Room A9	Room A1	A5|B11|A9|A1
Room A5	Room B11	Room A11	Room A1	A5|B11|A11|A1
Train	Train	Train	Test	Room C, breakdown, buildup
Room C5	Room B11	Room A3	Room A1	C5|B11|A3|A1
Room C5	Room B11	Room A7	Room A1	C5|B11|A7|A1
Room C5	Room B11	Room A11	Room A1	C5|B11|A11|A1
Train	Train	Train	Test	Breakdown, buildup
	Room B11	Room A3	Room A1	B11|A3|A1
	Room B11	Room A5	Room A1	B11|A5|A1
	Room B11	Room A7	Room A1	B11|A7|A1
	Room B11	Room A9	Room A1	B11|A9|A1
	Room B11	Room A11	Room A1	B11|A11|A1

the anechoic room B condition. It does not appear to matter which direction the delayed sound came from, as room A versus C had no effect after 11 presentations of room B clicks.

The effect of re-exposure to room A following 11 presentations of B can be stated simply: it does not appear to matter whether A or C came before, at least with the statistical power available with a small *N*. Although the B11|An curve appears to be lower than the other curves, an analysis testing whether the three curves in the right panel of Fig. 5 were different showed that they were not. Thresholds for condition A5|B11|An did not differ from those for C5|B11|An [$F(1,3)<1.0$,N.S.] or from B11|An [$F(1,3)=4.06$, $p>0.10$]. Level of exposure (or re-exposure) to room A was, of course, highly significant for both follow-up ANOVAs [$F(3,9)=9.14$, $p<0.004$ for the ABA and CBA comparison and $F(4,12)=11.55$, $p<0.0001$ for the ABA and BA comparison].

IV. DISCUSSION

In the course of a normal day, people will experience not two or three acoustic spaces, but dozens, as they move about their environment. In two experiments we have shown that

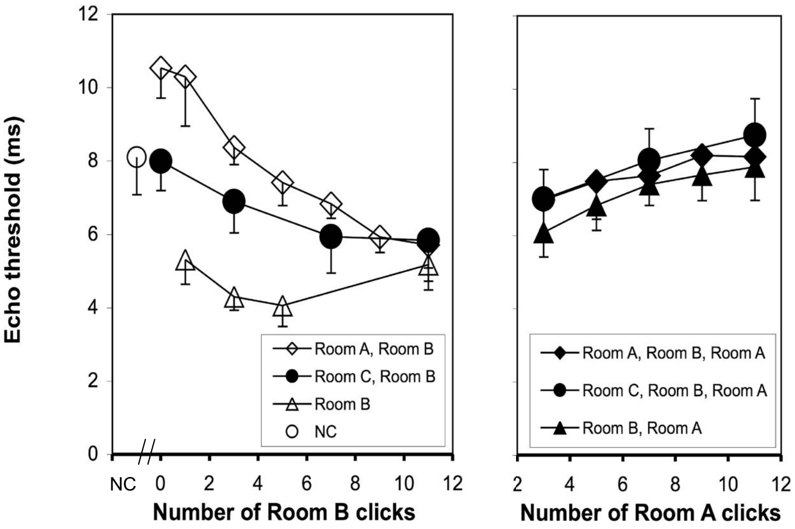


FIG. 5. Mean echo thresholds as a function of the number of clicks presented with (diamonds, circles) or without (triangles) an initial train of five clicks. Error bars indicate one standard error of the mean. Data in the left panel show the effect of breakdown with increasing number of room B clicks and are, with the exception of the filled symbols, replotted from experiment 1. The right panel shows data from experiment 2 indicating buildup/re-buildup from the additional room A clicks that followed the room B clicks. See Table II for a description of all conditions.

people form models of an acoustic space when exposed to sound and its attendant delayed reflections in that space. Furthermore, the model for a particular space is disrupted when the listener is exposed to a new space having different acoustic properties. The processes of building up and breaking down models are dependent on the amount of input from each spatial configuration of sound. In experiment 1 we charted the course of disruption of a room model once built up. Following five presentations of sounds in room A (which was sufficient to build up echo threshold to about 10.5 ms), we inserted varying amounts of exposure to room B (1, 3, 5, 7, 9, and 11 inputs). Echo threshold for room A steadily decreased so that after 11 presentations of room B, threshold was depressed to about 5 ms, which was comparable to one exposure to room A preceded only by presentations of room B. It was as if room A had never been experienced. This process can be thought of as a competition between two room models, with the more recently experienced room B model ultimately winning out.

In experiment 2 we tested whether there were savings in the buildup process by re-exposing listeners to room A for a second time. This was done after 11 presentations of room B had driven echo threshold for room A down to a low level. We found that buildup to room A progressed similarly under three conditions: prior exposure to room A (ABA), prior exposure to a space with different echoes (CBA), or only anechoic exposure (BA). In other words, buildup was independent of whether the listener had previously experienced that room.

Although the rooms we simulated are unusual and would not be encountered very often outside the laboratory, their very simplicity allows us to draw certain conclusions about how the auditory system handles exposure to the varied spaces we do inhabit. First, just as there is buildup in echo threshold when one hears sound produced within a space that has reflections, there is breakdown of that threshold when the listener enters a new space. The breakdown does not occur immediately upon entry into the new space (Djelani and Blauert, 2001; Freyman and Keen, 2006) but depends on the amount of input. Buildup and breakdown are incremental processes, with each instance of new input having a cumulative effect. The processes can take place rapidly and reach asymptote quickly because only 9–11 inputs are required to reach maximum echo threshold for buildup (Freyman *et al.*, 1991) or to reach lowest echo threshold for breakdown, as in the current data set. Our experience in real rooms can seem like an immediate adjustment to echoes because most sounds (speech, music, and noise) have numerous onsets occurring rapidly. It is only when the listener is exposed to punctate, countable sounds like brief clicks that buildup and breakdown in echo threshold become noticeable and can be measured. We can slow the process down by presenting click pairs slowly; at rates of 1/s maximum echo threshold is reached after about 9 or 10 s, whereas fast rates of 8/s or 16/s produce seemingly instantaneous buildup (Freyman *et al.*, 1991). This rapid buildup/breakdown is at the core of how the nervous system forms and discards models of acoustic space. Our data indicate that models are formed rapidly from brief exposure and are discarded with equal rapidity. There is little cost to rapid abandonment of a model because it is quickly reformed on subsequent exposure. Storing old models seems useless and inefficient and would only be valuable if acquiring new models were slow or difficult. It is the ultimate "throwaway" economy, perhaps a necessary one in light of the numerous acoustic spaces we experience every day.

And yet, the possibility of memory for past acoustic spaces is real and plausible. The usefulness of long-term memory for a familiar acoustic space is obvious for informing a listener about whether that space had changed, perhaps in some dangerous way. Although we know of no research that has tested listeners' sensitivity to changes in a familiar room, discrimination of different rooms' reverberant properties has been investigated. Robart and Rosenblum (2005) tested whether listeners could discriminate among four different rooms whose reflective properties varied widely. The same sounds were recorded in a gymnasium, a classroom, a rest room, and a small laboratory. Untrained listeners looked at photographs of the four rooms while listening to sound recorded in one of the rooms. Subjects were able to select the correct room on 78 of 100 trials with no feedback. Several studies have demonstrated humans' ability when blindfolded to detect objects by means of reflected sound (for reviews see Rice, 1967; Stoffregen and Pittenger, 1995). This prior work, as well as research on binaural room simulation as reviewed in Blauert (1997), supports our contention that listeners are highly sensitive to the structure of sound produced by numerous reflective surfaces in a room, and experience in different environments may generate memory for common acoustic spaces.

The idea that localization of sound in a new space is influenced by comparison of the current acoustic cues with the listener's acquired knowledge of spatial cues was proposed by Plenge (1974). In discussing how dummy-head recordings presented over earphones were first heard intracranially but after further experience were heard extracranially, Plenge (1974) hypothesized that subjects used cues from prior experience to localize sound. He suggested that knowledge of how sound behaves in spaces is acquired over a lifetime of experience and that we carry "stored stimulus patterns" for comparison to new stimuli. It appears that Plenge (1974) proposed a more generic form of stored knowledge than retention of specific room parameters; as he stated: "Only a short-term storage of such knowledge of sound sources and room conditions is useful" (p. 951). He more clearly described the process as follows: "The moment the hearer leaves the room, the stored information concerning its peculiarities is cleared, and information concerning the new situation is stored" (p. 951). A better description of the essence of our current data would be hard to find, although Plenge (1974) wrote this more than 3 decades ago.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Deafness and Other Communicative Disorders (Grant No. DC01625). The authors would like to thank Jamie Chevalier, Ackland Jones, Leah Novotny, Amanda LePine, and April Teehan for their assistance with the project. The authors gratefully acknowledge the helpful comments of D. Wesley Grantham, an anonymous reviewer, and Associate Editor Brian Moore.

ANSI. (**1996**). ANSI S3.6-1996, "Specifications for audiometers," American National Standards Institute, New York.

Bech, S. (**1995**). "Timbral aspects of reproduced sound in small rooms. I," J. Acoust. Soc. Am. **97**, 1717–1726.

Bech, S. (**1996**). "Timbral aspects of reproduced sound in small rooms. II," J. Acoust. Soc. Am. **99**, 3539–3549.

Bech, S. (**1998**). "Spatial aspects of reproduced sound in small rooms," J. Acoust. Soc. Am. **103**, 434–445.

Benade, A. (**1976**). *Fundamentals of Musical Acoustics* (Oxford University Press, London).

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Blauert, J., and Col, J. P. (**1992**). "A study of temporal effects in spatial hearing," in *Auditory Physiology and Perception*, edited by Y. Cazal, L. Demany, and K. Korner (Pergamon, Oxford), pp. 531–538.

Clifton, R. K. (**1987**). "Breakdown of echo suppression in the precedence effect," J. Acoust. Soc. Am. **82**, 1834–1835.

Clifton, R. K., and Freyman, R. L. (**1989**). "Effect of click rate and delay on breakdown of the precedence effect," Percept. Psychophys. **46**, 139–145.

Clifton, R. K., and Freyman, R. L. (**1997**). "The precedence effect: Beyond echo suppression," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. B. Anderson (Lawrence Erlbaum, Hillsdale, NJ), pp. 233–255.

Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**). "Listener expectations about echoes can raise or lower echo threshold," J. Acoust. Soc. Am. **95**, 1525–1533.

Clifton, R. K., Freyman, R. L., and Meo, J. (**2002**). "What echoes tell us about the auditory environment," Percept. Psychophys. **64**, 180–188.

Djelani, T., and Blauert, J. (**2000**). "Some new aspects of the buildup and breakdown of the precedence effect," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebart, A. J. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht, The Netherlands), pp. 200–207.

Djelani, T., and Blauert, J. (**2001**). "Investigations into the build-up and breakdown of the precedence effect," Acta Acust. Acust. **87**, 253–261.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). "Dynamic processes in the precedence effect," J. Acoust. Soc. Am. **90**, 874–884.

Freyman, R. L., and Keen, R. (**2006**). "Constructing and disrupting listeners' models of auditory space," J. Acoust. Soc. Am. **120**, 3957–3965.

Grantham, D. W. (**1996**). "Left-right asymmetry in the buildup of echo suppression in normal-hearing adults," J. Acoust. Soc. Am. **99**, 1118–1122.

Hartmann, W. M., and Rakerd, B. (**1989**). "Localization of sound in rooms IV: The Franssen effect," J. Acoust. Soc. Am. **86**, 1366–1373.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

McCall, D. D., Freyman, R. L., and Clifton, R. K. (**1998**). "Sudden changes in room acoustics influence the precedence effect," Percept. Psychophys. **60**, 593–601.

Plenge, G. (**1974**). "On the differences between localization and lateralization," J. Acoust. Soc. Am. **56**, 944–951.

Rakerd, B., and Hartmann, W. M. (**1985**). "Localization of sound in rooms, II: The effects of a single reflecting surface," J. Acoust. Soc. Am. **78**, 524–533.

Rice, C. E. (**1967**). "Human echo perception," Science **155**, 656–664.

Robart, R. L., and Rosenblum, L. D. (**2005**). "Hearing space: Identifying rooms by reflected sound," in *Studies in Perception and Action XIII*, edited by H. Heft and K. L. Marsh (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ).

Sanders, L. D., Joh, A. S., Keen, R. E., and Freyman, R. L. (**2008**). "One sound or two? Object-related negativity indexes echo perception," Percept. Psychophys. **70**, 1558–1570.

Stoffregen, T. A., and Pittenger, J. B. (**1995**). "Human echolocation as a basic form of perception and action," Ecological Psychol. **7**, 181–216.

Yang, X., and Grantham, D. W. (**1997**). "Echo suppression and discrimination suppression aspects of the precedence effect," Percept. Psychophys. **59**, 1108–1117.

Yost, W. A., and Guzman, S. (**1996**). "Auditory processing of sound sources: Is there an echo in here?" Curr. Dir. Psychol. Sci. **5**, 125–131.

Numerical study on source-distance dependency of head-related transfer functions

Makoto Otani[a)] and Tatsuya Hirahara
Faculty of Engineering, Toyama Prefectural University, Kurokawa 5180, Imizu, Toyama 939-0398, Japan

Shiro Ise
Graduate School of Engineering, Kyoto University, Katsura, Nishigyo-ku, Kyoto, Kyoto 615-8540, Japan

(Received 16 November 2007; revised 12 March 2009; accepted 12 March 2009)

This paper investigates the source-distance dependency of head-related transfer functions (HRTFs) on the horizontal and median sagittal planes using the boundary-element method and a dummy head scanned with laser and computer tomography scanners. First, the HRTF spectra are compared among various source positions in a head-centered coordinate system, confirming that the major HRTF spectral features vary with source distance as stated in previous works. Furthermore, the HRTF spectra are compared in an ear-centered coordinate system, revealing how the outer ear angle of incidence affects the source-distance dependency of the HRTFs. Next, the comparison across coordinate systems reveals that the source-distance dependency of the ipsilateral HRTFs on the horizontal plane is mainly attributable to the outer ear angle of incidence, whereas the contralateral HRTFs vary with the source distance mainly due to the head's presence. Finally, results also show that, in an ear-centered coordinate system, the ipsilateral HRTFs do not depend strongly on a source distance greater than 0.2 m from the center of the head, whereas the contralateral HRTFs depend on source distance less than 1.8 m. Results also show that HRTFs on the median sagittal plane depend on a source distance of less than 0.4 m.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3111860]

PACS number(s): 43.66.Qp, 43.66.Pn, 43.20.Fn [JCM] Pages: 3253–3261

I. INTRODUCTION

Head-related transfer functions (HRTFs), which contain naturally occurring spatial auditory cues of interaural time differences (ITDs), interaural level differences (ILDs), and spectral cues, require individualized measurements to allow realistic virtual auditory space simulations.[1]

Generally, HRTFs are measured in an anechoic chamber for a finite number of source positions, yielding HRTF databases.[2–6] Most HRTF measurements are performed for various source directions, but for only a single distance from the head, for example, 1 m, which limits HRTF-based virtual auditory displays. Assuming that HRTFs do not depend on the source distance, one approach for presenting a virtual sound source of various distances is to give distance-level decay to HRTFs, i.e., distance extrapolation. However, as Morimoto *et al.*[7] reported, a HRTF spectrum for horizontal sources within 1 m of the head shows considerable variation with source distance. Araki *et al.*[8] suggested a precise HRTF measurement technique for nearby sources using a spark source and reported that, compared to conventional measurements using loudspeakers, the resulting HRTFs enhance sound localization performance for nearby sources.

The research described above suggests that HRTFs of nearby sources depend on the source distance, thereby producing cues for source-distance perception and nearby source localization. With a rigid sphere analysis, Brungart *et al.*[9] showed that HRTFs within 1-m distance vary markedly with source distance, whereas HRTFs over 1-m distance are almost independent of source distance. In addition, ITDs are essentially independent of the source distance, whereas ILDs depend strongly on it. The ITDs depend on the distance difference, whereas ILDs depend on a distance ratio between both ears because changing the source distance to the head with a fixed incident angle to a center of the head yields a considerable change in distance ratio, but little change in the distance difference. Duda and Martens[10] confirmed these findings using HRTF measurements with a bowling ball. By measuring HRTFs on a Knowles Electronics Mannequin for Acoustic Research (KEMAR), Brungart *et al.*[9] reported that a nearer source yields a low-pass filtering effect; as the source distance decreases, an overall gain of ipsilateral HRTFs increases, whereas that of contralateral HRTFs decreases; the frequencies of spectral notches at higher frequencies, and the azimuthal angle at which the notches occur, vary with the source distance.[11] However, rigid sphere analysis does not take the outer ear into account, and the work of Brungart *et al.*[9] was conducted using coarse source-distance sampling (0.12, 0.25, 0.5, and 1.0 m).

This study investigates the source-distance dependency of the HRTFs in detail, including the effect of the outer ear, using numerically simulated HRTFs for point sources located at a 1-cm spatial resolution. Furthermore, this study discusses what causes the source-distance dependency of the HRTFs by clarifying the effect of the angle of incidence of the outer ear using head-centered and ear-centered coordinate systems.

[a)] Present address: Research Institute of Electrical Communication, Tohoku University. Electronic mail: otani@ais.riec.tohoku.ac.jp

II. NUMERICAL SIMULATION OF HRTFs

The authors developed a modification of the boundary-element method (BEM), which enables rapid HRTF computation for various positions[12,13] of the point and planar sources. The results obtained using this method are identical to those obtained using conventional BEM. Furthermore, the HRTFs can be simulated not only for various source positions but also for spherical and planar waves.[14] The HRTFs for planar waves are approximately identical to HRTFs for spherical waves originating from an infinite distance, and are therefore independent of the source distance. For this reason, the HRTFs for planar waves provide a benchmark when investigating the source-distance dependency of HRTFs for spherical waves.

HRTFs are calculable by applying numerical analysis to a wave equation whose boundary models a human head.

A. BEM-based HRTF calculation method

The numerical simulation method discussed here is based on the BEM, which discretizes a boundary of the wave equation. The most time-consuming process in the BEM is solving simultaneous equations and yielding surface pressures $\hat{\mathbf{P}}$. Based on the BEM formulation,[12] the simultaneous equations can be written as

$$\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)\hat{\mathbf{P}} = g. \tag{1}$$

In the conventional BEM, the sound pressure at a point $\mathbf{s}$ is calculated by solving the simultaneous equations in Eq. (1), giving

$$P(\mathbf{s}) = g_s - (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})\hat{\mathbf{P}}. \tag{2}$$

Assuming that the boundary condition and a receiver point are constant but that source position varies, an increase in the speed of the HRTF calculation can be attained as follows. As a pre-process, simultaneous equations written as

$$\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)^T\mathbf{Q}^T = (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})^T \tag{3}$$

are first solved to yield a vector $\mathbf{Q}$, which represents transfer functions from boundary elements to a receiver point,[13] where T represents a transposition. Next, as a post-process, substituting $\mathbf{Q}$ into the equation yields

$$P(\mathbf{s}) = g_s - \mathbf{Q}g, \tag{4}$$

where g_s and g denote a source-to-receiver transfer function and source-to-boundary-elements transfer function, respectively. This method enables rapid computation of the HRTFs for an arbitrary source position.

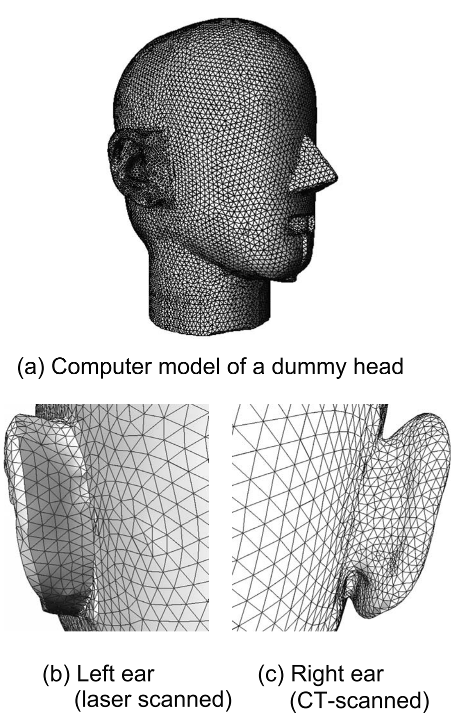
Furthermore, calculating g_s and g as

$$g_s = \frac{e^{-jk\mathbf{r}_s}}{4\pi\mathbf{r}_s} \quad \text{and} \quad g = \frac{e^{-jk\mathbf{r}}}{4\pi\mathbf{r}} \tag{5}$$

and

$$g_s = e^{-jk\mathbf{r}_s} \quad \text{and} \quad g = e^{-jk\mathbf{r}} \tag{6}$$

gives HRTFs for point and planar sources, respectively,[14] where j is the square root of -1 and $k(=\omega/c)$ is a wave number. In addition, $\mathbf{r}_s$ and $\mathbf{r}$ are source-to-receiver distance and source-to-boundary-elements distance, respectively. The HRTFs for point and planar sources located at an arbitrary position can be obtained through post-processing alone. Only if the boundary condition or receiver point change, pre-processing solution [Eq. (4)] be recomputed.


(a) Computer model of a dummy head

(b) Left ear (laser scanned)

(c) Right ear (CT-scanned)

FIG. 1. (a) Computer model of a dummy head (4128C; B&K) constructed from laser-scanned and micro-CT-scanned surface geometrical data. Both ear canals are blocked at their entrances. Back of (b) laser=scanned ear (left ear) and (c) CT-scanned ear (right ear). The back of the right pinna is modeled accurately, whereas that of the left pinna is not modeled accurately.

The HRTF calculation with the BEM is categorized as an exterior problem, which handles a scattering sound field around an object (the head in this case). The BEM for an exterior problem produces inaccurate solutions at eigenfrequencies of a corresponding interior problem (solution inside the object) for which Eq. (1) has no unique solution. To resolve this so-called "non-uniqueness problem," a combined Helmholtz integral equation formulation (CHIEF) method[15] is applied. In the CHIEF method, equations satisfying the constraint that pressures at arbitrary points inside the boundary equal to zero are added to the usual BEM formulation. The resulting over-determined linear system has a least-squares solution, enhancing the solution accuracy at eigenfrequencies that are underdetermined otherwise.

B. Head model

The HRTFs of a dummy head (4128C; Brüel & Kjær) are simulated without a torso. A head model [Fig. 1(a)] was constructed from laser-scanned and micro-computer-tomography (CT)-scanned surface geometrical data.[12] Both ear canals are blocked at their entrances. Figure 1(b) depicts a laser-scanned outer ear (left ear). The back of the pinna and its fine details (e.g., concha detail) are not modeled accurately because the laser scanner was unable to measure shadowed parts that the laser did not reach. Consequently, the

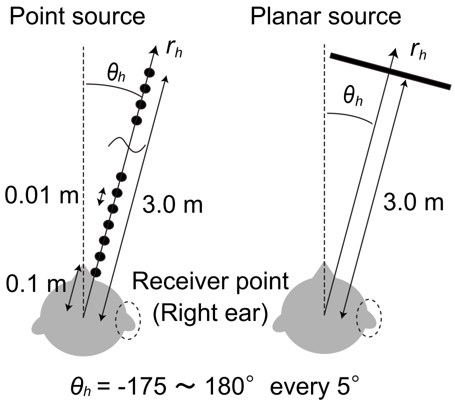


FIG. 2. Numerical configuration of point and planar sources and receiver point in head-centered coordinate system (θ_h, r_h). The HRTFs are calculated for the right ear. Point sources are located in azimuthal direction (θ_h) between −175° and 180° at increments of 5° and at distance from origin (r_h) between 0.1 and 3.0 m at increments of 0.01 m.

right ear was replaced with a much more accurately modeled ear scanned using a micro-CT scanner [Fig. 1(c)].[13]

The measurement resolutions of the laser and micro-CT scanners are approximately 1 and 0.1 mm, respectively. However, the measured surface geometrical data were decimated when finding the numerical solutions. The head model comprises 28 000 triangular elements whose maximum length is 5.64 mm. This element length corresponds to 1/4 of the wavelength of 15 kHz and 1/3 of the wavelength of 20 kHz.

C. Numerical condition

HRTFs for spherical and planar waves originating from a sound source located on the horizontal plane were investigated. The head surface is assumed to be acoustically rigid. Transfer functions are calculated up to 20 kHz at intervals of 86 Hz. Figure 2 depicts the position of the point and planar sources. The origin is set to the midpoint between the ears. Here, θ_h represents an azimuthal direction, in degrees. Furthermore, r_h represents a distance from the origin, in meters. The right ear HRTFs are calculated for point sources located at θ_h between −175° and 180° at increments of 5°, and r_h between 0.1 and 3.0 m at increments of 0.01 m. If $\theta_h=60°$, then $r_h=0.1$ m corresponds to a point approximately 1 cm away from the surface of the dummy head. The HRTFs for a planar source are independent of the source distance; therefore, only the azimuthal direction varies in planar source solutions. The receiver is located adjacent to the ear canal entrance. All calculated HRTFs are normalized using transfer functions at the origin in the absence of the head. These transfer functions can be calculated using Eqs. (5) and (6), respectively.

The rigid sphere HRTFs are also boundary-element simulated. The diameter of the rigid sphere is equal to the interaural distance of the head model, and receivers are equal to those in the head model computation.

III. RESULTS

A. HRTF spectra variation due to the source distance

1. Overall features

Figure 3(a) shows HRTFs for point sources (spherical HRTF) located at $\theta_h=60°$ and $r_h=0.13$, 0.25, 0.5, 1.0, and 3.0 m, and HRTFs for planar sources (planar HRTF) located at the same azimuthal direction. Transfer functions up to 20 kHz are shown. The maximum element length of the head model used (5.64 mm) corresponds to 1/3 wavelength of 20 kHz. It is generally believed that discretization of four or five elements per wavelength yields accurate numerical results. However, discretization of 1/3 wavelength does not markedly reduce the numerical accuracy because it is reported that the maximum difference between 1/3 and 1/5 discretizations is less than 1 dB.[16] Consequently, this paper demonstrates numerical results up to 20 kHz.

As the source distance increases, the magnitude spectra of spherical HRTFs become closer to those of planar HRTFs. As the source distance decreases, the spherical HRTF mag-

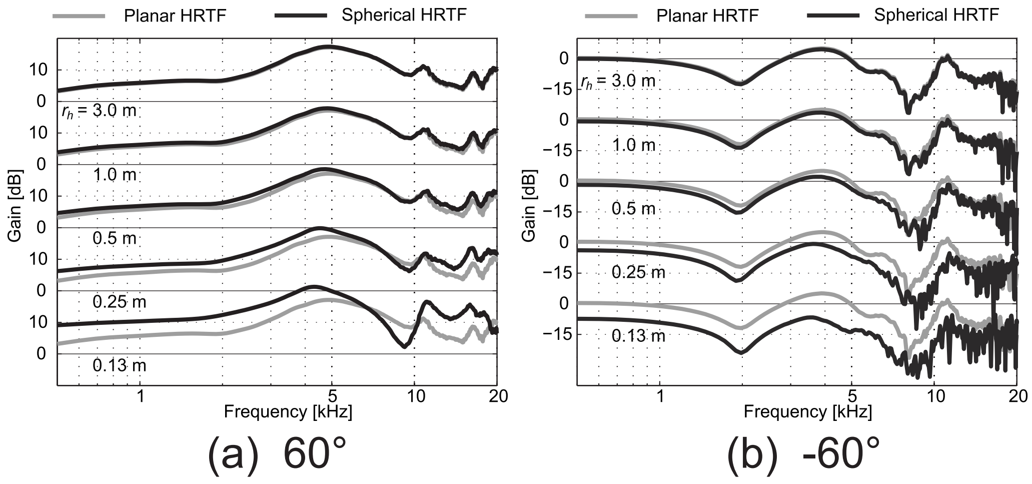


FIG. 3. Planar and spherical HRTFs up to 20 kHz for (a) $\theta_h=60°$ and (b) $\theta_h=-60°$, and $r_h=0.13$, 0.25, 1.0, and 3.0 m. Gray lines represent planar HRTFs, which are identical for all distances. Black lines represent spherical HRTFs. (a) As the source distance increases, the spectra of spherical HRTFs become closer to those of planar HRTFs for $\theta_h=60°$. As the source distance decreases, the spectral magnitude of the spherical HRTFs increases for frequencies below 6 kHz. Above 6 kHz, the amplitude and frequency of spectral peaks and notches vary. (b) As the source distance decreases, the magnitude spectra of the spherical HRTFs decrease at all frequencies for $\theta_h=-60°$. In addition, the frequencies of notches around 2 and 7–8 kHz vary with source distance.

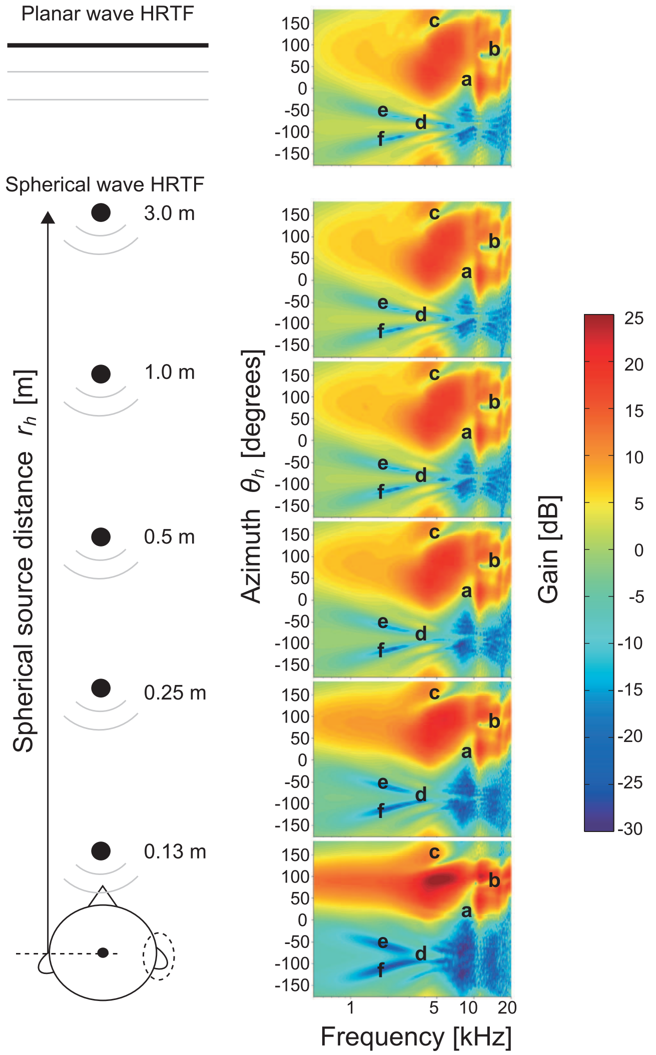


FIG. 4. Planar HRTFs and spherical HRTFs for various source distances (r_h=0.13, 0.25, 0.5, 1.0, and 3.0 m). Abscissa represents frequency and ordinate represents azimuthal direction. Colors represent gain in decibels. As illustrated in the color bar next to the figure, red colors indicate higher gain and blue colors indicate lower gain. Labels (**a**)–(**f**) in figures are reference points where HRTF spectra show noticeable features. See text for details.

nitude increases for frequencies below 6 kHz. At frequencies greater than 6 kHz, the amplitude and frequency of spectral peaks and notches in the magnitude spectra vary.

Figure 3(b) presents the results for θ_h=−60°. As the source distance decreases, the spherical HRTF magnitude spectra decrease at all frequencies. Furthermore, the frequencies of notches around 2 and 7–8 kHz vary with source distance. The dependency of these features of the spherical-wave HRTF magnitude spectra on source distance has previously been observed in measured HRTFs from KEMAR.[11]

2. Variation in spectral peaks and notches

Figure 4 shows spherical HRTFs for r_h=0.13, 0.25, 0.5, 1.0, and 3.0 m, as well as planar HRTFs. The abscissa represents frequency and the ordinate represents the azimuthal direction. The labels (**a**–**f**) in the panels are reference points for which the HRTF spectra show noticeable features. These reference points are plotted at the same frequency and azimuthal direction in all panels. Figure 4 shows that spectral peaks and notches markedly vary their frequencies, heights/depths, and azimuthal directions with source distance. For example, reference point **a** (θ_h=20°, 9 kHz) marks a spectral notch in the planar wave HRTF. As source distance decreases, the notch width increases, and the notch becomes apparent for a larger extent of source azimuths (e.g., from 20° to 60° at the 0.13-m distance).

Figure 5 portrays the HRTFs of a rigid sphere head every 30° in the azimuthal direction. The abscissa represents the frequency and the ordinate represents the point source distance. The angles shown around the sphere denote azimuthal direction θ_h. This figure clearly highlights that, for θ_h from 60° to 120°, the magnitude spectra of HRTFs from a rigid sphere increase at lower frequencies as the source distance decreases; for θ_h from −150° to −30°, the magnitude spectra decrease at all frequencies. Furthermore, spectral notches, which appear for θ_h from −150° to −30°, vary their frequencies with source distance [Fig. 3(b)]. The figure also shows that, for other azimuthal angles, HRTFs of the rigid sphere do not vary markedly with source distance.

In the same manner as Fig. 5, Fig. 6 portrays the spherical HRTFs of the full head model, including pinnae. The labels written at the top of each figure correspond to those portrayed in Fig. 4. The gray shaded area at smaller distance indicates HRTFs that were not calculated because the point source would have been located inside the head (θ_h=0°, r_h $\leq$0.12 m). Firstly, the frequencies of spectral notches in the ipsilateral HRTFs (θ_h=0°−150°), such as those at reference points *a*, *a*′, and *c*, vary with source distance. Generally, notch depths deepen and peak heights increase with decreasing source distance, yielding HRTF spectra with markedly emphasized peaks and notches. In contrast, the ipsilateral HRTFs of the rigid sphere have no such notches at higher frequencies (see Fig. 5), indicating that these notches are caused by the pinna. A spectral notch at reference point **c** demonstrates a typical example of a notch frequency that varies with source distance. The notch at **c** appears at 5 kHz for r_h=3.0 m, then moves to 6–7 kHz for r_h=0.1 m. This phenomenon arises from a source-distance dependency of the notch on the azimuth relative to the ear, θ_e, rather than the head, θ_h (see further analysis in Sec. III A 3). For r_h =3.0 m, θ_h=150° corresponds to approximately θ_e=151°, then for r_h=0.1 m, the same θ_h corresponds to θ_e equaling approximately 193°, generating a 42° space difference in θ_e. In contrast, the frequency of the spectral notch at around 19 kHz for the same θ_h does not vary for $r_h$$\geq$0.2 m. Secondly, a spectral notch in the HRTFs for θ_h=90°, which is at reference point **b**, appears at essentially the same frequency across all source distances. This can be understood by noting that the incident angle to the outer ear (θ_e) is constant, irrespective of the source distance, because sources are located exactly to the right. Finally, spectral notches in the contralateral HRTFs (θ_h from −30° to −150°), such as those at **d**–**f**, vary their depths and frequencies markedly with source distance.

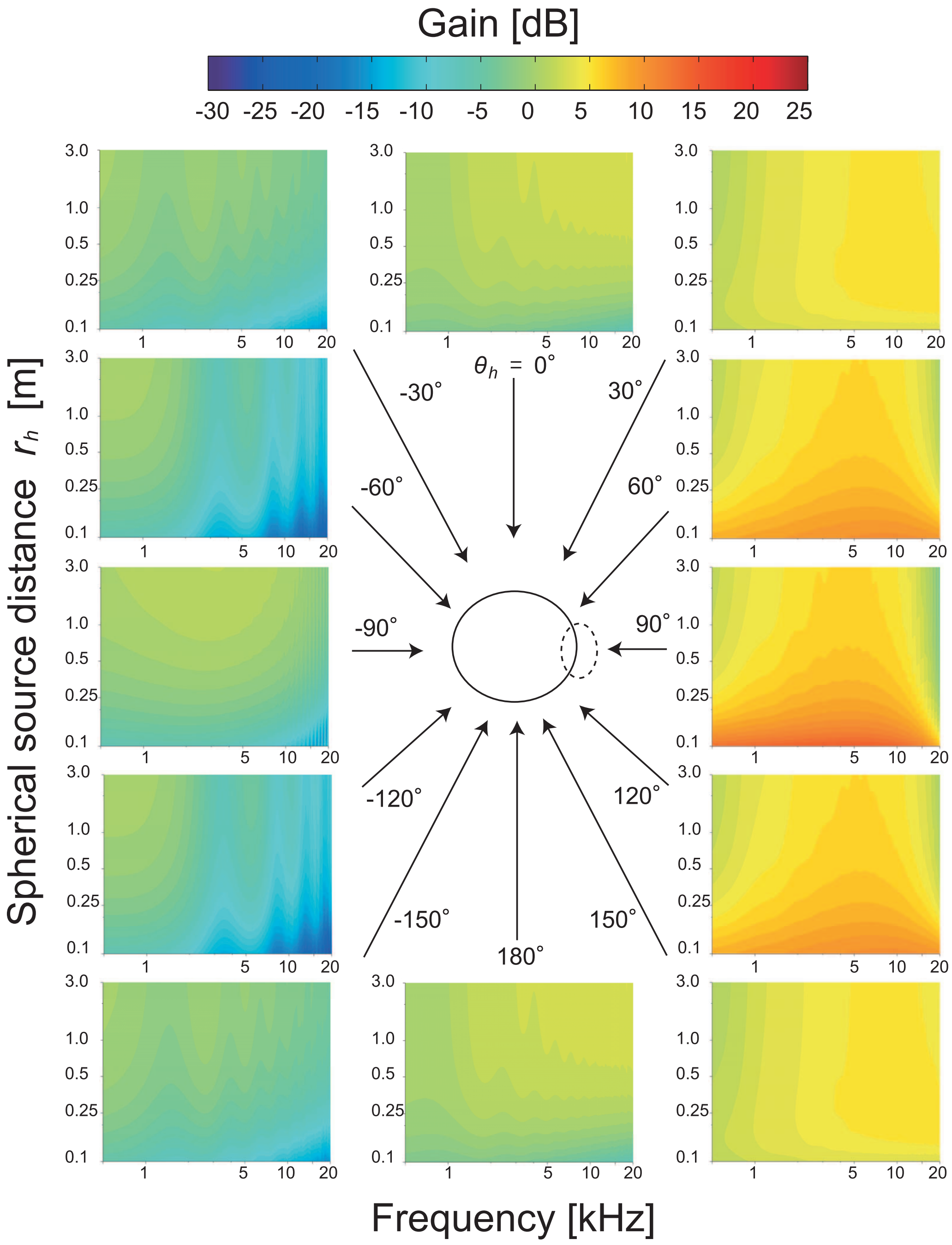


FIG. 5. HRTFs of a rigid sphere every 30° azimuthal direction as function of source distance in a head-centered coordinate system (θ_h, r_h). Abscissa represents frequency and ordinate represents point source distance. Colors present same manner as Fig. 4. Angles denote azimuthal direction θ_h relative to the median plane of the head. No higher-frequency spectral notches were observed.

Such spectral notches are also observed in rigid sphere head results (Fig. 5). However, the spectral notches in rigid sphere head results appear at different frequencies with different depths from those in full head model results (Fig. 6). Furthermore, the frequencies of the spectral notches in rigid sphere head results do not depend on the source distance, although the depths of these spectral notches deepen as the source distance decreases. This comparison indicates that, in the contralateral HRTFs, complicated effects of pinnae and a head yield the spectral notches as well as variations in the frequencies and depths of the spectral notches with source distance.

Figure 5 demonstrates that the rigid sphere head also yields the spectral notches in the contralateral HRTFs. However, there are prominent discrepancies between the rigid sphere head and full head model results; the spectral notches in rigid sphere head appear at different frequencies from those in full head model results; especially, rigid head model results do not demonstrate spectral notches at **d**–**f**. This comparison indicates that the pinnae yield the deep spectral notches at **d**–**f**, or shift the frequencies of the shallow spectral notches produced by the head. No matter which the origin of the spectral notches at **d**–**f** is, the presence of pinnae has effects on the frequencies of the spectral notches in the contralateral HRTFs. (See further discussion in Sec. III A 3.)

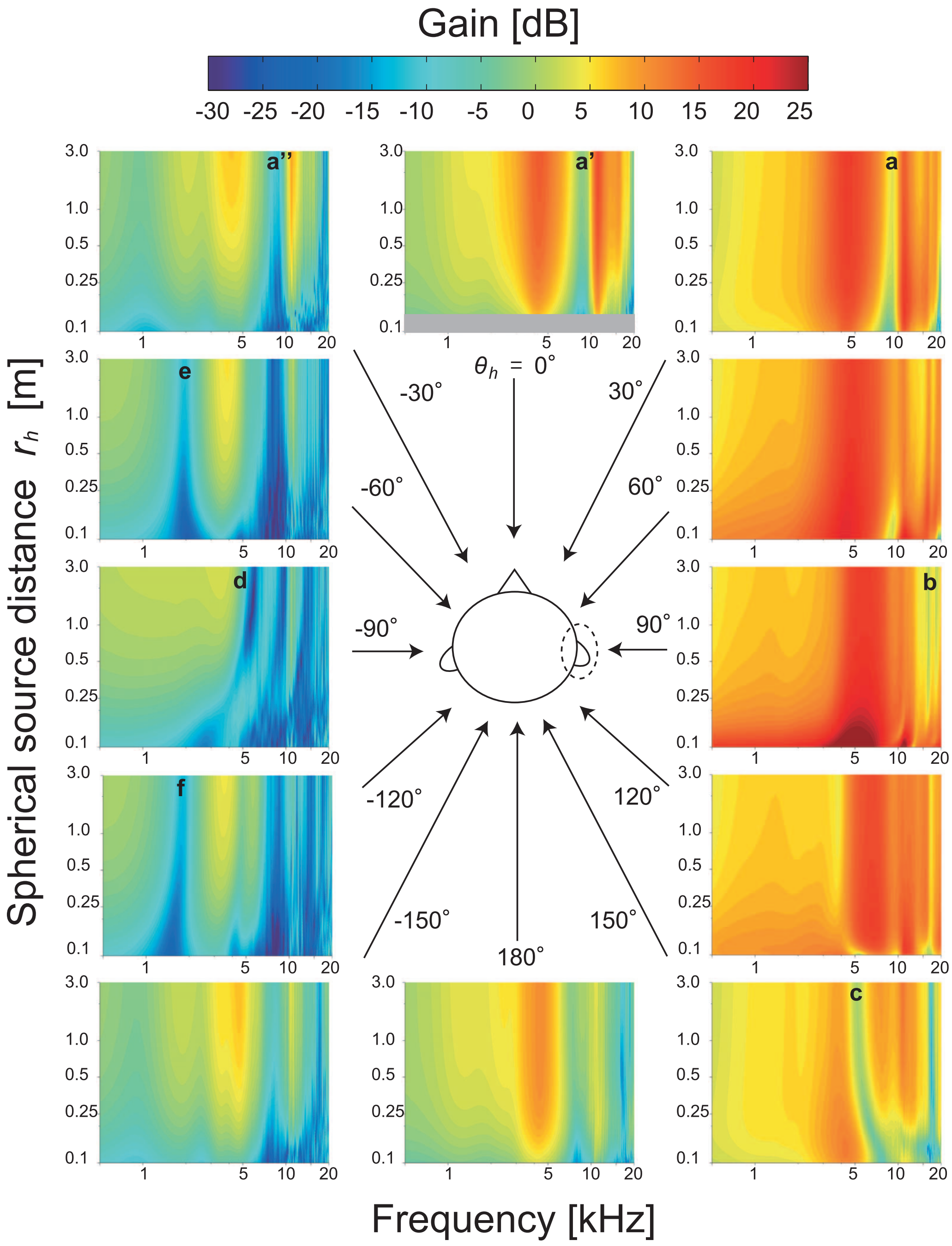


FIG. 6. HRTFs of a head model with pinnae every 30° azimuthal direction as function of source distance in a head-centered coordinate system (θ_h, r_h). Abscissa represents frequency and ordinate represents point source distance. Red and blue colors indicate higher and lower gains, respectively. Labels written at the top of each figure correspond to those in Fig. 4. Gray shaded areas at smaller distance indicate HRTFs that could not be calculated because the point source would have been located inside head (θ_h=0°, $r_h \leqslant 0.12$ m). See text for details.

3. Effect of incident angle to the outer ear

The source-distance dependency of the HRTFs, especially in the ipsilateral HRTFs, is thought to be attributable to variations in θ_e. To investigate the source-distance dependency of the HRTFs further, an ear-centered coordinate system is used, in which θ_e does not vary with the source distance. The numerical configuration is the same as that for a head-centered coordinate system shown in Fig. 2, except that the origin is set to the receiver point, i.e., the entrance of the ear canal. The distance from a point source to the entrance of ear canal is between 0.01 and 3.0 m at increments of 1 cm. The planar HRTFs in this coordinate system are identical to those in the head-centered coordinate system.

Figure 7 shows the spherical HRTFs calculated in the ear-centered coordinate system. The gray shaded areas indicate HRTFs that cannot be calculated, because they fall inside the head (note that these areas are much greater than the corresponding areas in Fig. 6). Unlike Fig. 6, the ordinate (r_e) ranges from 0.01 to 3.0 m. When r_e=3.0 m, the spherical HRTFs for both coordinate systems (Figs. 6 and 7) are almost identical to the planar HRTFs, indicating that these two coordinate systems yield no remarkable differences in

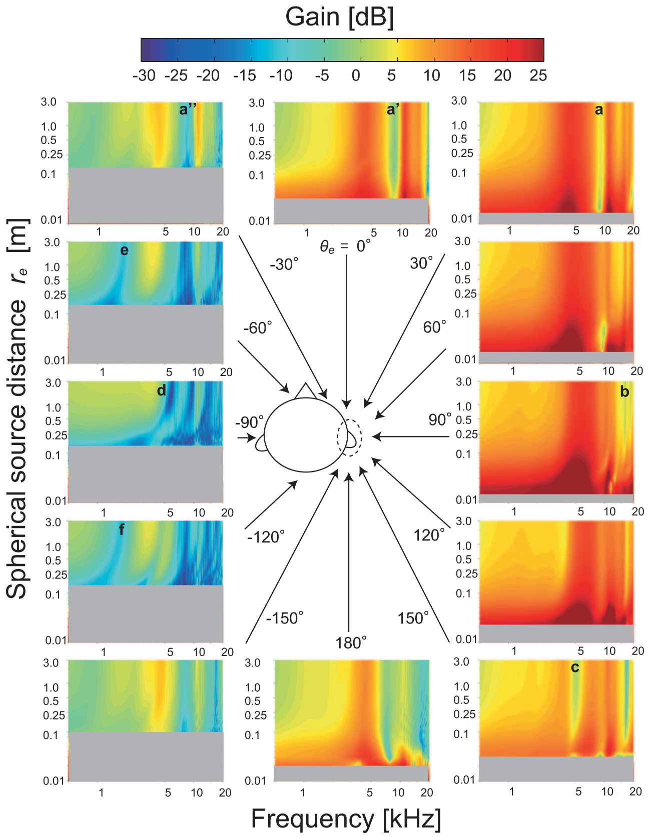


FIG. 7. HRTFs of a head model every 30° azimuthal direction as function of source distance in an ear-centered coordinate system (θ_e, r_e). Red and blue colors indicate higher and lower gains, respectively. Gray shaded area represents HRTFs that could not be calculated; this area is much larger than corresponding area in Fig. 6. Unlike Fig. 6, ordinate (r_e) is 0.01–3.0 m. See text for details.

this case. Firstly, the frequencies of spectral notches in the ipsilateral HRTFs $(\theta_e = 0°)$, such as those at reference points **a**, 9 kHz, and **c**, 5 kHz, vary less with source distance compared with the head-centered coordinate system results shown in Fig. 6. This comparison reveals that the incident angle to the outer ear explains the notch frequency variations that arise in the head-centered coordinate analysis. A spectral notch at reference point **a** moves slightly for $r_e \leqslant 0.1$ m, likely due to a source-distance dependency of interactions with the outer ear, especially given that the outer ear's frequency response varies with source distances within 4 cm of the outer ear.[17] Secondly, spectral notches in the HRTFs for $\theta_e = 90°$ are almost identical to those shown in Fig. 6 (note that the ordinate values differ between these figures). Thirdly, spectral notches in the contralateral HRTFs $(\theta_e = -60°$ and $-120°)$, such as those at **e** and **f**, move from 2 to 1 kHz as the source distance decreases; however, when analyzed in head-centered coordinate system, the spectral notches shift much less. This indicates that the spectral notches at **e** and **f** depend primarily on the head's presence. However, considering that the spectral notches **e** and **f** are produced because of the pinnae, or be influenced by the pinnae, as analyzed in Sec. III A 2, the frequencies of the spectral notches, such as those at **e** and **f**, depend on the incident angle to both the head and the outer ear, indicating a complex source-distance dependency of the contralateral HRTFs. On the other hand, the result for $\theta_e = -90°$ is similar to that shown in Fig. 6 because the sources are located on the interaural axis, and $\theta_e = \theta_h$.

For $\theta_e = 150°$, spectral notches appear at 4, 8, and 11–20 kHz for $\theta_e = 150°$ and $r_e = 0.03$ m. A source location of $\theta_e = 150°$ and $r_e = 0.03$ m corresponds to locations behind the pinna. The HRTF for $\theta_e = 180°$ has a similar spectral notch. Thus, these notches appear to arise when a source is located behind the pinna.

As described above, spectral notches at higher frequencies of the ipsilateral HRTF spectra (θ_e from 0° to 180°), which are produced because of the pinna, do not depend strongly on source distance for distances of tens of centimeters or more from the outer ear. This indicates that a source-distance dependency, in a head-centered coordinate system, of the frequency of high-frequency spectral notches in the ipsilateral HRTFs is attributable to change in the incident angle to the outer ear (θ_e). In contrast, the impact of the outer ear on the source-distance dependency of the HRTFs increases as the source becomes closer to the ears, generating spectral notches that are not observable for more distant sources.

4. Median sagittal plane

HRTFs for sources located on the median plane are also investigated. In this work, effects of the body are not incorporated because the numerical head model does not include the body, although the presence of the body is important for elevation angles between −60° and −120°. For a frontal source, HRTFs are identical to those for $\theta_h = 0°$ shown in Fig. 6, and spectral notch variations are also identical to those for $\theta_h = 0°$. Furthermore, on the median sagittal plane, variations in θ_e with source distance are almost the same as for a frontal source across all elevation angles. As a result, there is a similar source-distance dependency across all elevation angles, even though spectral notches have different frequencies and depths for different elevation angles.

B. Objective evaluation of the spectrum variation

Spectral distances (SDs) between the spherical HRTFs and the reference planar HRTFs are computed in both coordinate systems. The SDs are calculated as follows:

$$\mathrm{SD} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(20\log_{10}\frac{H_{\mathrm{spherical}}(r,\theta,\omega_k)}{H_{\mathrm{planar}}(\theta,\omega_k)}\right)^2},$$

where K is the number of calculated frequencies, ω_k is the angular frequency, $H_{\mathrm{spherical}}(r,\theta,\omega_k)$ is a HRTF for a point source located at a distance of r and an azimuthal angle of θ, and $H_{\mathrm{planar}}(\theta,\omega)$ is a HRTF for a planar source located at an azimuthal angle of θ. In addition, $\theta = \theta_h$ and $r = r_h$ for a head-centered coordinate system, whereas $\theta = \theta_e$ and $r = r_e$ for an ear-centered one. The frequencies are from 86 Hz to 20 kHz at an interval of 86 Hz. Therefore, $K = 232$.

Results for the coordinate systems are shown in Figs. 8(a) and 8(b), respectively. The abscissa represents the azimuthal angle, and the ordinate represents the distance from each origin point (i.e., the center of the head or the entrance

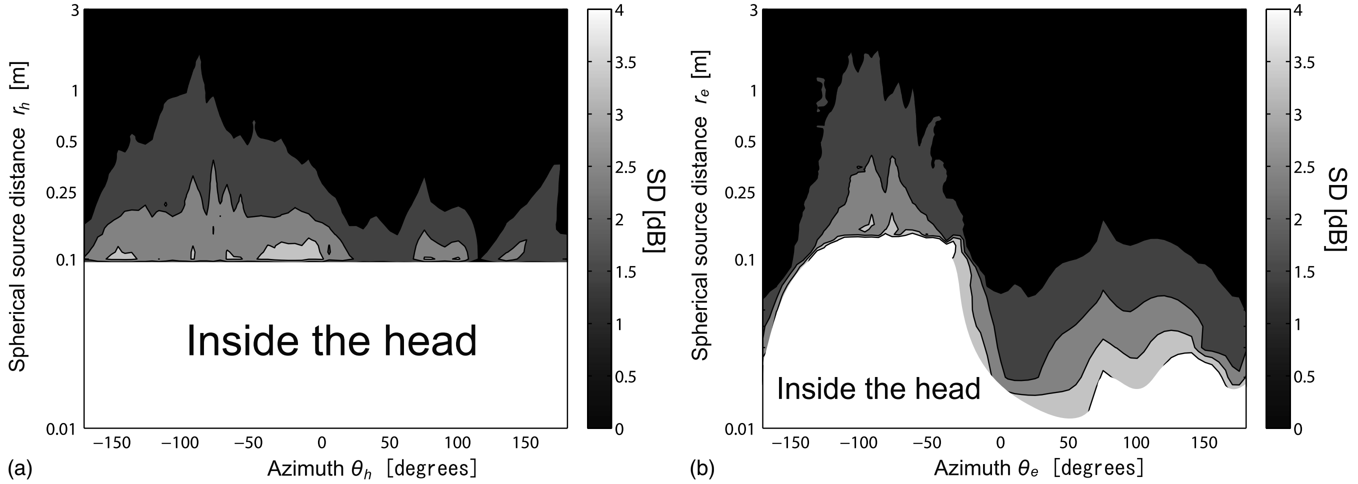


FIG. 8. SD between spherical and planar HRTFs on the horizontal plane in (a) head-centered (θ_h, r_h) and (b) ear-centered (θ_e, r_e) coordinate systems. Abscissa represents azimuthal angle; ordinate represents distance from each origin point, which is either the center of the head or the entrance to the ear canal. Colors closer to black indicate smaller SD values; black represents SDs less than 1 dB. White areas at smaller distances indicate SDs that were not calculated because the source would have been located inside head. See text for details.

to the ear canal). Colors closer to black represent smaller SD values, and the black area represents a SD of less than 1 dB. The white areas at smaller distance indicate a SD that was not calculated because the source would have been located inside the head.

In the head-centered coordinate system, as shown in Fig. 8(a), SDs are larger for contralateral HRTFs than for the ipsilateral HRTFs, revealing that the contralateral HRTFs depend primarily on the source distance (the source-distance dependency of the head has a large effect since the head is located between the source and the receiver ear). In contrast, SDs are generally smaller for the ipsilateral HRTFs because the head's presence has a smaller effect. For the ipsilateral HRTFs, the largest distance when SDs are less than 1 dB is r_h=0.38 m at θ_h=170°. For the contralateral HRTFs, the corresponding distance is r_h=1.69 m at θ_h=−90°.

In the ear-centered coordinate system [Fig. 8(b)], the corresponding distance above which distance effects are small is r_e=1.79 m at θ_e=−110° for the contralateral HRTFs, indicating that the contralateral HRTFs show a considerable dependency on source distance. In contrast, the corresponding distance for the ipsilateral HRTFs is r_e=0.18 m at θ_e =75°. In fact, r_e=0.18 m at θ_e=75° corresponds to a source distance of about 0.25 m in the head-centered coordinate system. Therefore, according to these SD evaluations, the ipsilateral HRTFs depend less on the source distance in the head-centered coordinate system than in the ear-centered one. From a perceptual viewpoint, the contralateral HRTFs are less important than the ipsilateral HRTFs if ITDs and the ipsilateral HRTFs are accurate. Consequently, the source-distance dependency of the contralateral HRTFs would be much smaller in perceptual meaning.

Assuming that a source-distance dependency of the HRTFs is essentially imperceptible when SDs to the planar HRTFs are less than 1 dB, source distance has negligible effects on the contralateral HRTFs for sources located further than 1.7 m in the head-centered coordinate system and 1.8 m in the ear-centered one. For the ipsilateral HRTFs, the criterion decreases to approximately 0.4 m in the head-centered coordinate system and approximately 0.2 m in the ear-centered coordinate system.

Additionally, the corresponding results for the median sagittal plane show that the relation between SD and source distance does not depend greatly on the elevation angle compared to the dependence on an angle in the horizontal plane, although rear sources yield smaller SDs, which indicates a weak source-distance dependency of the rear HRTF. In this case, the largest distance when SDs are less than 1 dB is 0.44 m at elevation angle ϕ=65°. These results indicate that a source distance has negligible impact on the HRTFs on the median sagittal plane for sources located further than 0.4 m in the head-centered coordinate system. However, for a nearby source located less than approximately 0.4 m from the center of the head, the source distance has a considerable effect on the HRTFs.

IV. CONCLUSION

This study investigated the source-distance dependencies of the HRTFs using fine spatial resolution with numerically simulated HRTFs.

Simulation in a head-centered coordinate system confirmed results of previous works using a finer spatial resolution than had been used in previous studies.[9–11] Namely, the HRTF spectra vary markedly with decreasing source distance. As the source distance decreases, the magnitude spectra increase at lower frequencies for ipsilateral HRTFs and decrease at all frequencies for contralateral HRTFs. Spectral notches vary in their depths, widths, and frequencies markedly with source distance.

Simulation in an ear-centered coordinate system clarified that ipsilateral HRTF spectra for sources on the horizontal plane do not vary markedly with decreasing source distance compared to in a head-centered coordinate system. Higher-frequency spectral notches, which are caused by the pinna, do not vary markedly for sources located further than several tens of centimeters from the outer ear, and show a weak source-distance dependency. The source-distance dependency of the pinna is considerable for a source adjacent to the pinna, producing spectral notches that were not observed in the head-centered simulation.

Moreover, a comparison across both coordinate systems reveals that the source-distance dependency of the higher-frequency spectral notches in the frontal and ipsilateral HRTFs might be attributable to the changing angle of incidence with respect to the outer ear in the head-centered coordinate system. On the other hand, the source-distance dependency of the contralateral HRTFs is attributed to a combination of the presence of the head and the outer ears, showing complexities compared to the ipsilateral HRTFs.

On the median sagittal plane, HRTF spectra also depend on source distances for nearby sources in a manner similar to the spectral variation for the frontal source in the horizontal plane.

SDs from the reference planar HRTFs were calculated to objectively evaluate the source-distance dependency of the spherical-wave HRTFs. For sources on the horizontal plane, results show that SDs are less than 1 dB for a point source located further than 1.7 and 1.8 m in the head- and ear-centered coordinate systems, respectively. For ipsilateral HRTFs, they are, respectively, 0.4 and 0.2 m. For sources on the median sagittal plane, they are about 0.4 m for all elevation angles.

ACKNOWLEDGMENTS

The computer model of the head used in this work was developed in conjunction with the National Institute of Advanced Industrial Science and Technology (AIST) and Matsushita Electric Industrial Co., Ltd. at Kyoto University. The authors would like to thank them for their cooperation. This research was partly supported by a Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan: Grant-in-Aid for Scientific Research (B), 2006 (Grant No. 18300042).

[1] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recording?," J. Audio Eng. Soc. **44**, 451–469 (1996).

[2] W. Gardner and K. Martin, "HRTF measurement of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).

[3] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, Mohonk Mountain House, New Paltz, NY, 21–24 October (2001).

[4] S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki, and T. Sone, "A database of head-related transfer functions in whole directions on upper hemisphere," Acoust. Sci. & Tech. **23**, 160–162 (2002).

[5] A. Kudo, H. Hokari, and S. Shimada, "A study of switching of the transfer functions focusing on sound quality," Acoust. Sci. & Tech. **26**, 267–278 (2005).

[6] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolation of the head related transfer function on the horizontal plane," J. Acoust. Soc. Jpn. **55**, 91–99 (1999).

[7] M. Morimoto, N. Joren, Y. Ando, and Z. Maekawa, "On head-related transfer function," Transactions on Technical Committee of Psychological and Physiological Acoustics, Acoustical Society of Japan, **H-31-1-2**, 4–9 (1976).

[8] J. Araki, T. Nishino, K. Takeda, and F. Itakura, "Measurement of the head related transfer function using the spark noise," J. Acoust. Soc. Jpn. **60**, 314–318 (2004).

[9] D. S. Brungart, W. M. Rabinowitz, and N. Durlach, "Auditory localization of a nearby point source," J. Acoust. Soc. Am. **100**, 2593–2593 (1996).

[10] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **105**, 3048–3058 (1998).

[11] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479 (1999).

[12] M. Otani and S. Ise, "A fast calculation method of the head-related transfer functions for multiple source points based on the boundary element method," Acoust. Sci. & Tech. **24**, 259–266 (2003).

[13] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," J. Acoust. Soc. Am. **119**, 2589–2598 (2006).

[14] M. Otani, T. Hirahara, and S. Ise, "A fast calculation method of HRTF for line and plane wave source," Proceedings of the Autumn Meeting of Acoustic Society of Japan (2006), pp. 475–476.

[15] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).

[16] Architectural Institute of Japan, *Prediction of Sound Field in Rooms: Theory, Application and Recent Development* (Maruzen, Tokyo, 2001), pp. 83–84 (in Japanese).

[17] D. S. Brungart, "Near-field auditory localization," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts (1996).

Results of the National Institute for Occupational Safety and Health—U.S. Environmental Protection Agency Interlaboratory Comparison of American National Standards Institute S12.6-1997 Methods A and B

William J. Murphy[a] and David C. Byrne
Hearing Loss Prevention Team, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS C-27, Cincinnati, Ohio 45226-1998

Dan Gauger
Bose Corporation, MS271E, 145 Pennsylvania Avenue, Framingham, Massachusetts 01701-9168

William A. Ahroon
U.S. Army Aeromedical Research Laboratory, 6901 Andrews Avenue, P.O. Box 620577, Fort Rucker, Alabama 36362-0577

Elliott Berger
Aearo Technologies, 7911 Zionsville Road, Indianapolis, Indiana 46268-1657

Samir N. Y. Gerges
Departamento de Engenharia Mecânica, Laboratório de Vibrações e Acústica, (LARI and LAEPI), Universidade Federal de Santa Catarina, Cx. P. 476, CEP 88040-900, Florianópolis, SC, Brazil

Richard McKinley
Human Effectiveness Directorate, AFRL/HECB, 2610 7th Street, Building 441, Dayton, Ohio 45433

Brad Witt
Howard Leight Industries, 7828 Waterville Road, San Diego, California 92154

Edward F. Krieg
Division of Applied Research and Technology, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS C-22, Cincinnati, Ohio 45226

(Received 22 September 2008; revised 8 January 2009; accepted 9 February 2009)

The National Institute for Occupational Safety and Health and the Environmental Protection Agency sponsored the completion of an interlaboratory study to compare two fitting protocols specified by ANSI S12.6-1997 (R2002) [(2002). American National Standard Methods for the Measuring Real-Ear Attenuation of Hearing Protectors, American National Standards Institute, New York]. Six hearing protection devices (two earmuffs, foam, premolded, custom-molded earplugs, and canal-caps) were tested in six laboratories using the experimenter-supervised, Method A, and (naïve) subject-fit, Method B, protocols with 24 subjects per laboratory. Within-subject, between-subject, and between-laboratory standard deviations were determined for individual frequencies and *A*-weighted attenuations. The differences for the within-subject standard deviations were not statistically significant between Methods A and B. Using between-subject standard deviations from Method A, 3–12 subjects would be required to identify 6-dB differences between attenuation distributions. Whereas using between-subject standard deviations from Method B, 5–19 subjects would be required to identify 6-dB differences in attenuation distributions of a product tested within the same laboratory. However, the between-laboratory standard deviations for Method B were −0.1 to 3.0 dB less than the Method A results. These differences resulted in considerably more subjects being required to identify statistically significant differences between laboratories for Method A (12–132 subjects) than for Method B (9–28 subjects).

© 2009 *Acoustical Society of America.* [DOI: 10.1121/1.3095803]

PACS number(s): 43.66.Vt, 43.50.Hg, 43.66.Yw [BSF] Pages: 3262–3277

I. INTRODUCTION

In March 2003, the United States Environmental Protection Agency (EPA) conducted a workshop to bring together parties involved in the manufacture, sale, testing, and use of hearing protection devices (HPDs). These parties included government, academia, manufacturers, and testing laboratories. During this two-day workshop, possible changes in the EPA's regulation for hearing protector labeling, 40 CFR 211 subpart B (EPA, 1978), were discussed during lecture and facilitated discussions. One of the more controversial discussion topics was the choice of testing methods for measuring

[a] Electronic mail: wmurphy@cdc.gov

the attenuation of HPDs among those specified in the American National Standard Method for Measuring the Real-Ear Attenuation of Hearing Protectors (ANSI S12.6-1997 (R2002), 2002).

The two methods under consideration were an experimenter-supervised fit (Method A) and a naïve subject-fit (Method B). Method A allows the use of subjects who have received training or have experience fitting HPDs. However, Method B requires subjects who have not received previous one-on-one training and have limited experience with protector testing, with wearing HPDs, and with computer-based or video training in the correct fitting of HPDs. The workshop participants identified the need for a direct comparison of the two methods as a priority before revising the regulation. For insert earplugs, the role of the experimenter has proven to be a significant factor in the amount of attenuation achieved by a particular laboratory. According to Berger *et al.* (1998), attenuations from naïve subjects were representative of the upper quartile of the real-world attenuation measurements. Under Method B, naïve subjects are recruited to assess the performance expected from real-world users.

Interlaboratory studies of this scope are few. In the early 1990s, two studies of informed-user fit and subject-fit protocols were completed. The results of the first study were unpublished and motivated the design of the second study in 1992. Four papers resulted from the second study: Royster *et al.* (1996), Berger *et al.* (1998), Murphy *et al.* (2002), and Murphy *et al.* (2004). As a consequence of the interlaboratory study, ANSI S12.6-1984 was revised in 1997 and included the Method B subject-fit test real-ear attenuation at threshold (REAT). In the second study, as documented in Royster *et al.* (1996), four hearing protectors were evaluated: the Bilsom Blue earmuff, the E•A•R® Classic® foam earplug, the Allsafe V-51R single-flanged premolded earplug, and the Willson EP100 triple-flanged earplug. The devices were selected for laboratory attenuation testing based on the availability of real-world studies for comparison. Berger *et al.* (1998) examined the ability to predict real-world protector performance of the naïve subject-fit protocol as compared to the experimenter-fit protocol on which the current US Noise Reduction Rating (NRR) is based. They concluded that the naïve subject-fit test corresponded more closely, approximating the upper quartile of the real-world studies. Similarly, Berger *et al.* (1998) showed that the naïve subject-fit test rank-ordered relative protector performance almost identically to the real-world studies. Murphy *et al.* (2002) developed mathematical models to describe the distribution of attenuations as a function of frequency and protector. The attenuation distributions from premolded earplugs tended to be bimodal at frequencies below 1000 Hz due to poorly-fit earplugs for some subjects. The distributions were accurately modeled for all frequencies and protectors by a mixed Gaussian distribution. For other protectors and frequencies, the data were normally distributed and did not require a mixed distribution. Murphy *et al.* (2004) reported the statistical analysis of the within-subject, between-subject, and between-laboratory variabilities necessary to estimate sample sizes to determine repeatability and reproducibility. Repeatability characterizes the expected variability if the protectors were to be tested in the same laboratory with the same subject panel. Reproducibility characterizes the expected variability if the protectors were to be tested with a different panel of subjects in the same laboratory or in a different laboratory. This analysis formalized the statistical justification for the standard's use of 10 subjects for earmuffs and 20 subjects for earplugs and semi-aural insert HPDs.

In 2004, European acoustic research and HPD testing laboratories reported a round-robin study where subjects were tested with the same products in several laboratories (Poulsen and Hagerman, 2004). The Nordic round-robin found no significant differences among the laboratories for the several devices tested. In Royster *et al.* (1996), the results from four studies reported between 1976 and 1986 were summarized, finding low variability for repeated measures of subjects' attenuations, large intersubject variability, and larger interlaboratory variability. Beyond this limited set of studies, no additional studies have significant bearing on the results to be presented here.

In November 2004, a test protocol was prepared for the current study. Six testing laboratories agreed to participate: Aearo/E•A•R® E•A•RCAL^{SM} test laboratory (E•A•RCAL), Howard Leight Industries (HLI) test laboratory, Brazil's Laboratory for Acoustic Research Institute (LARI), the National Institute for Occupational Safety and Health Robert A. Taft Laboratory (NIOSH), the U.S. Army Aeromedical Research Laboratory (USAARL), and the U.S. Air Force Research Laboratory (AFRL). The test protocol specified six products to be tested by each laboratory with 24 naïve subjects recruited from the respective local communities, for a total of 144 subjects. Subjects were tested first according to ANSI S12.6-1997 (R2002) (2002) Method B then, after individual instruction, according to Method A. The laboratories began testing in January 2005 and all tests were completed by August 2006.

The subjects' mean hearing thresholds, anthropometric information, mean attenuations, and statistical analyses of the results for repeatability and reproducibility are given in this manuscript.

II. METHODS

A. ANSI S12.6-1997 (R2002)

ANSI S12.6 specifies two protocols to assess the REAT for a hearing protection device: Method A experimenter-supervised fit and Method B subject fit. The experimental protocol required each laboratory to recruit naïve subjects (i.e., no prior experience with testing and limited experience in using hearing protection devices) for the Method B testing first. Each subject was then trained in the fitting of the different protectors and tested according to the Method-A protocol. The protocol deviated from S12.6 in that the standard allows continued use of experienced subjects for Method A whereas only newly-trained naïve subjects were used in this experiment. The participating laboratories decided *post hoc* that this difference had small effect on the reproducibility conclusions reached. The Method-A protocol allows the experimenter to instruct the subject with any variety of training

TABLE I. Subject recruitment and retention by laboratory. Participating laboratories were AFRL, E•A•RCAL, HLI, LARI, NIOSH, and USAARL

Laboratory	Subjects recruited	Subjects tested	Subjects rejected	Subjects dropped
AFRL	36	24	5	7
E•A•RCAL	47	24	17	6
HLI	30	24	6	0
LARI	37	24	13	0
NIOSH	27	24	3	0
USAARL	30	24	0	6

TABLE III. Means and standard deviations of ear canal size (diameters), bitragion width, and head height.

Lab	Right canal size (cm)	Left canal size (cm)	Bitragion width (cm)	Head height (cm)
AFRL	0.88 ± 0.10	0.87 ± 0.09	13.48 ± 0.64	12.75 ± 0.74
EARCal	0.89 ± 0.10	0.89 ± 0.10	14.17 ± 0.73	12.34 ± 1.04
HLI	1.01 ± 0.09	1.01 ± 0.09	13.44 ± 1.62	13.79 ± 0.99
LARI	0.95 ± 0.08	0.95 ± 0.08	13.39 ± 0.85	12.38 ± 0.95
NIOSH	0.95 ± 0.10	0.93 ± 0.10	13.83 ± 0.81	13.97 ± 1.20
USAARL	0.94 ± 0.12	0.94 ± 0.12	13.83 ± 1.31	14.21 ± 1.35

materials and personal demonstration. However, as specified by Method A, the experimenter was not allowed to fit the protector on the subject, though if the fit were judged to be inadequate the experimenter could instruct the subject to refit the product. Laboratories followed their normal practice for instruction; no attempt was made *a priori* to standardize this aspect of the test.

Subjects were informed of any potential risk that they might face during the testing in the laboratory. The human subject use protocols were approved by the CDC-NIOSH human subjects review board, were reviewed by local review boards at USAARL and WPAFB, and complied with the ethical principles of the Acoustical Society of America.

B. Subject selection

Within each of the six laboratories, 24 adult subjects (12 females, 12 males) were recruited. In Table I, the statistics for the number of subjects who were recruited, tested, rejected, or dropped from the study are reported. As specified by ANSI S12.6, the subjects were prohibited from having received prior one-on-one training, were allowed only limited experience with hearing protector testing, were prohibited from having received computer-based or video training in the use of hearing protectors, and were allowed limited experience in wearing hearing protectors during the previous 2 year period. In addition, subjects were required to have normal anatomy of the external ear and ear canal (i.e., no obvious physical deformities), normal otoscopy, and hearing thresholds better than 25 dB HL (re ANSI S3.6-1996) at all test frequencies (125, 250, 500, 1000, 2000, 4000, 8000 Hz). The hearing thresholds were measured using standard audiometric procedures.

The ear canal size, bitragion width (width of the head at the tragus), and head height (tragus to crown) were measured. The ear canal size was measured according to ANSI S12.6 Annex D using the EARGAGE™ which has five different diameters: 7.62, 8.48, 9.27, 10.46, and 11.53 mm. The hearing thresholds and standard deviations are reported in Table II. The anthropometric results are reported in Table III.

C. Products under test

All product samples were purchased on the open market and were provided to the participating laboratories by the NIOSH organizers. Two earmuffs were selected: the Peltor Tactical-Pro and the 3M 1427. The Tactical-Pro earmuff is a sound-restoration electronic muff, which is intended to be worn with the headband over the head. Testing was conducted with batteries in the battery compartment; however, the electronics were turned off. The headband for the 3M 1427 passive earmuff can be worn in three different positions: over the head, under the chin, and behind the head. In this study, the headband was worn behind the head with the

TABLE II. Mean hearing threshold levels and standard deviations (dB HL) for right and left ears.

Laboratory	Ear	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AFRL	Right	8.8 ± 4.9	3.3 ± 4.1	5.0 ± 3.9	4.2 ± 4.3	2.5 ± 3.6	4.0 ± 4.9	7.9 ± 8.2
	Left	8.5 ± 6.2	3.3 ± 3.5	4.2 ± 4.3	4.4 ± 4.0	2.9 ± 4.1	5.6 ± 6.0	4.8 ± 4.8
EARCal	Right	5.2 ± 6.8	3.9 ± 7.1	2.1 ± 6.1	5.0 ± 5.7	5.4 ± 6.1	10.2 ± 7.1	4.1 ± 7.0
	Left	6.6 ± 7.2	4.4 ± 7.5	2.0 ± 7.9	5.8 ± 6.3	6.2 ± 7.5	7.7 ± 6.9	4.7 ± 7.6
HLI	Right	7.1 ± 6.2	6.5 ± 5.2	3.5 ± 5.2	4.2 ± 5.2	2.9 ± 3.9	3.8 ± 4.9	6.7 ± 4.8
	Left	6.3 ± 6.8	5.0 ± 4.7	2.7 ± 3.9	3.1 ± 3.8	3.1 ± 4.4	3.5 ± 5.6	12.3 ± 8.5
LARI	Right	10.2 ± 4.5	8.1 ± 3.6	7.9 ± 4.4	7.5 ± 5.1	6.0 ± 4.7	8.1 ± 7.2	9.6 ± 5.7
	Left	11.7 ± 4.8	10.0 ± 4.2	9.4 ± 4.3	6.0 ± 4.4	5.4 ± 4.4	7.1 ± 5.3	7.7 ± 4.7
NIOSH	Right	6.7 ± 5.2	3.8 ± 4.2	5.2 ± 4.3	1.5 ± 3.5	1.0 ± 5.7	3.3 ± 7.9	6.5 ± 9.0
	Left	5.0 ± 5.3	3.1 ± 5.7	4.4 ± 5.4	2.7 ± 5.5	2.9 ± 7.9	4.2 ± 7.2	7.1 ± 9.1
USAARL	Right	9.8 ± 5.0	8.5 ± 4.8	6.9 ± 5.1	3.3 ± 4.3	0.4 ± 3.9	5.2 ± 7.6	7.3 ± 6.4
	Left	9.8 ± 4.8	9.0 ± 4.7	6.7 ± 5.5	4.6 ± 5.9	2.7 ± 4.2	4.0 ± 5.1	7.5 ± 4.4

TABLE IV. Means and standard deviations of headband clamping force measured in Newtons. (*N* is the number of hearing protectors.)

		Peltor TacticalPro		3M 1427		Moldex Jazzband	
Lab	Force system	Force	*N*	Force	*N*	Force	*N*
AFRL	INSPEC	11.9 ± 0.5	3	11.9 ± 0.5	3	2.5 ± 0.1	24
EARCal	INSPEC	10.2 ± 0.3	3	10.5 ± 0.3	3	2.5 ± 0.1	24
HLI	Load Cell	11.9 ± 0.5	3	10.9 ± 0.6	3	2.3 ± 0.1	24
LARI	Force Gauge	10.2 ± 0.2	3	11.1 ± 0.6	3	3.6 ± 0.2	24
NIOSH	Michael Assoc.	11.2 ± 0.1	3	10.5 ± 0.3	3	2.7 ± 0.1	24
USAARL	Force Gauge	10.5 ± 0.9	3	10.5 ± 0.8	3	2.5 ± 0.5	24

adjustable crown strap placed over the head to prevent the muff from sliding downward. The two muffs were chosen to represent an over-the-head and a behind-the-head earmuff. Twenty pairs of each earmuff were purchased with each laboratory receiving three pairs to be tested in a balanced manner across the 24 test subjects. The two remaining pairs were kept as spares in the event that parts required replacement. Only the vinyl crown strap for the 1427 needed replacement for three of the earmuffs during the course of the study.

Four earplugs were selected: E•A•R® Classic® foam plugs, Howard Leight AirSoft premolded plug, Custom Protect Ear dB Blocker custom silicon earplug, and the Moldex-Metric JazzBand banded hearing protector. The Classic was included to provide reference with the previous studies. The AirSoft was selected because it is a flanged premolded protector. One box of premolded earplugs was provided to each laboratory. The dB Blocker was selected to examine the variability of a custom-molded product across laboratories. Each subject's earmold impression was collected and one pair of custom-molded earplugs were manufactured for each test subject. The Moldex-Metric JazzBand, a banded hearing protector, achieves a seal at the entrance to the ear canal. Twenty-four JazzBands were provided to each laboratory with sufficient replacement tips.

The headband force of each earmuff and canal-cap protector was measured by the participating laboratories. AFRL and E•A•RCAL used the commercially available INSPEC system. NIOSH used the Michaels and Associates headband force system. The other laboratories, Howard Leight, LARI, and USAARL, utilized either a force gauge or a load cell as a part of a custom-built system. Means and standard deviations of headband force were reported in Table IV. The results were not statistically different across laboratories. Although AFRL measured the highest headband force for the earmuffs, it did not measure the highest force for the JazzBand products. Increased headband force can improve attenuation; however, the differences measured here were uncorrelated with the attenuation measurements to be discussed later.

D. Test procedure

Each subject was trained to perform the threshold test in the diffuse sound field at the respective laboratory. Subjects were required to produce three open-ear (unoccluded) thresholds that had a range of no more than 5 dB. Product testing was counter-balanced for product order and occluded/unoccluded order. A subject was required to complete the tests in occluded/unoccluded pairs. All of the Method-B tests for a subject were completed before the subject proceeded with any Method-A tests. Depending on the laboratory, each subject spent between five to seven visits to complete qualification and all product tests.

III. RESULTS

A. Attenuations by frequency

For each laboratory, device, and frequency, the mean Method-A attenuations are given in Table V and the mean Method-B attenuations are given in Table VI. The Method-B attenuations were generally less than the Method-A attenuations measured after the subjects had received instruction and training. For the earmuffs, the low-frequency attenuations ranged from about 10 to 20 dB at 125 and 250 Hz and from about 20 to 30 dB at 500 Hz. At high frequencies (1000–8000 Hz), the attenuations were about 30–40 dB. The Method-A earmuff attenuations were nearly the same or were a few decibels greater than the Method-B earmuff attenuations at all frequencies. For the earplugs tested under Method A, the protectors exhibited about 15–30 dB of attenuation in the low frequencies, 125–250 Hz and ranged from about 25–45 dB at the higher frequencies (500–8000 Hz).

B. *A*-weighted attenuation

Gauger and Berger (2004) conducted a comprehensive analysis of hearing protector rating methods and found that the use of an *A*-weighted statistic provided more accurate estimates of the effective exposure level when hearing protectors are worn. Current methods such as the NRR or the single number rating (SNR) are *C*-weighted statistics in that they are designed to be subtracted from a *C*-weighted exposure level. Following Gauger Berger's (2004) analysis, the Acoustical Society of America's accredited standards committee on noise, Working Group 11 developed ANSI S12.68-2007 that estimates the attenuation for each subject in a test panel across a population of noises (ANSI S12.68, 2007; Gauger and Berger, 2004; Johnson and Nixon, 1974; Kroes *et al.*, 1975). For each noise, the *A*-weighted attenuation is computed and the average and standard deviation of the at-

TABLE V. Method-A mean attenuations and standard deviations in dB at each test frequency, by laboratory and protector.

Lab	Protector	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AFRL	TacPro	10.8 ± 2.5	18.2 ± 2.3	26.8 ± 2.7	35.1 ± 2.0	34.6 ± 2.5	35.4 ± 2.7	38.1 ± 2.6
EARCAL	TacPro	14.2 ± 5.2	19.2 ± 6.1	28.6 ± 6.9	32.9 ± 5.3	32.9 ± 4.7	35.0 ± 8.0	36.1 ± 6.4
HLI	TacPro	15.0 ± 2.7	21.9 ± 2.4	32.0 ± 2.6	34.2 ± 2.8	33.7 ± 3.1	37.9 ± 2.6	40.8 ± 2.1
LARI	TacPro	11.4 ± 3.9	20.3 ± 5.4	31.4 ± 2.2	32.6 ± 2.9	34.2 ± 1.9	35.3 ± 4.5	36.2 ± 2.9
NIOSH	TacPro	15.0 ± 3.8	21.4 ± 3.8	31.1 ± 3.4	32.3 ± 4.3	35.1 ± 3.2	38.1 ± 2.4	40.6 ± 3.8
USAARL	TacPro	12.6 ± 3.7	19.5 ± 3.7	24.9 ± 4.0	28.2 ± 3.8	28.9 ± 4.5	30.4 ± 4.8	35.0 ± 4.7
AFRL	1427	9.4 ± 6.6	13.4 ± 6.1	21.3 ± 5.8	36.5 ± 6.5	35.6 ± 2.7	34.5 ± 4.7	33.7 ± 4.6
EARCAL	1427	15.5 ± 7.3	16.3 ± 7.0	25.9 ± 6.6	36.0 ± 5.3	33.5 ± 3.3	36.2 ± 6.9	33.7 ± 6.8
HLI	1427	18.1 ± 4.0	20.3 ± 2.9	28.9 ± 4.3	36.0 ± 5.1	35.2 ± 2.7	38.6 ± 3.1	38.7 ± 2.4
LARI	1427	11.4 ± 5.8	16.3 ± 4.9	24.0 ± 6.1	33.1 ± 5.8	34.7 ± 3.0	32.3 ± 6.8	33.6 ± 4.3
NIOSH	1427	13.5 ± 8.2	16.8 ± 8.2	24.9 ± 6.8	32.1 ± 7.4	34.2 ± 4.3	36.6 ± 6.9	35.9 ± 5.1
USAARL	1427	13.6 ± 5.9	16.7 ± 6.4	22.5 ± 7.0	31.3 ± 6.7	29.8 ± 4.1	29.0 ± 6.8	30.9 ± 5.0
AFRL	dB Blocker	25.6 ± 6.2	25.7 ± 6.9	26.5 ± 6.0	30.7 ± 5.9	36.9 ± 4.5	41.0 ± 3.9	44.4 ± 4.2
EARCAL	dB Blocker	28.5 ± 9.0	27.4 ± 8.9	28.4 ± 9.7	27.8 ± 7.7	32.6 ± 5.4	39.9 ± 7.1	40.6 ± 6.6
HLI	dB Blocker	25.4 ± 6.5	24.9 ± 5.6	25.5 ± 5.2	25.5 ± 5.5	33.2 ± 4.5	40.8 ± 4.8	40.5 ± 7.1
LARI	dB Blocker	24.6 ± 9.1	27.1 ± 9.4	29.1 ± 8.6	28.4 ± 7.2	32.7 ± 6.8	39.7 ± 7.0	39.4 ± 6.3
NIOSH	dB Blocker	25.7 ± 7.8	27.2 ± 7.6	27.6 ± 7.1	25.7 ± 5.8	32.6 ± 4.2	40.0 ± 5.0	40.5 ± 7.0
USAARL	dB Blocker	24.9 ± 9.8	24.3 ± 8.4	24.4 ± 8.3	23.8 ± 5.9	27.7 ± 6.2	30.4 ± 6.7	34.2 ± 10.3
AFRL	JazzBand	19.0 ± 5.2	17.3 ± 5.0	16.6 ± 5.0	23.6 ± 3.9	32.6 ± 3.8	35.8 ± 3.3	38.3 ± 6.3
EARCAL	JazzBand	16.6 ± 9.1	15.5 ± 8.5	14.8 ± 7.6	18.0 ± 6.5	27.4 ± 5.7	34.9 ± 6.3	33.6 ± 7.6
HLI	JazzBand	24.7 ± 6.0	22.8 ± 5.5	21.4 ± 5.0	22.7 ± 4.4	32.3 ± 4.0	40.0 ± 4.3	42.9 ± 4.0
LARI	JazzBand	15.6 ± 8.7	15.0 ± 7.6	17.1 ± 6.2	18.4 ± 5.3	27.8 ± 5.9	34.8 ± 6.2	32.6 ± 6.8
NIOSH	JazzBand	20.9 ± 6.7	20.7 ± 6.7	18.4 ± 4.9	20.2 ± 4.9	30.3 ± 5.5	37.7 ± 4.4	37.0 ± 8.3
USAARL	JazzBand	19.1 ± 5.9	19.9 ± 5.6	17.4 ± 5.3	19.1 ± 5.7	26.8 ± 5.5	31.4 ± 6.1	33.5 ± 8.8
AFRL	Classic	30.9 ± 5.0	33.1 ± 4.7	35.4 ± 4.4	38.3 ± 3.6	36.9 ± 3.1	41.0 ± 2.2	47.3 ± 2.4
EARCAL	Classic	27.5 ± 8.9	28.5 ± 8.9	32.6 ± 9.7	31.0 ± 7.0	32.0 ± 4.1	41.2 ± 4.0	43.0 ± 5.0
HLI	Classic	32.1 ± 8.3	33.9 ± 7.6	36.9 ± 7.7	33.4 ± 6.0	34.2 ± 4.0	43.4 ± 3.3	46.3 ± 4.6
LARI	Classic	22.9 ± 4.9	26.6 ± 5.3	29.6 ± 5.9	28.7 ± 5.0	32.0 ± 3.3	40.7 ± 2.8	41.1 ± 4.8
NIOSH	Classic	24.4 ± 8.2	27.9 ± 8.7	31.0 ± 9.4	28.3 ± 7.0	33.0 ± 4.3	40.5 ± 3.2	43.5 ± 6.9
USAARL	Classic	19.7 ± 5.0	19.8 ± 4.4	19.6 ± 5.0	19.4 ± 4.9	26.9 ± 5.3	32.9 ± 5.3	35.5 ± 5.5
AFRL	AirSoft	18.8 ± 7.9	19.5 ± 7.6	20.5 ± 7.4	26.5 ± 6.5	32.0 ± 5.6	32.6 ± 7.5	40.0 ± 9.1
EARCAL	AirSoft	26.6 ± 7.5	25.8 ± 8.7	27.7 ± 9.4	27.1 ± 7.9	29.4 ± 4.7	34.9 ± 7.5	39.5 ± 9.3
HLI	AirSoft	29.4 ± 7.1	29.4 ± 6.6	31.3 ± 7.5	31.0 ± 5.6	32.0 ± 4.0	37.2 ± 6.9	44.7 ± 5.5
LARI	AirSoft	23.6 ± 8.8	25.2 ± 8.7	29.0 ± 9.7	27.8 ± 8.8	32.9 ± 5.0	37.7 ± 8.7	38.9 ± 7.8
NIOSH	AirSoft	24.9 ± 8.1	25.2 ± 8.1	27.3 ± 9.4	25.2 ± 8.1	30.3 ± 6.6	35.1 ± 8.9	40.6 ± 9.7
USAARL	AirSoft	20.3 ± 11.2	20.5 ± 10.4	20.9 ± 11.4	21.2 ± 10.7	26.1 ± 7.8	28.1 ± 8.8	33.1 ± 11.7

tenuations are used to estimate the SNR. To simplify the analysis, the *A*-weighted attenuation for pink noise is used to perform all of the subsequent analyses. The difference between the levels for *C*-weighted and *A*-weighted pink noise, $L_C - L_A$, is about 1.0 dB, which is close to the median difference, 1.8 dB, for the NIOSH 100 noises. Thus the results can be related to the expected results when the S12.68 standard is applied to determine a rating for the protector. The *A*-weighted attenuation using pink noise removes the additional element of variance across noises and allows more direct comparison of the attenuations due to the subject panels as well as retrospective comparison to the previous interlaboratory studies (Murphy *et al.*, 2004).

The overall *A*-weighted attenuation for each subject/protector combination was calculated using the following equation:

$$\text{Atten} = 10\log\left(\sum_{f=125}^{8000} 10^{L_f + A_f}\right) - 10\log\left(\sum_{f=125}^{8000} 10^{L_f + A_f - \text{Atten}_{f,\text{avg}}}\right), \quad (1)$$

where the test frequencies were $f = 125, 250, 500, 1000, 2000, 4000, 8000$ Hz. The *A*-weighting correction factors were $A_f = -16.1$, -8.6, -3.2, 0.0, 1.2, 1.0, -1.1 at the respective test frequencies. The noise spectrum levels were pink noise, $L_f = 100$ at all frequencies. The first summation yields 107.0 dB rounded to a tenth of a decibel. The attenuation, $\text{Atten}_{f,\text{avg}}$, measured from each subject's two paired occluded and unoccluded trials were averaged at each frequency. For example, rounded to a tenth of a decibel, one subject's Method-A attenuations were

TABLE VI. Method-B mean attenuations and standard deviations in dB at each test frequency, by laboratory and protector.

Lab	Protector	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AFRL	TacPro	8.7 ± 3.2	16.5 ± 4.6	24.4 ± 4.1	34.4 ± 2.1	33.7 ± 2.6	34.5 ± 3.9	37.4 ± 4.3
EARCAL	TacPro	13.6 ± 5.9	19.2 ± 7.4	28.7 ± 9.1	32.9 ± 6.9	32.2 ± 5.7	35.1 ± 9.2	35.5 ± 7.8
HLI	TacPro	13.0 ± 5.1	19.2 ± 6.3	30.0 ± 5.6	33.1 ± 4.4	33.2 ± 2.9	36.0 ± 4.3	38.6 ± 3.5
LARI	TacPro	10.5 ± 4.2	20.3 ± 5.8	30.1 ± 2.7	31.9 ± 3.9	33.7 ± 3.2	34.2 ± 4.2	36.0 ± 5.1
NIOSH	TacPro	14.1 ± 4.8	21.1 ± 4.0	29.5 ± 5.0	31.3 ± 4.4	34.4 ± 3.7	36.5 ± 4.4	39.8 ± 3.8
USAARL	TacPro	12.7 ± 4.1	18.9 ± 4.0	25.0 ± 5.0	28.0 ± 4.4	29.3 ± 4.0	30.7 ± 5.5	34.0 ± 5.0
AFRL	1427	10.3 ± 4.0	14.9 ± 3.6	22.8 ± 4.2	36.9 ± 3.8	36.0 ± 3.1	34.7 ± 4.7	34.1 ± 5.0
EARCAL	1427	12.7 ± 6.8	16.1 ± 6.5	25.9 ± 6.2	34.9 ± 6.5	32.6 ± 4.1	33.5 ± 6.9	31.6 ± 7.0
HLI	1427	14.9 ± 7.4	18.0 ± 7.3	26.9 ± 6.7	34.6 ± 6.3	34.5 ± 3.0	35.4 ± 4.7	34.6 ± 5.2
LARI	1427	9.6 ± 5.6	14.5 ± 6.9	24.7 ± 5.5	33.4 ± 5.5	34.2 ± 3.7	31.1 ± 7.2	32.2 ± 5.9
NIOSH	1427	12.3 ± 7.6	15.9 ± 7.2	24.7 ± 7.3	33.1 ± 7.5	34.3 ± 4.9	34.6 ± 6.9	32.7 ± 6.7
USAARL	1427	13.6 ± 4.9	17.7 ± 5.6	23.1 ± 7.0	31.2 ± 5.6	28.9 ± 5.7	28.2 ± 7.4	27.8 ± 6.8
AFRL	dB Blocker	23.2 ± 9.0	24.5 ± 8.5	25.2 ± 9.2	29.2 ± 8.0	34.3 ± 8.1	38.7 ± 7.2	41.4 ± 8.8
EARCAL	dB Blocker	25.1 ± 13.0	25.9 ± 14.0	25.8 ± 14.1	24.8 ± 12.5	27.7 ± 11.7	34.9 ± 13.1	34.3 ± 15.1
HLI	dB Blocker	21.1 ± 7.8	21.7 ± 7.7	23.6 ± 7.8	23.3 ± 5.7	30.7 ± 4.9	39.4 ± 5.7	38.7 ± 6.1
LARI	dB Blocker	22.4 ± 10.4	24.1 ± 10.9	25.8 ± 10.4	25.5 ± 8.5	30.9 ± 8.5	35.5 ± 8.4	35.3 ± 9.9
NIOSH	dB Blocker	24.5 ± 9.8	25.4 ± 10.0	25.0 ± 10.4	23.1 ± 8.1	30.2 ± 7.2	37.2 ± 7.4	35.4 ± 9.9
USAARL	dB Blocker	21.0 ± 9.4	21.0 ± 10.3	21.7 ± 9.5	21.7 ± 9.2	27.0 ± 8.7	29.3 ± 9.0	31.9 ± 11.5
AFRL	JazzBand	16.4 ± 8.1	15.1 ± 7.9	14.6 ± 7.0	21.6 ± 5.6	29.7 ± 6.5	33.1 ± 6.5	33.4 ± 10.1
EARCAL	JazzBand	11.2 ± 9.3	11.3 ± 8.2	11.5 ± 8.0	14.8 ± 6.9	24.1 ± 7.4	31.2 ± 6.3	28.6 ± 8.6
HLI	JazzBand	17.6 ± 8.6	17.5 ± 8.1	17.3 ± 6.1	18.5 ± 6.0	27.1 ± 6.8	35.0 ± 7.2	33.0 ± 8.8
LARI	JazzBand	13.9 ± 9.0	13.7 ± 8.9	13.8 ± 9.7	14.6 ± 9.3	23.9 ± 10.0	30.4 ± 10.2	29.4 ± 11.7
NIOSH	JazzBand	18.3 ± 8.1	18.5 ± 7.4	17.3 ± 6.5	18.4 ± 5.2	27.8 ± 5.9	33.2 ± 6.2	33.4 ± 8.3
USAARL	JazzBand	17.0 ± 7.5	15.7 ± 7.3	15.2 ± 6.8	16.2 ± 7.2	23.6 ± 7.5	28.1 ± 9.3	30.0 ± 10.8
AFRL	Classic	18.5 ± 7.0	19.0 ± 6.8	21.2 ± 7.2	25.6 ± 5.7	31.7 ± 4.2	38.0 ± 4.1	40.9 ± 7.2
EARCAL	Classic	18.5 ± 9.8	19.0 ± 9.8	20.6 ± 11.8	20.8 ± 9.9	28.8 ± 6.0	37.1 ± 7.5	34.5 ± 10.7
HLI	Classic	19.9 ± 6.0	21.1 ± 6.3	23.1 ± 7.2	21.9 ± 6.3	29.3 ± 4.2	39.5 ± 5.1	40.2 ± 6.1
LARI	Classic	16.6 ± 6.5	19.0 ± 6.5	22.6 ± 7.3	21.5 ± 5.2	29.6 ± 3.8	37.8 ± 3.7	36.9 ± 5.2
NIOSH	Classic	17.3 ± 5.7	18.1 ± 5.2	19.5 ± 5.8	19.1 ± 4.5	27.9 ± 5.8	36.5 ± 4.8	34.6 ± 7.4
USAARL	Classic	16.6 ± 5.9	17.5 ± 5.5	17.7 ± 5.8	17.5 ± 6.0	25.4 ± 5.9	31.1 ± 7.6	31.5 ± 7.5
AFRL	AirSoft	16.2 ± 8.4	16.1 ± 8.6	18.1 ± 9.2	24.0 ± 8.8	30.7 ± 5.3	31.1 ± 7.4	38.7 ± 10.2
EARCAL	AirSoft	21.6 ± 11.3	21.6 ± 11.9	22.1 ± 13.9	23.0 ± 12.4	26.3 ± 10.0	31.2 ± 9.6	34.0 ± 13.9
HLI	Airsoft	20.7 ± 9.8	21.0 ± 10.2	22.3 ± 9.7	23.2 ± 8.1	27.3 ± 6.5	31.1 ± 8.3	35.9 ± 11.7
LARI	AirSoft	19.5 ± 9.4	22.1 ± 10.1	24.9 ± 11.3	24.6 ± 10.0	29.4 ± 8.4	36.1 ± 10.0	38.1 ± 11.7
NIOSH	AirSoft	20.3 ± 10.6	20.5 ± 11.1	21.9 ± 11.2	20.4 ± 9.7	27.2 ± 8.3	32.6 ± 10.2	36.0 ± 13.5
USAARL	AirSoft	16.5 ± 12.2	16.3 ± 12.0	17.3 ± 12.0	17.1 ± 11.3	22.5 ± 10.5	23.8 ± 10.9	27.3 ± 14.3

$$\text{Atten}_{f,1} = [19.2, 27.7, 37.7, 35.7, 33.0, 36.0, 41.7],$$

$$\text{Atten}_{f,2} = [19.0, 23.0, 32.3, 37.8, 34.3, 36.3, 37.8],$$

$$\text{Atten}_{f,A,\text{avg}} = [19.1, 25.3, 35.0, 36.8, 33.6, 36.2, 39.8].$$

The second summation of Eq. (1) yielded 73.1 dB. The *A*-weighted attenuation for this subject and device was 107.0−73.1=33.9 dB. The Method-B attenuations were

$$\text{Atten}_{f,B,\text{avg}} = [15.3, 24.1, 31.1, 32.2, 32.8, 30.1, 38.5],$$

which yields an *A*-weighted attenuation of 30.6 dB for this same subject.

The differences in the *A*-weighted attenuations between Methods A and B were determined for each subject to facilitate comparison between methods,

$$\Delta_{\text{A-B}} = A_{\text{A}} - A_{\text{B}}, \qquad (2)$$

where A_A and A_B are the attenuations from Eq. (1). For the example data, the difference was 3.3 dB. The differences in *A*-weighted attenuation were analyzed by grouping the subjects within a laboratory and also by pooling all of the subjects together across all laboratories. The averaged differences and the statistical analyses are presented in Table VII.

In Fig. 1 the *A*-weighted attenuations are compared graphically with box-whisker plots. Method-A data are shown on the left of each pair and Method-B data on the right in accordance with standard alphabetical preference; note that the actual sequence of testing was Method B and then Method A. The vertical length of each box represents the interquartile range, which extends from the 25th to the 75th percentiles (i.e., 50% of all data points are contained within the box). The horizontal line inside the box indicates the median value, and the whiskers depict the 10th and 90th

TABLE VII. *A*-weighted attenuations in dB for Methods A and B, attenuation differences (Δ_{A-B}), standard error, student's *t* test, probability associated with *t*, and lower and upper 95% confidence interval boundaries, across all laboratories and by individual laboratory for each protector.

Protector	Lab	*A*-weighted attenuation Method A	*A*-weighted attenuation Method B	Difference Δ_{A-B}	Standard error	Stud. value	Prob. of *t*	Conf. Int. (lower, upper)
TacticalPro	All Labs	29.2	28.1	1.1	0.33	3.350	0.0203	(0.26, 1.95)
TacticalPro	AFRL	28.5	26.7	1.8	0.71	2.612	0.016	(0.39, 3.32)
TacticalPro	EARCAL	29.0	28.7	0.3	1.35	0.227	0.823	(−2.48, 3.09)
TacticalPro	HLI	31.5	29.2	2.3	0.89	2.588	0.016	(0.46, 4.14)
TacticalPro	LARI	29.1	28.3	0.8	0.46	1.925	0.067	(−0.07, 1.82)
TacticalPro	NIOSH	30.8	29.9	0.9	0.64	1.464	0.157	(−0.39, 2.25)
TacticalPro	USAARL	26.5	26.1	0.4	0.74	0.482	0.634	(−1.18, 1.89)
3M 1427	All Labs	27.3	26.5	0.8	0.58	1.080	0.3296	(−0.87, 2.12)
3M 1427	AFRL	24.9	26.4	−1.5	0.86	−1.743	0.095	(−3.26, 0.28)
3M 1427	EARCAL	28.0	27.1	0.9	1.15	0.807	0.428	(−1.45, 3.31)
3M 1427	HLI	31.4	28.6	2.8	1.07	2.668	0.014	(0.64, 5.05)
3M 1427	LARI	26.1	24.9	1.2	0.79	1.434	0.165	(−0.50, 2.76)
3M 1427	NIOSH	27.1	26.9	0.2	1.06	0.150	0.882	(−2.03, 2.35)
3M 1427	USAARL	25.4	25.2	0.2	1.37	0.145	0.886	(−2.64, 3.04)
AirSoft	All Labs	27.8	23.7	4.1	0.64	6.359	0.0014	(2.43, 5.72)
AirSoft	AFRL	26.2	23.6	2.6	1.21	2.142	0.0430	(0.09, 5.10)
AirSoft	EARCAL	28.6	24.4	4.2	1.85	2.289	0.0316	(0.41, 8.04)
AirSoft	HLI	31.7	24.7	7.0	1.57	4.454	0.0002	(3.74, 10.24)
AirSoft	LARI	29.9	26.9	3.0	1.30	2.372	0.0264	(0.39, 5.78)
AirSoft	NIOSH	27.9	23.7	4.2	1.46	2.912	0.0079	(1.23, 7.26)
AirSoft	USAARL	22.4	19.1	3.3	1.85	1.789	0.0867	(−0.52, 7.13)
Classic	All Labs	31.4	23.8	7.6	1.32	5.696	0.0023	(4.12, 10.91)
Classic	AFRL	37.3	26.4	10.9	1.32	8.231	0.0000	(8.13, 13.59)
Classic	EARCAL	31.9	23.5	8.4	1.47	5.764	0.0000	(5.43, 11.50)
Classic	HLI	35.0	25.3	9.7	1.12	8.731	0.0000	(7.46, 12.09)
Classic	LARI	30.7	24.6	6.1	1.09	5.588	0.0000	(3.83, 8.34)
Classic	NIOSH	30.6	22.5	8.1	1.40	5.789	0.0000	(5.21, 11.01)
Classic	USAARL	22.6	20.8	1.8	1.00	1.800	0.0851	(−0.27, 3.87)
JazzBand	All Labs	22.7	19.6	3.1	0.36	8.591	0.0004	(2.15, 3.99)
JazzBand	AFRL	23.7	21.5	2.2	1.06	2.062	0.0506	(−0.01, 4.38)
JazzBand	EARCAL	20.4	16.8	3.6	1.13	3.164	0.0043	(1.24, 5.93)
JazzBand	HLI	25.9	21.5	4.4	1.00	4.463	0.0002	(2.38, 6.50)
JazzBand	LARI	21.0	17.9	3.1	1.43	2.195	0.0385	(0.18, 6.09)
JazzBand	NIOSH	23.4	21.3	2.1	0.78	2.735	0.0118	(0.52, 3.73)
JazzBand	USAARL	21.7	18.8	2.9	1.40	2.130	0.0441	(0.09, 5.86)
dB Blocker	All Labs	29.1	26.7	2.4	0.31	7.554	0.0006	(1.56, 3.18)
dB Blockers	AFRL	31.7	30.0	1.7	1.57	1.083	0.2901	(−1.55, 4.94)
dB Blockers	EARCAL	30.1	26.5	3.6	2.37	1.519	0.1424	(−1.30, 8.51)
dB Blockers	HLI	28.7	26.2	2.5	1.15	2.163	0.0411	(0.11, 4.87)
dB Blockers	LARI	30.3	27.5	2.8	1.43	1.851	0.0771	(−0.31, 5.61)
dB Blockers	NIOSH	28.8	26.4	2.4	2.07	1.133	0.2687	(−1.93, 6.61)
dB Blockers	USAARL	25.1	23.7	1.4	1.85	0.774	0.4469	(−2.39, 5.25)

percentiles. The points depict the individual subject results outside the 10th and 90th percentiles. In panel A of Fig. 1, the Peltor Tactical Pro earmuff yielded essentially identical results for Methods A and B for four laboratories and statistically significant increases in attenuation for Method A over Method B for the AFRL and HLI laboratories. In most cases, the lowest attenuations were eliminated with the training of the naïve subjects. Therefore, the differences can be attributed to the test subjects' lack of experience with hearing protector fitting and testing. The difference for the Peltor product was small yet statistically significant with $\Delta_{A-B} = 1.10$, $p = 0.0203$, $\mathrm{CI} = (0.26, 1.95)$, where p is the probability of significance and CI are the upper and lower limits of the 95% confidence interval.

Attenuation values for the 3M 1427 earmuff were higher for Method A, although the differences between the two methods were slight. The HLI laboratory exhibited the greatest difference between Methods A and B in the lower frequencies. The HLI laboratory was the only laboratory where the difference from zero was statistically significant at p

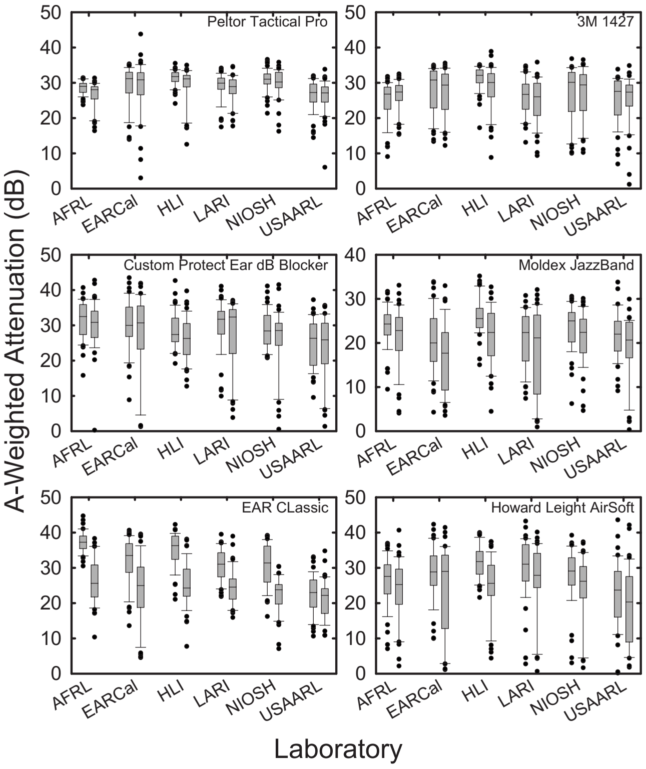


FIG. 1. The A-weighted attenuation estimated from the REAT for each method, protector, and laboratory. In each set of paired results, Method A attenuations are displayed on the left and Method B on the right. Each box plot gives the median, 25th and 75th percentiles; the error bars indicate the 10th and 90th percentiles; individual point represent the attenuations observed on individual trials for subjects outside the 10th and 90th percentiles.

<0.05: $\Delta_{A-B}=2.84$, $p=0.014$, CI$=(0.64,5.05)$. Overall, the 3M product did not exhibit a statistically significant difference between Methods A and B.

The Custom Protect Ear dB Blocker results for the overall laboratory and individual laboratory difference analyses exhibited different trends in the improvement of the A-weighted attenuation between Methods B and A. Five of the laboratories showed no statistically significant improvement; the confidence interval includes 0. The HLI laboratory exhibited a significant improvement: $\Delta_{A-B}=2.49$, $p=0.04$, CI$=(0.11,4.87)$; the confidence interval excludes 0. The overall laboratory difference analysis exhibited a statistically significant improvement, $\Delta_{A-B}=2.37$, $p=0.0006$, CI $=(1.56,3.18)$. The attenuations at individual frequencies improved between 1 and 5 dB while the standard deviations decreased about 1–2 dB (see Tables V and VI). The change in attenuation based on 24 samples from a single laboratory was not significant; however, pooling the subjects' results (thereby increasing the sample size to 144) causes incremental improvements in attenuation to become significant when exhibited in a larger population.

The JazzBand canal caps exhibited significant improvement between Methods B and A. Only the AFRL subjects did not demonstrate statistically significant improvement at the $p=0.05$ level, but the probability was nearly significant: $\Delta_{A-B}=2.19$, $p=0.0506$, CI$=(-0.01,4.38)$. Δ_{A-B} for the other laboratories ranged between 2.12 dB at NIOSH and 4.44 dB at Howard Leight. Overall, the improvement was $\Delta_{A-B}=3.07$, $p=0.004$, CI$=(2.15,3.99)$. The additional instruction in Method A significantly improved the performance of naïve subjects.

The E•A•R® Classic® earplugs exhibited the greatest effect due to testing under Method A versus Method B. Overall, $\Delta_{A-B}=7.52$, $p=0.0023$, CI$=(4.12,10.91)$. The Classic also exhibited the greatest variability across laboratories. The AFRL laboratory measured A-weighted Method-A attenuations above 30 dB for all of its test subjects: $\Delta_{A-B}=10.88$, $p<0.0001$, CI$=(8.13,13.59)$. However, USAARL exhibited no significant improvement in Method-A testing: $\Delta_{A-B}=1.80$, $p=0.851$, CI$=(-0.27,3.87)$. From the working group's discussions of the results, the AFRL laboratory was determined to have scrutinized the subjects under Method A the most carefully prior to commencing Method-A testing. However at the USAARL laboratory, subjects were trained to properly fit the E•A•R® earplugs but the experimenter did not monitor the subjects' fit prior to Method-A testing. Thus, experimenter involvement was an essential element for increasing the attenuation under Method-A protocol. As a result, ANSI S12 Working Group 11 deliberated extensively to improve the procedure to make the requirements more explicit and less open to interpretation.

Finally, the Howard Leight AirSoft earplug showed significant improvement between Methods B and A: $\Delta_{A-B}=4.07$, $p=0.0014$, CI$=(2.43,5.72)$. Like the Classic, only one laboratory did not demonstrate significant improvement. The HLI laboratory had the greatest improvement, $\Delta_{A-B}=6.99$, $p=0.0002$, CI$=(3.74,10.24)$.

C. Statistical analysis

The statistical power calculation developed by Murphy *et al.* (2004) for the previous interlaboratory study was used in the present data analysis. Repeatability is defined as testing the same product under identical testing conditions (i.e., same subjects, equipment, and environment). Reproducibility can be defined in two ways. Between-subject reproducibility is the expected variability if a different group of subjects were to be recruited and tested in the same laboratory with the same equipment and environment. Between-laboratory reproducibility is the expected variability if the product were to be tested in a different laboratory that complies to the same standard testing protocol.

A multi-level analysis of variance (ANOVA) (Netter *et al.*, 1990) was used to estimate the standard deviations for laboratory, subject, and trial effects. The statistical model was

$$Y_{ijk} = \mu + \mathrm{Trial}_{k(ij)} + \mathrm{Subject}_{j(i)} + \mathrm{Lab}_i, \qquad (3)$$

where Y_{ijk} is the measured attenuation of a given trial, subject, and laboratory, μ is the overall attenuation, $\mathrm{Trial}_{k(ij)}$ is the random term for the kth trial within the jth subject and ith laboratory, $\mathrm{Subject}_{j(i)}$ is the random term for the jth subject within the ith laboratory, and Lab_i is the random term for the ith laboratory. Variance components and means were estimated with the MIXED procedure in SAS® (SAS, 2007). The method of estimation was residual maximum likelihood. The standard deviations for within-subject repeatability,

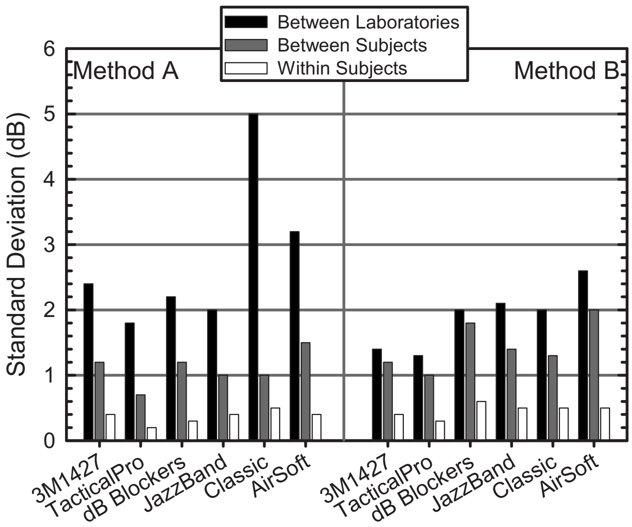


FIG. 2. The within-subject, between-subject, and between-laboratory standard deviations for the *A*-weighted attenuations of each protector using Methods A and B.

between-subject reproducibility, and between-laboratory reproducibility have particular meaning when considering repeated tests of the same product. Between-subject reproducibility is the variability one might expect if the product were tested in the same laboratory with the same subjects and identical conditions. Between-laboratory reproducibility evaluates the testing of a different group of subjects in another laboratory applying the same testing protocol (i.e., similar equipment, psychophysical paradigm, and training).

Within-subject standard deviation was calculated with the equation,

$$\sigma_{\text{within-subject}} = \sqrt{\frac{\sigma_{\text{trial}}^2}{(n_s n_t)}}, \tag{4}$$

where σ_{trial}^2 was the trial-to-trial variance, $n_s=24$ was the number of subjects, and $n_t=2$ was the number of trials per subject. Between-subject standard deviation was calculated with the equation,

$$\sigma_{\text{between-subject}} = \sqrt{\frac{\sigma_{\text{subject}}^2}{n_s} + \frac{\sigma_{\text{trial}}^2}{(n_s n_t)}}, \tag{5}$$

where $\sigma_{\text{subject}}^2$ was the subject-to-subject variance. Between-laboratory standard deviation was calculated with the equation,

$$\sigma_{\text{between-laboratory}} = \sqrt{\sigma_{\text{laboratory}}^2 + \frac{\sigma_{\text{subject}}^2}{n_s} + \frac{\sigma_{\text{trial}}^2}{(n_s n_t)}}, \tag{6}$$

where $\sigma_{\text{laboratory}}^2$ was the laboratory-to-laboratory variance. The standard deviation estimates of $\sigma_{\text{within-subject}}$, $\sigma_{\text{between-subject}}$, and $\sigma_{\text{between-laboratory}}$ were calculated in the multi-level ANOVA, and are presented in Fig. 2. By definition, the standard deviation for any given protector increases progressively as more variance terms are added to the computation. Thus, within-subject standard deviations are the smallest and between-laboratory standard deviations are the largest.

1. Within-subject variability

The within-subject standard deviations were between 0.2 and 0.8 dB for both Methods A and B (see Table VIII). This performance statistic was slightly higher for Method B compared to Method A. The JazzBand and dB Blockers exhibited less variability at all frequencies for Method A. For individual frequencies, the agreement between the Method-A and Method-B within-subject standard deviations were between 0.0 and 0.2 dB. With the exception of the dB Blocker, the *A*-weighted attenuation within-subject standard deviations exhibit either no difference or 0.1-dB differences. The standard deviations of the *A*-weighted attenuations tend to be close to the values in the 1000–4000 Hz region. The improvements of the standard deviations from Method B to Method A were 0.0 dB for the Classic and 3M 1427 devices to 0.1 dB for the AirSoft, JazzBand, and Tactical Pro and 0.3 dB for the dB Blocker. The small decreases for the within-subject standard deviations could be a random effect, a learning effect that results from the subjects having more

TABLE VIII. Within-subject standard deviations in dB for Methods A and B for each protector.

Protector	Method	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	*A*-weight Atten
TacticalPro	A	0.4	0.4	0.3	0.4	0.4	0.3	0.4	0.2
	B	0.4	0.4	0.5	0.3	0.4	0.4	0.4	0.3
3M 1427	A	0.5	0.5	0.5	0.5	0.4	0.4	0.5	0.4
	B	0.5	0.6	0.5	0.5	0.4	0.5	0.4	0.4
dB Blocker	A	0.6	0.5	0.5	0.5	0.4	0.4	0.5	0.3
	B	0.7	0.7	0.7	0.6	0.5	0.6	0.7	0.6
JazzBand	A	0.6	0.5	0.5	0.5	0.4	0.5	0.5	0.4
	B	0.7	0.7	0.6	0.5	0.5	0.6	0.7	0.5
Classic	A	0.7	0.7	0.6	0.5	0.4	0.5	0.5	0.5
	B	0.8	0.8	0.7	0.6	0.5	0.6	0.7	0.5
AirSoft	A	0.6	0.5	0.6	0.6	0.5	0.5	0.6	0.4
	B	0.7	0.6	0.7	0.5	0.5	0.6	0.7	0.5

TABLE IX. Between-subject standard deviations in dB for Methods A and B for each protector.

Protector	Method	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	*A*-weight Atten
TacticalPro	A	0.8	0.9	0.8	0.8	0.7	0.9	0.8	0.7
	B	0.9	1.1	1.1	0.9	0.8	1.1	1.0	1.0
3M 1427	A	1.3	1.3	1.3	1.3	0.7	1.2	1.0	1.2
	B	1.3	1.3	1.3	1.2	0.9	1.3	1.3	1.2
dB Blocker	A	1.7	1.6	1.6	1.3	1.1	1.2	1.5	1.2
	B	2.0	2.1	2.1	1.8	1.7	1.8	2.2	1.8
JazzBand	A	1.4	1.3	1.2	1.1	1.0	1.1	1.5	1.0
	B	1.7	1.6	1.5	1.4	1.5	1.6	2.0	1.4
Classic	A	1.4	1.4	1.5	1.2	0.8	0.7	1.0	1.0
	B	1.4	1.4	1.6	1.3	1.0	1.2	1.5	1.3
AirSoft	A	1.7	1.7	1.9	1.7	1.2	1.7	1.8	1.5
	B	2.1	2.2	2.3	2.1	1.7	1.9	2.6	2.0

experience with the threshold identification task, or an effect of the training in fitting the protector. The study was not designed to test for such effects.

2. Between-subject variability

The variability between subjects is summarized in Table IX. For earmuffs, the standard deviations were 0.7–1.3 dB. For the Peltor TacticalPro, the standard deviations were consistently 0.1–0.3 dB lower for Method A compared to Method B. For frequencies 125–1000 Hz, the between-subject variabilities for the 3M 1427 were essentially the same for both methods, and had slightly less variability in the high frequencies. For the JazzBand and dB Blocker, the Method-B between-subject deviations were 0.3–0.7 dB greater than Method-A deviations. For the Classic at frequencies 125–1000 Hz, the standard deviations were the same or 0.1 dB lower. At higher frequencies (2000–8000 Hz), the Classic's standard deviations for Method A were 0.2–0.5 dB less than the standard deviations observed for Method B. For the AirSoft, the between-subject variability was larger by 0.2–0.8 dB for Method B than for Method A. Differences in standard deviations between Methods A and B less than 1 dB may exhibit statistical significance but have no practical importance.

The standard deviations for the *A*-weighted attenuations tended to correlate with the 1, 2, and 4 kHz estimates and not the 8 kHz deviations. This trend was most evident in the JazzBand, dB Blocker, and AirSoft results. With the exception of the 3M 1427 muff, the *A*-weighted standard deviations were all less for the Method-A than the Method-B testing. For the Classic, the higher frequencies exhibited less variability for Method A than for Method B. The variability approached the lowest for all products at the higher frequencies.

3. Between-laboratory variability

The greatest difference in standard deviations between Methods A and B occurred for the between-laboratory variability (see Table X). Whereas the Method-A standard deviations were smaller than those for Method-B standard deviations for within- and between-subjects, the between-laboratory variability was greater for Method A than for

TABLE X. Between-laboratory standard deviations in dB for Methods A and B for each protector.

Protector	Method	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	*A*-weight Atten
TacticalPro	A	1.8	1.4	2.8	2.4	2.2	2.8	2.4	1.8
	B	2.1	1.6	2.6	2.2	1.9	2.0	2.1	1.5
3M 1427	A	3.1	2.2	2.7	2.3	2.1	3.5	2.6	2.4
	B	2.0	1.4	1.6	1.9	2.5	2.8	2.4	1.3
dB Blocker	A	1.4	1.3	1.8	2.5	2.9	4.1	3.3	2.2
	B	1.7	2.0	1.6	2.6	2.6	3.6	3.4	2.0
JazzBand	A	3.3	3.1	2.2	2.3	2.6	2.9	3.9	2.0
	B	2.7	2.6	2.2	2.7	2.5	2.4	2.2	2.1
Classic	A	4.8	5.1	6.1	6.3	3.3	3.6	4.2	5.0
	B	1.3	1.2	2.0	2.8	2.1	2.9	3.6	2.0
AirSoft	A	3.9	3.7	4.4	3.3	2.5	3.5	3.7	3.2
	B	2.3	2.7	2.9	2.8	2.8	4.0	4.1	2.6

TABLE XI. Calculated numbers of subjects necessary to achieve 6 dB resolution in the *A*-weighted attenuation from between-subject and between-laboratory variances, for Methods A and B. (Results are shown with one decimal place for illustrative purposes.)

Method	Variance	TacticalPro	3M 1427	dB Blocker	JazzBand	Classic	AirSoft
Method A	Between subject	2.5	6.6	7.3	5.2	5.1	11.1
Method B	Between subject	4.3	7.0	15.3	9.3	7.8	18.5
Method A	Between lab	11.7	26.7	19.0	19.0	131.1	52.2
Method B	Between lab	9.2	8.2	16.7	20.2	17.6	27.5

Method B at most frequencies. The Classic, AirSoft, and 1427 devices had greater standard deviations for Method A than Method B at 1000 Hz and below. The Classic had the greatest standard deviations of all the protectors tested. The standard deviations of the *A*-weighted attenuation for these three devices differed across methods by as little as 0.6 dB to as much as 3 dB. In general, the Method-B standard deviations for earplugs increased with frequency, while the standard deviations for earmuffs and canal-caps did not vary appreciably with frequency. For the Classic and AirSoft earplugs, Method-A standard deviations are less at the higher frequencies relative to the lower frequencies. For the dB Blocker, the Method-A standard deviations increased with frequency.

The standard deviations between Methods A and B for the Tactical Pro, JazzBand, and dB Blocker tended to differ by less than 0.5 dB at most frequencies. The standard deviations for the *A*-weighted attenuations were between 0.1 and 0.3 dB for these three devices. The small differences between the standard deviations for the *A*-weighted attenuations suggest that the low frequencies (125–500 Hz) and the highest frequency (8000 Hz) do not contribute significantly to the overall variance for these protectors. The *A*-weighting applied in the computation has de-emphasized the contribution from the lower and highest frequencies.

D. Number of subjects necessary for a desired resolution

The concept of resolution comes from the ability to distinguish between the central tendency of two distributions. In astronomy, the light from two stars can be resolved only if there is sufficient angular separation between the respective images in the telescope. Noise inherent in the image, either due to atmospheric effects or distortions of the equipment, can increase the apparent size of an object. In the analogous case for hearing protectors, the "image" is the distribution of attenuations. The separation of the central modes, the width of the distributions, and the choices for power and confidence level affect the resolution. In essence, the resolution determines the ability to distinguish between the central tendencies of the attenuation distributions in different tests. Because resolution is inextricably linked to the width of the distribution, a wider distribution will require either a greater separation or more subjects to increase the statistical power of the measurement to permit resolving differences between two sets of data. Resolution is dependent on the power and confidence level. If one chooses a low power (0.8) and confidence level (0.84), the resolution will be lower than if a higher power (0.9) and confidence level (0.99) are chosen. The choice of the confidence level and resolution implies that the observed decibel difference in the means of the respective attenuation distributions can be distinguished with 84% or 99% confidence.

In Murphy *et al.* (2004), the number of subjects was estimated using a 6-dB resolution from ANSI S12.6-1997 (R2002) (2002) and the greatest between-subject standard deviation, which typically occurred at 8000 Hz. The frequency with the greatest variance (largest standard deviation) was assumed to dominate the variance of any rating. Since the present analysis examines the repeatability and reproducibility of an *A*-weighted attenuation, the frequencies having the greatest contribution to the *A*-weighted protected and unprotected sound pressure levels (1, 2, and 4 kHz) dominate the variance. This effect is evident in Table XI and Fig. 3 where the numbers of subjects necessary to achieve a 6 dB resolution (i.e., the ability to detect differences between two means of 6 dB) are plotted. From Murphy *et al.* (2004), the required sample size is

$$N = n_s\left(\sqrt{2}(\text{probit}(1-\alpha) + \text{probit}(1-\beta))\frac{\sigma}{R}\right)^2 = n_s\left(\frac{2.5966\sigma}{R}\right)^2, \qquad (7)$$

where n_s is the number of subjects tested, σ is the standard deviation for repeatability or reproducibility as estimated by Eqs. (5) and (6), and R is the desired resolution (6 dB). The

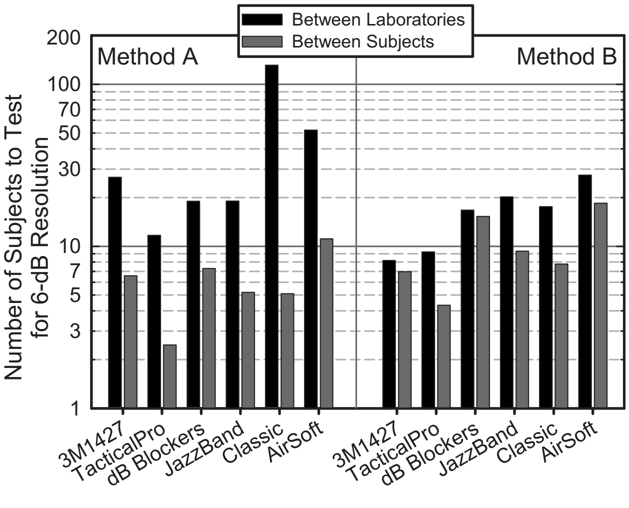


FIG. 3. The power estimates for the number of subjects necessary to achieve a 6 dB resolution for the *A*-weighted attenuations of each protector Methods A and B.

probit function provides the appropriate percentile value from a standard normal distribution for the confidence level, $1-\alpha=0.84$ and power, $1-\beta=0.8$ (see Murphy *et al.* 2004 for further explanation of this approach).

For the Method-A between-subject repeatability (essentially testing the same subject population in the same laboratory), the estimated sample sizes were less than 12 subjects needed to separate the *A*-weighted attenuation results regardless of the protector being tested. The TacticalPro may require fewer than 3 subjects while the AirSoft needs about 11 subjects to achieve a 6-dB resolution.

For the Method-B between-subject repeatability, the AirSoft, JazzBand, and dB Blocker products increased the most. Typically, the estimated sample size was between 10 and 30 subjects if the standard deviations for individual frequencies were used. If the standard deviations based on *A*-weighted attenuations are used, the number of subjects necessary to achieve 6-dB resolution was between 4 and 19 subjects. For the Classic, Tactical Pro, and 3M 1427, 4–8 subjects were required.

The reproducibility between laboratories is considerably poorer than the between-subject repeatability. One should note that the between-laboratory reproducibility includes the between-subject variability and therefore must be greater. From the sample size calculations, Method-B testing would require 9–28 subjects for all products (see Table XI). For the Method-A data, the numbers of subjects required for earmuffs, dB Blocker, and JazzBand were less than 30. However, for the AirSoft and Classic earplugs, 53 and 132 subjects were necessary to achieve a 6-dB resolution, respectively.

The effect of the *A*-weighting can be seen in the power calculation. Note that the between-subject and between-laboratory standard deviations for the *A*-weighted attenuation were more closely correlated with the individual frequency results around 2000 Hz. Particularly the Method-A and Method-B data for the dB Blocker exhibited a minimum at 2000 Hz which correlated closely with the *A*-weighted power estimate. For the AirSoft, the minimum was at 2000 Hz, and the *A*-weighted power was greater than the 2000 Hz value. This suggests that the *A*-weighted power estimate and its associated standard deviation were dominated by those bands (1000, 2000, and 4000 Hz) which have the strongest contribution to the *A*-weighted energy.

IV. DISCUSSION

A. Procedural differences

The testing was completed in August 2006 and the results were initially analyzed in October and November of 2006. ANSI Working Group 11 held three meetings in late 2006 and early months of 2007 to discuss the results and determine modifications to the ANSI S12.6 standard. Laboratory representatives, manufacturers, EPA, and government representatives participated in the deliberations which revealed differences in the implementation of Method-A tests by the six participating laboratories. Particularly, the insertion of the foam earplug was identified as the contributing factor related to different laboratory-to-laboratory experimenter influence under Method A. Requiring subjects to achieve a minimum of 75% insertion is allowed under experimenter-supervised tests. From earlier in-house tests of the E•A•R® Classic®, the experimenter and the laboratory manager at AFRL determined that this depth was necessary to protect personnel in high noise environments (Hall *et al.*, 2005). The fact that AFRL achieved the greatest attenuation of all the laboratories highlights the ability of the product to achieve maximum attenuation with deeper insertion.

The AirSoft also benefited from increased insertion depth. One might be tempted to attribute the highest Method-A attenuations at the Howard Leight Laboratory to its ability to use the product; the effect was not due to subject selection since the same initially naïve subjects were used for both Method-A and Method-B tests. The manufacturer bias was not evident in the test results for the Classic tested at E•A•RCAL. Thus, an appeal to manufacturer bias cannot be substantiated. If the outlier subjects (those who clearly had the least attenuation) for the other laboratories were removed from the distribution, the mean would be increased and standard deviations reduced, yielding better agreement with the results from the Howard Leight Laboratory. Rather, the improved performance at the Howard Leight Laboratory was likely a function of the poorly performing naïve subjects being better motivated and more conscientious of the insertion process during Method-A testing.

B. Fitting of protector types

In an effort to improve the uniformity of results collected with the different protocols, various schemes were investigated to remove outlier subjects. However, the scrutiny of the working group did not yield adequate methods to algorithmically remove outlier data. Instead, the outliers must remain in the data sets and explanations for the aberrant results must be sought.

First, the type of protector must be considered. Earmuffs comprise one category; the custom protectors and semi-aural inserts (i.e., canal-caps), although seemingly different, comprise another. Insert earplugs can be separated into two categories. Essentially, all existing protectors can be fit into these categories regardless of the electronic enhancements or other features of the protector.

Earmuffs must make a seal with the skull around the ear. So long as hair, jewelry, and head shape do not interfere with the ability of the ear cushion to seal against the skull, the protector's attenuation is governed by the volume and mass of the earcup, the compliance of the cushion, the ability to conform to an irregular shape, and the headband clamping force. Differences in these characteristics influence the overall attenuation and possibly have small effects on the variability of the attenuations across subjects. For example, the Peltor Tactical Pro has a slightly more compliant cushion than the 3M 1427 muff. As well, the headbands for the two muffs were worn in different positions (the 3M 1427 was worn behind the head, and the Peltor was worn over the head). However, these characteristics are largely independent of the user, and therefore the differences between the trained and untrained tests were minimal.

The second category, canal-caps and custom products, has a similarity in that once the canal is sealed, little additional attenuation can be achieved (i.e. deeper insertion is not possible). The canal-cap seals the entrance of the ear canal, and the custom protector is designed to extend into the canal. If a custom product is manufactured from a deeper impression, then an increased attenuation will be likely (Hall *et al.*, 2005). Assuming the dB Blockers were manufactured to the same tolerances, the nominal length of the ear canal portion should provide about the same amount of attenuation. For those subjects already achieving good attenuations, the additional experimenter instruction (Method A) did not produce an appreciable improvement compared to the attenuations measured with Method B (see Fig. 1, 10% whisker). For the lower attenuations obtained with Method B, the poorly-sealed ear canals were able to be sealed once the subjects were instructed by the experimenter.

Some banded devices are designed to have the earplug inserted into the ear canal. In this case, attenuation testing results would be expected to resemble those of earplugs rather than canal-caps. If products have a limitation on how far they might be inserted or how the seal with the ear canal is created, then training will improve the poorly-fit protectors but little improvement can be expected for the well-fit protectors.

A third category of protector, premolded earplugs, typically achieves a seal through pressing a single flange or multiple flanges against the canal walls. The seal created by the flanges is the main determiner of the product's attenuation rather than transmission through the body of the earplug. If the flanges fail to adequately seal to the ear canal walls, then the dominant transmission path will be through the leak(s) around the flange. Once the flanges have achieved a patent seal, insertion of the product further into the ear canal will do little to improve or increase attenuation.

The AirSoft product tested in this study is not the same product currently on the market. The product tested in the study had an unbaffled, air-filled bladder within the body of the plug; the current product has been redesigned to incorporate internal baffles. The results may generalize to other similar premolded earplugs. In particular, the previous interlaboratory study examined the V-51R and EP100 earplugs (Royster *et al.*, 1996). The maximum attenuations published by Murphy *et al.* (2004) did not change appreciably between the informed-user and naïve subject-fit tests. Thus, users of premolded plugs in the lower quartile of the study will benefit most from training and instruction.

A fourth category of protector, formable earplugs, creates a seal between the lateral surface of the earplug and the ear canal walls. For the Classic®, the attenuations for both the lowest and highest quartiles improved significantly between Method-A and Method-B tests (see Fig. 1). The improvements for the subjects in the highest quartiles indicate the training was effective and necessary to achieve better attenuations from a foam roll-down earplug. As more of the lateral surface of the plug makes contact with the ear canal wall, the attenuation increases. This finding has been observed in the previous interlaboratory study (Murphy *et al.*, 2004) and explicitly tested in a study conducted by the Air Force (Hall *et al.*, 2005).

During the working group deliberations, the test methods used at AFRL were found to require the subjects insert at least 75% of the Classic into the ear canal. The tester marked the protectors at approximately three-fourths of the length and watched the subject during insertion to ensure that the mark was obscured after insertion. Subjects were required to refit the plug if the mark was visible to the experimenter. While the methods employed by AFRL are permitted under the Method-A protocol, they do not reflect the typical use of the product by a trained or untrained user. The user would need a mirror to examine the insertion depth and a second mirror to inspect whether the mark had been obscured. Furthermore, most users will not ask a co-worker to inspect the fit or insertion depth of the earplug. Although the AFRL method was effective in achieving greater attenuation, it does not reflect typical or practical use.

One key result of the current study was the realization that greater consistency of experimenter instruction of subjects would improve the between-laboratory reproducibility of Method A. The working group has invested substantial time in revising ANSI S12.6 in an attempt to accomplish this, changing Method A from an an experimenter-supervised fit to a trained-subject fit protocol, removing the influence of the experimenter on the fitting of the device during the actual attenuation measurement.

C. Application to regulatory issues

Under the OSHA Hearing Conservation Amendment (OSHA, 1983), workers are enrolled in a hearing conservation program when noise exposures equal or exceed an 8 h time-weighted average sound level of 85 dB(A). Training in the use of hearing protection must be provided, and workers are required to be refitted and retrained whenever they suffer a standard threshold shift and provided with hearing protectors offering greater attenuation if necessary. While the quality of training may vary greatly across companies, those workers that have received it would no longer qualify as naïve subjects. The majority of hearing protection is sold to industrial hearing conservation programs and not to the average consumer (Frost and Sullivan, 2005). This alone suggests that hearing protectors are being used by trained workers and not untrained, uninformed consumers. Method-A data should be more representative of the typical user than Method-B data, yet Berger *et al.* (1998) provides contradictory evidence suggesting that Method B is the more appropriate technique.

In Table XII, several advantages and disadvantages for Methods A and B are presented side-by-side. One driving force is the creation of a reproducible test method. Widely varying product test results with the same subjects are unacceptable. If the EPA were to audit a manufacturer's product, then reproducibility within a laboratory is paramount. This study found better between-subject reproducibility with Method A yet better between-laboratory reproducibility with

TABLE XII. Comparison of advantages and disadvantages between Methods A and B.

Topic	Method A	Method B
Between-subject reproducibility	X	
Between-laboratory reporducibility		X
Testing cost	X	
Testing expediency	X	
Real-world applicability		X
Rank-ordering of result		X
Necessity of derating NRR		X
Explainability of test results	X	
Inherent device performance	X	
Dual number rating	?	?
International applicability	?	?

Method B, though the latter apparent advantage of Method B is helped by the inherently greater between-subject variance that results from using naive subjects.

Testing cost and expediency of testing a product either in an in-house laboratory or an independent laboratory are important considerations for the EPA. The cost and speed of testing will be lower with Method A since the time to recruit, qualify, and maintain a Method-B panel constitutes the initial visit of the subject to the testing laboratory. The six laboratories in this study had varying success in getting subjects to complete the entire series of tests. In some cases, a subject chose to exit the study with only one more Method-A test to complete. A new subject had to be qualified, trained in the task, and run through all of the protector conditions to replace the subject who left. If experienced subjects were permitted, then the subject could have been replaced from the pool of previously qualified subjects.

The EPA should specify a policy regarding the reuse of subjects when testing a product for labeling purposes. Under Method B, the working group has determined that,

> "Once a subject has been accepted in an inexperienced-subject fit evaluation in a given facility, s/he may participate for a lifetime maximum of 30 separate inexperienced-subject fit tests, each test consisting of 2 trials. Of those 30 tests, the total number permissible for earplugs and semi-inserts, or both, shall not exceed 12, and there shall not be more than 4 tests on any one of the following categories: foam, premolded, malleable, semi-insert, and other earplugs, and no more than one test on a custom-molded plug. Subjects shall be excluded from any further inexperienced-subject fit testing of earplugs or semi-inserts once they have viewed video or computer-based fitting instructions during a product test." (ANSI S12.6-2008, 2008).

For Method A, the same subjects may be reused many times for testing products. As subjects develop expertise with the testing paradigm and use of the products under test, they would be expected to provide more consistent test results. With respect to subsequent audit tests or mandated retesting, the reuse of the subject is a topic for debate. If EPA mandates periodic retesting, the lack of any overlap of the testing panel from the initial rating to the retest ensures that the results are statistically independent. Since the period for retesting would likely be more than 2 years, retention of the entire subject panel over that duration is unlikely. An additional argument in favor of nonoverlapping panels would be to increase the numbers of subjects on which the product has been tested. In Murphy *et al.* (2004), the statistical power of the data set increased more by adding subjects than by performing repeated measurements on the same subjects.

With respect to real-world applicability, the Method-A results are expected to be similar to the European SNR [in accordance with ISO 4869-1 (1990) and ISO 4869-2 (1994)] as the subjects would be experienced users of protectors, while workers that regularly use HPDs are undoubtedly experienced, many wear the protectors only to comply with company policy as mandated by federal and state regulations. Those workers achieving inadequate levels of protection (poorly-fit) are at increased risk of developing noise-induced hearing loss. Method-B data have been demonstrated to provide a better correlation with real-world attenuations than the current experimenter-fit NRR or attenuations based on ISO 4869-1 (1990) experimenter-supervised fit data (Gauger and Berger, 2004). Knowledge of how a protector is likely to be worn is useful in predicting rates of hearing loss. In addition to being more applicable to real-world performance, the Method-B data would not need to be derated to assess the likely performance in the real world.

Method-A data are more useful in explaining the performance of a device. While we have not focused on the attenuation by frequency for the various products, the improved low-frequency performance of the Classic® can be better understood when the results are more consistent across subjects. If a manufacturer designs a protector to achieve maximum attenuation for a predominantly low-frequency, high-noise environment, then using an untrained naïve subject is inappropriate. The quality of the seal becomes more critical as the low-frequency attenuation is decreased dramatically by the presence of small leaks in the seal of a protector. The workers in these environments should be specifically trained in the use of the personal protective equipment and ought to be required to demonstrate adequacy of protection through fit-testing of the protectors. Thus the Method-A data should be able to better assess the inherent performance that a HPD is capable of providing.

The last two elements in the table are indeterminate at the present time. ANSI has recently published a method to calculate the effective A-weighted sound pressure level when a hearing protector is worn (ANSI S12.68, 2007). In this new standard, the use of two numbers to describe the attenuation that "most users can achieve" and that which "motivated users might be able to achieve" is novel. Essentially the rating describes the variation about the mean, recommending the use of ±0.84 standard deviations corresponding to the 80th and 20th percentiles. When the working group developed the standard and deliberated the results from this study, the initial intent was to create a two-number rating that was applicable to Method-B data. However, the two-number rating is equally applicable to the Method-A data. Instead of a motivated user, the user may be a trained, expert user. The language associated with the higher and lower ratings is nuanced; regardless, one can learn from the spread of the two values, with a smaller spread indicating less variation across users and noise spectra. Thus, the occupational hearing conservationist will have another means to judge the performance of a particular protector.

V. CONCLUSIONS

Whether the Method-A or Method-B data are more similar to international standards and ratings is open for consideration. Currently Brazil, Canada, and Australia/New Zealand have adopted the essential elements of a Method-B procedure from the ANSI S12.6-1997 (2002) standard into occupational hearing conservation standards and directives. Almost identical to Method B, ISO has developed a method using naïve test subjects (ISO 4869-5, 2006). The European Union has adopted the ISO 4869-1 and 4869-2 standards for HPD testing and rating (ISO 4869-1, 1990; ISO 4869-2, 1994). The ISO 4869-1 standard is similar to the Method-A approach. Until tests conducted under Method-A conditions are available, the similarity of attenuations with those of ISO 4869-1 is unknown. Since both standards utilize trained subjects, the attenuations should be similar.

The U.S. EPA is anticipated to propose a revision to the hearing protector labeling regulation, 40 CFR 211 Subpart B. The Method-A test protocol is expected to be the proposed method for assessing the performance of passive hearing protectors. At the heart of the Noise Control Act is the need *to provide accurate and understandable information to product purchasers and users regarding the acoustic properties of designated products so that meaningful comparisons with respect to noise emission or noise reduction can be made as a part of a product purchase or use decision* (EPA, 1972). Method B can more nearly represent the anticipated protection of uninformed users. Once a user has experienced the higher attenuation and tactile sensation of a well-fit earplug and has learned techniques to check the fit, the assumption of performance typified by naïve users is potentially erroneous. Furthermore, the question remains as to when an uninformed or naïve user becomes an experienced user with respect to repeated fittings of earplugs. While the attenuation achieved by uninformed, naïve users is important for estimating the risk of noise-induced hearing loss of a group of persons exposed to a given noise, the purpose of the label is to provide the manufacturer a means to inform the user of the acoustic properties of the product. Underestimating the NRR by applying a Method-B protocol could present a disservice to the public and could unfairly disadvantage manufacturers who create high attenuation protectors.

This research, the participating laboratories, and the ANSI S12 Working Group 11 have provided invaluable assistance to the EPA in the preparation of a revised regulation. The original promulgation of 40 CFR 211 Subpart B occurred in 1979. Changes to the fundamental protocol for rating hearing protector performance are now being contemplated in a notice of proposed rule making by the EPA. Future changes may not occur again for several decades. Thus the effort to quantify the differences between ANSI S12.6 Method A and Method B has led to revisions of the test standard that will affect the policy of the United States for years to come.

ACKNOWLEDGMENTS

Portions of this work were supported by the U.S. EPA interagency Agreement No. 75921973-01-0. Each participating laboratory funded the subject payments from their respective internal budgets. Each laboratory had employees without whose assistance and dedication this study would not have been able to be completed: Paul Schley, AFRL; Ron Keiper, E•A•RCAL; Jesse Duran, Howard Leight; Racheal de Favacorte, LARI; Pamela Graydon, Dayna Richards and Benson Davis, NIOSH; Adrian Houtsma, Melinda Hill and Kimberly Vasquez, USAARL. NIOSH also wishes to note the contributions of Christa Themann and Edward Zechmann for their thorough review of the manuscript. As well, the authors wish to thank Lee Hager, Torben Poulsen, and Rickie Davis for their thoughtful comments during the NIOSH peer-review process.

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not represent any official policy of the Centers for Disease Control and Prevention, The National Institute for Occupational Safety and Health, the Environmental Protection Agency, the U.S. Air Force or the U.S. Army. Mention of company names and products does not constitute endorsement by CDC, NIOSH, EPA, U.S. Air Force or the U.S. Army.

ANSI S12.6-1997 (R2002) (**2002**). American National Standard for the Measuring Real-Ear Attenuation of Hearing Protectors, American National Standards Institute, New York.

ANSI S12.68 (**2007**). American National Standard Methods of Estimating Effective *A*-Weighted Sound Pressure Levels When Hearing Protectors, are Worn, American National Standards Institute, New York.

ANSI S12.6-2008 (**2008**). American National Standard for the Measuring Real-Ear Attenuation of Hearing Protectors, American National Standards Institute, New York.

Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (**1998**). "Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part III. The validity of using subject-fit data," J. Acoust. Soc. Am. **102**, 665–672.

EPA (**1972**). P. L. 92-574 Noise Control Act of 1972, 86 Stat. 1234, U.S. Environmental Protection Agency, October 27.

EPA (**1978**). CFR 40 211B Hearing Protective Devices. U.S. Environmental Protection Agency, September 29.

Frost and Sullivan (**2005**). "U.S. markets for industrial hearing protection products," Technical Report, December 13 (Frost and Sullivan, San Antonio).

Gauger, D., and Berger, E. H. (**2004**). "A new hearing protector rating: The noise reduction statistic for use with *A*-weighting (NRS_A)," Technical Report No. E-A-R 04-01/HP, American National Standards Institute, New York. This report can be found online at the following URL web URL http://www.e-a-r.com/pdf/hearingcons/TO4_01EPA.pdf.

Hall, J., Mobley, F., McKinley, R., and Schley, P. (**2005**). "New directions for custom earplugs," in 2005 Conference and Exposition on Noise Control Engineering Rio de Janeiro, Brazil, p. 89.

ISO 4869-1 (**1990**). "Acoustics—Hearing Protectors Part 1: Subjective method for the measurement of sound attenuation," International Organization for Standardization, Geneva.

ISO 4869-2 (**1994**). "Acoustics—Hearing Protectors Part 2: Estimation of effective *A*-weighted sound pressure levels when hearing protectors are worn," International Organization for Standardization, Geneva.

ISO 4869-5 (**2006**). "Acoustics—Hearing Protectors Part 5: Method for estimation of noise reduction using fitting by inexperienced test subjects," International Organization for Standardization, Geneva.

Johnson, D. L., and Nixon, C. W. (**1974**). "Simplified methods for estimating hearing protector performance," Sound Vib. **8**, 20–27.

Kroes, P., Fleming, R., and Lempert, B. (**1975**). "List of personal hearing protectors and attenuation data," Technical Report Publication No. 76-120, U.S. Department of Health Education and Welfare, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati.

Murphy, W. J., Franks, J. R., and Krieg, E. F. (**2002**). "Hearing protector attenuation: Models of attenuation distributions," J. Acoust. Soc. Am. **111**, 2109–2116.

Murphy, W. J., Franks, J. R., Berger, E. H., Behar, A., Casali, J. G., Dixon Ernst, C., Krieg, E. F., Mozo, B. T., Ohlin, D. W., Royster, J. D., Royster, L. H., Simon, S. D., and Stephenson, C. (**2004**). "Development of a new standard laboratory protocol for estimation of the field attenuation of hearing protection devices: Sample size necessary to provide acceptable reproducibility," J. Acoust. Soc. Am. **115**, 311–323.

Netter, J., Wasserman, W., and Kutner, M. H. (**1990**). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, 3rd ed., Richard D. Irwin Inc., Boston, pp. 970–1001.

OSHA (**1983**). CPL 2-2.35A-29 CFR 1910.95(b)(1) Guidelines for noise enforcement: Appendix A, U.S. Department of Labor, Occupational Safety and Health Administration, December 19.

Poulsen, T., and Hagerman, B. (**2004**). "A Nordic round robin test on hearing protectors. The influence of the sound field on measured REAT attenuation," Acta Acust. **90**, 838–846.

Royster, J. D., Berger, E. H., Merry, C. J., Nixon, C. W., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Mozo, B. T., Ohlin, D., and Royster, L. H. (**1996**). "Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part I. Research of Working Group 11, Accredited Standards Committee S12, Noise," J. Acoust. Soc. Am. **99**, 1506–1526.

SAS (**2007**). SAS/STAT Software, 9.0.1 ed., SAS Institute Inc., Carey, NC.

How to stretch and shrink vowel systems: Results from a vowel normalization procedure

Christian Geng
Linguistics and English Language, The University of Edinburgh, Midlothian EH8 9AD, United Kingdom

Christine Mooshammer
Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511

(Received 17 May 2007; revised 20 February 2009; accepted 25 February 2009)

One of the goals of phonetic investigations is to find strategies for vowel production independent of speaker-specific vocal-tract anatomies and individual biomechanical properties. In this study techniques for speaker normalization that are derived from Procrustes methods were applied to acoustic and articulatory data. More precisely, data consist of the first two formants and EMMA fleshpoint markers of stressed and unstressed vowels of German from seven speakers in the consonantal context /t/. Main results indicate that (a) for the articulatory data, the normalization can be related to anatomical properties (palate shapes), (b) the recovery of phonemic identity is of comparable quality for acoustic and articulatory data, (c) the procedure outperforms the Lobanov transform in the acoustic domain in terms of phoneme recovery, and (d) this advantage comes at the cost of partly also changing ellipse orientations, which is in accordance with the formulation of the algorithms. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106130]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt [AL] Pages: 3278–3288

I. INTRODUCTION

A. Background

One of the major challenges in experimental phonetics is to overcome the consequences of speaker-specific variability because individual differences obscure the distinction between categories in acoustic and articulatory spaces. Johnson *et al.* (1993), for example, tested the hypothesis that speakers use the same set of articulatory features for the production of the American English vowel system by analyzing tongue contours and jaw movements. They concluded that speakers were very consistent within themselves in the strategies applied for producing different vowels. Between speakers, however, there was a great amount of variability in the way they increased speech tempo, distinguished between tense and lax vowels and also in their overall strategies. This led the authors to the conclusion that the targets of speech production must be specified in terms of the acoustic output. However, as was discussed in Disner (1980), variability in the formant space still reflects speaker-dependent differences due to vocal-tract shapes and sizes, which makes it impossible to compare vowel inventories of different languages by means of formant frequencies taken from natural utterances of human speakers.

The major aim of the current study is to test the usefulness of a normalization procedure heavily inspired by generalized Procrustes analysis (Gower, 1975). Specifically, we applied this normalization procedure to acoustic, articulatory, and anatomic data in order to reveal speaker-independent strategies for the production of German vowels. The motivation for speaker normalization has also been expressed with a stronger conceptual and theoretical bias in the aforementioned paper by Johnson *et al.* (1993). Their discussion of what they term as the "universal articulatory phonetics hypothesis" puts in question the classical tenet that linguistic equivalence classes should be built on the basis of articulatory substance unless it is lawfully possible to relate different speakers' articulatory performance in a systematic way such that the constitution of traditional units such as the phoneme becomes possible. Johnson *et al.* (1993) worked through a whole catalog of influences, which have the potential of putting the universal articulatory phonetics hypothesis in danger: measured fleshpoint, speaking rate, dialect, palate shape, dental occlusion, and articulatory strategy. In fact, in their study, systematic individual differences were retained, making the articulatory definition of equivalence classes a problematic undertaking. There are several ways out of this dilemma: one solution would propose to locate invariance in speech perception (Kingston and Diehl, 1994) or the sensory periphery (Guenther *et al.*, 1998), which assumes that producing phoneme sequences involves the activation of invariant auditory goals. Alternatively, a more practical approach is the use of normalization procedures. Normalization procedures have been successfully applied to articulatory data, e.g., in Harshman *et al.* (1977), Hashi and Westbury (1998), and Beckman *et al.* (1995) or to acoustical data [for reviews see Disner (1980) and Johnson (2005)]. The goals of such procedures can be to abstract from speaker-specific vowel locations and variation patterns in order to compare linguistically different vowel systems in terms of their formant (e.g., Clopper *et al.* 2005) or articulatory spaces (e.g., Jackson, 1988). Within a given language, vowel normalization can be applied to tongue configurations in order to find speaker-independent strategies for speech tempo variation (Hoole, 1999) and for the lexical stress distinction (Geng and Mooshammer, 2000) or to test against predictions made by quantal theory (Beckman *et al.*, 1995).

Traditionally, acoustical vowel normalization methods have been divided into "intrinsic" and "extrinsic" methods for vowel normalization, a division which dates back to Joos (1948). Adank *et al.* (2004) argued though that this division might not be fine-grained enough to describe all essential differences between procedures. They extended the usage of the extrinsic vs intrinsic dichotomy to both (a) vowels and (b) formants. This work does not consider vowel intrinsic normalization because it is not clear how it could be related to parallel articulatory data. The same holds for all vowel-extrinsic/formant-extrinsic schemes known to us. The only class containing published normalizations, which correlate well with the aims of the current work, are formant-intrinsic, vowel-extrinsic normalizations. According to Adank *et al.* (2004), this kind of normalization has been the most efficient. Within this class, the Lobanov (1971) transform has been described as one of the most efficient procedures. Still, this kind of normalization is not without drawbacks: As pointed out by Nearey (1989), despite its success for practical reasons, i.e., the most substantial scatter reduction, its cognitive plausibility is questionable because the listener would have to know the formant frequencies of the complete vowel system spoken by this particular speaker in order to recognize a single vowel. This points to the discussion about the psychological reality of entities such as mean formant frequencies and scale factors. In contrast to Nearey, Adank *et al.* (2004, p. 3105) viewed such factors as possibly accounting for listeners' life-long experiences with listening to different types of speakers. Another criticism was noted by Apostol *et al.* (2004), according to which statistical methods cannot be directly related to anatomical differences between speakers, which ultimately underlie the large variation in the formant spaces. The basic hypothesis of their model is that inter-speaker variability in formant spaces "arises from differences among speakers in the respective lengths of their back and front vocal-tract cavities" (p. 337) because of the formant-cavity affiliation.

In a similar vein, namely, to relate articulatory positions to formant frequencies, a number of geometrical normalization procedures has been applied to tongue contours derived during vowel production. These procedures are based on re-expressing fleshpoints on the tongue—usually acquired by the x-ray microbeam system (Westbury, 1994)—as distances to the palate (e.g., Beckman *et al.*, 1995; Hashi and Westbury, 1998; Perkell and Nelson, 1985). Therefore, the pellet positions are translated into a palate-based coordinate system in order to minimize the effects of differences in vocal-tract size and shape on mean articulatory postures. The advantage is that the new palate-based coordinate system is more closely related to the oral part of the area function than the original coordinate space and can therefore more easily be related to spectral properties. Within this framework, as well as for the cavity-affiliation model by Apostol *et al.* (2004), the major aim is to explain and reduce the speaker-dependent acoustical variability by anatomical differences, e.g., due to gender (see Simpson, 2002).

A very different approach is the factor analytic treatment of vowel production in which the underlying control mechanisms from the highly correlated coordinates of the tongue-pellets are extracted. In their work, Harshman *et al.* (1977) subjected multi-speaker x-ray tongue contours to the PARAFAC algorithm, yielding two factors consisting of three matrices: (a) the speaker weights, (b) a speaker-independent vowel space, and (c) the so-called articulator weights. A consistent and robust interpretation of the application of PARAFAC arises from these first two factors extracted from flesh-point data as well as from contours and for a number of different languages (see, e.g., Harshman *et al.*, 1977; Hoole, 1999; Jackson, 1988). The first factor, usually dubbed front raising, distinguishes low vowels from high front vowels. In articulatory terms, front raising is a forward movement of the root of the tongue and an upward movement of the front of the tongue. The second factor, back raising, is associated with the formation of back vowels and characterized by an upward and backward movement of the tongue. The current study will make an attempt to evaluate a technique conforming to that of Harshman *et al.* in its applicability to both acoustic and articulatory data but differing in its computational procedure as well as its general orientation. The normalization procedure applied here is based on a method frequently applied in morphometrics/zoology in order to solve the problem of the superimposition of geometrical landmarks. In zoology, this is often helpful for shape comparison between species abstracting from uninformative scaling, translation, and rotations. These methods are often termed "Procrustes" methods (see Gower, 1975, and with a special background in morphometrics Rohlf and Slice, 1990 or Goodall and Green, 1986). Due to its central importance for the current work, this approach will be described in greater detail in the next section.

B. The normalization procedure: A modified generalized Procrustes analysis

These superimposition techniques can be distinguished according to the following aspects: (a) The number of objects to be aligned: Procrustes analysis has originally been designed for the alignment of only two specimens but has later been extended to handle any number of objects in Gower (1975), (b) the nature of the transformation terms to be applied: If only rigid rotations are allowed, orthogonal methods—i.e. Procrustes methods in a more narrow sense—are used, which preserve the angles between data points. If uniform affine deformation is to be applied, the class of methods are called oblique and (c) the optimization strategy applied: If no local shape change is allowed, least-squares fitting methods are appropriate; if one wishes to account for local shape change, more advanced, non-parametric methods based on the median have to be applied. The acoustic and articulatory data corpora in our study contain more than two speakers, such that our algorithm a priori has to refer to the "generalized" case. With respect to the nature of the rotation terms to be applied, a distinction has to be made between articulatory and acoustic data sets. While for the articulatory data, a considerable amount of affine deformation would already be expected due to different vocal-tract morphologies, this is not so clear for the acoustic data set. Goodall and Green (1986) devised a method for checking the amount of affine deformation necessary to superimpose landmarks in the two-dimensional (2D) case. In cases in which the amount of affine deformation necessary for superimposition is not substantial, superimposition applying orthogonal transformations yields similar results. Concerning the optimization scheme, we used the least-squares technique and not the nonparametric method described in Rohlf and Slice (1990) because we were aiming at uniform rather than local shape change. The approach used here comprises two separate steps: First, across different speakers, a "consensus configuration" representing an average subject is calculated, and, in a second step, this consensus is fitted to the data of individual speakers to result in normalized data.

1. Construction of the consensus object

The consensus configuration is calculated as follows [equivalent to the formulation in Rohlf and Slice (1990)]: First, the data of the n individual speakers are centered and scaled with a standard z-transform. Then, a first version of the consensus object is calculated using

$$A = \frac{1}{n}\sum_{i}^{n} X_i(X_i^T X_i)^{-1} X_i^T, \tag{1}$$

where the X_i are the individual speakers' data matrices after centering and scaling, n is the number of speakers, and superscript T denotes the transpose of a matrix. This version of the consensus object does not resemble the original objects though. In the bivariate case—like in the analysis of formant spaces—the $X(X^T X)^{-1} X^T$ operation transforms each object "so that the variance in the bivariate distribution of landmarks is the same in all directions in the plane for each object" (Rohlf and Slice, 1990, p. 49). The final consensus configuration is calculated as

$$C = A\left(\frac{1}{n}\sum_{i} X_i X_i^T\right)A. \tag{2}$$

Thereafter, C is subjected to a singular-value decomposition, and the final consensus configuration is a matrix of eigenvectors of C subjected to truncation; e.g., for planar configurations, the first two columns are taken (Rohlf and Slice, 1990, p. 49).

2. Calculation of normalized data

The second step consists of calculating reconstructed data for each subject's configuration. These are calculated by post-multiplying the consensus object with a transformation matrix, which in general is calculated as

$$H^* = (X_2'^T X_2')^{-1} X_2'^T X_1' \tag{3}$$

for two objects in the oblique case. X_2 here is the consensus configuration as calculated above, and X_1 is an arbitrary speaker's original configuration. This is equivalent to the equation for the least-squares estimates of partial regression coefficients in multivariate multiple regression. In the case of orthogonal rotation, the rotation matrices are calculated by performing a singular-value decomposition of the product of the object matrices to be superimposed:

$$H = VSU^T, \tag{4}$$

where U and V are such that $X_1'^T X_2' = U\Sigma V^T$ and Σ is a diagonal matrix. S is a diagonal matrix with $s_{ii} = \pm 1$, and the signs of the s_{ii} are taken from the corresponding elements of Σ. Summing up, the main outputs of the procedure are (a) a consensus object, which in our case is the configuration of an average speaker characterized by the statistical properties as described above, (b) the eigenvalues of the transformation matrices, i.e., the diagonal of Σ, as the amounts of uniform deformation in the directions guaranteeing optimal "superimposition" in terms of the least-squares criterion applied, and (c) the normalized data of each single speaker. The data sets described in the following section will be evaluated with respect to these statistics. The results of (b) are often displayed as "a pair of orthogonal axes with lengths proportional to the two eigenvalues and oriented so that the longer axis is parallel to the direction (…) of maximum stretching" (Rohlf and Slice, 1990, p. 48). In the present case, this makes sense for the acoustic data set: A so-called strain cross, defined by the first two eigenvalues and the angle ψ, of a given speaker deviating to a great degree from the unit circle, indicates that this speaker's vowel space needs a higher amount of affine transformation in order to fit to the consensus object as compared to a speaker with a smaller strain cross.

C. Aims of this study

As was pointed out above, the major aim of this study is to evaluate the usefulness of the algorithm we proposed for both acoustic and articulatory data. The rationale of speaker normalization is chosen with regard to the methodological corollary of the universal articulatory phonetics hypothesis, according to which it is not sufficient to report patterns of individual differences among a homogeneous group of speakers, but "we must also consider ways in which such variability is lawful, because this variability must be made to square with the fact that language is a shared system" (Johnson *et al.*, 1993, p. 702). Therefore, apart from demonstrating the success of the normalization suggested, the second aim of the current work is to explore whether the modes of affine deformation are correlated to aspects of vocal-tract morphology.

II. METHOD

A. Data acquisition

Seven native speakers of German (five males, M1–M5, and two females, W1 and W2) were recorded by means of electromagnetic midsagittal articulography (EMMA, AG 100, Carstens Medizinelektronik). All speakers spoke a standard variety of German with at best slight dialectal variations: three speakers (W1, M1, and M4) originally come from South Germany, one speaker (W2) from Saxonia, two speakers (M3 and M5) from Northeast Germany, and one speaker (M2) from Berlin. At the time of recording, the speakers were between 25 and 40 years old and had lived in

Berlin for at least 5 years. The speech material consisted of words containing /tVt/ syllables with the full vowels /iː,ɪ,yː,ʏ,eː,ɛ,ɛː,øː,œ,aː,a,oː,ɔ,uː,ʊ/ in stressed and unstressed positions. Stress alternations were fixed by morphologically conditioned word stress and contrastive stress. Each symmetrical CVC sequence was embedded in the carrier phrase *Ich habe* /'tVtɐ/, *nicht* /tV'taːl/ *gesagt*. (*I said* /'tVtɐ/, *not* /tV'taːl/) with the test syllable /tVt/ in the first word always stressed and in the second word always unstressed. All 15 sentences were repeated six (four speakers) or ten times (three speakers). Four sensors were attached to the tongue, one to the lower incisors, and one to the lower lip. The analyses in this study are limited to the four sensors on the tongue for the remainder of this text and are numbered T1–T4 going from front to back. Two sensors on the nasion and the upper incisors served as reference coils to compensate for head movements relative to the helmet and for the definition of a preliminary coordinate system. This served as the basis for the final reference coordinate system, which was defined by recordings of two sensors on a T-bar, manufactured individually for each subject in order to determine his or her bite plane. Simultaneously, the speech signal was recorded by a digital audio tape recorder. Original sampling frequencies were 400 Hz for the EMMA data and 48 kHz for the acoustical signals. For the analyses, the EMMA signals were lowpass filtered at 30 Hz and downsampled to 200 Hz, while the acoustical signals were downsampled to 16 kHz.

B. Measurements

Formant frequencies of the first and second formant were measured interactively close to the mid of the vowel at the moment of minimal formant movement or, for lax vowels, at a turning point in the F2 trajectory. For estimating the frequencies of the first and second formants, the default settings of the software package SIGNALYZE (⟨http://www.signalyze.com/⟩) were used, i.e., LPC with 15 ms smoothing. The same temporal markers as described for the acoustic analyzes were also used for extracting tongue positions. Both acoustic and articulatory data were then averaged over the six respectively ten repetitions of each vowel. Information about palate shapes was acquired by measuring the artificial EPG palates of all seven speakers by means of a sliding caliper. This procedure gave the 3D coordinates for all EPG electrodes and the 2D coordinates of the palate midline approximately located between the two most central columns. Since the location of the EPG electrodes is adjusted to the speakers' anatomy, e.g., the rear border is aligned with the rear wall of the second molars, the EPG-based palate midline was deemed to be more exact compared to the palate outline, traced by means of EMMA. For addressing the question whether speaker-dependent differences can be explained by their palate shape, two measures were used: the palate length and a doming index (see Johnson *et al.*, 1993), which was calculated as the ratio between the total midsagittal length of individual EPG palates and the vertical distance between the first and the last point on the palate. Higher values indicate a palate with a higher degree of doming. Three sets of data were subjected to the analysis: tongue configurations during the 15 vowels in two stress conditions, measured as X and Y coordinates of the four sensors (30×8 matrix), frequencies of the first and second formant during the vowels (30×2 matrix), and palate outlines specified by 11 x and y coordinates (11×2 matrix). For measuring the palatal outline, we adapted the method described in Fitzpatrick and Ni Chasaide (2002). In short the 3D coordinates of the Reading EPG palates were measured as described above. From this we estimated the midsagittal outline from the positions of the two inner columns of the electrode positions. The outline was then adjusted to the EMMA data by eye.

C. Statistical apparatus

1. Quantification of the relationship pools of quantitative variables

Throughout the current work, methods are used, which require the correlation of data sets with independent as well as independent variables containing more than two variables. The correlations between articulatory configurations and the morphological data set could serve as an example. In this example, the research interest lies in gaining insight about the nature of the main directions of isotropic shape change in these two data modalities, substantial correlations indicating that similar variance components are targeted in both data sets. In order to explore relationships as these, canonical correlation analysis (CCA) was applied. CCA can be seen as the multivariate generalization of product-moment correlation. Considering the two matrices X and Y, the CCA finds a linear combination of the variables of X and a linear combination of the variables of Y of maximal correlation. It has often been claimed that CCA needs many cases compared to the number of variables. But Stevens (1986) discussed sample size in CCA and states that if the canonical correlations are strong (i.e., $R > 0.7$), then even small samples (e.g., $n = 50$) can be sufficient to detect significant correlations most of the time. Another drawback often reported for canonical correlations is that it is reported to be sensitive to multicollinearity among variables. This issue can at least to some extent be settled by reporting redundancy indices, which provide a partitioning of the explained variance into predictors and criteria. The redundancies of individual canonical variates can be summed up to yield an R^2-like measure of the contribution of the predictor-side canonical variates in explaining the criterion-side canonical variates and vice versa.

2. Classification procedures

By means of statistical discrimination, it is possible to check how successful different normalization procedures are at preserving the phonemic identity of vowel tokens. Adank *et al.* (2004) focused on two such classification approaches of vowel tokens. The most basic model, linear discriminant analysis (LDA) assumes that the covariance of each of the classes is identical. A more advanced method, quadratic discriminant analysis (QDA), makes no such assumptions but has the drawback of estimating more parameters, making it more difficult to crossvalidate and more prone to overfitting. These two approaches are just two instances of classification

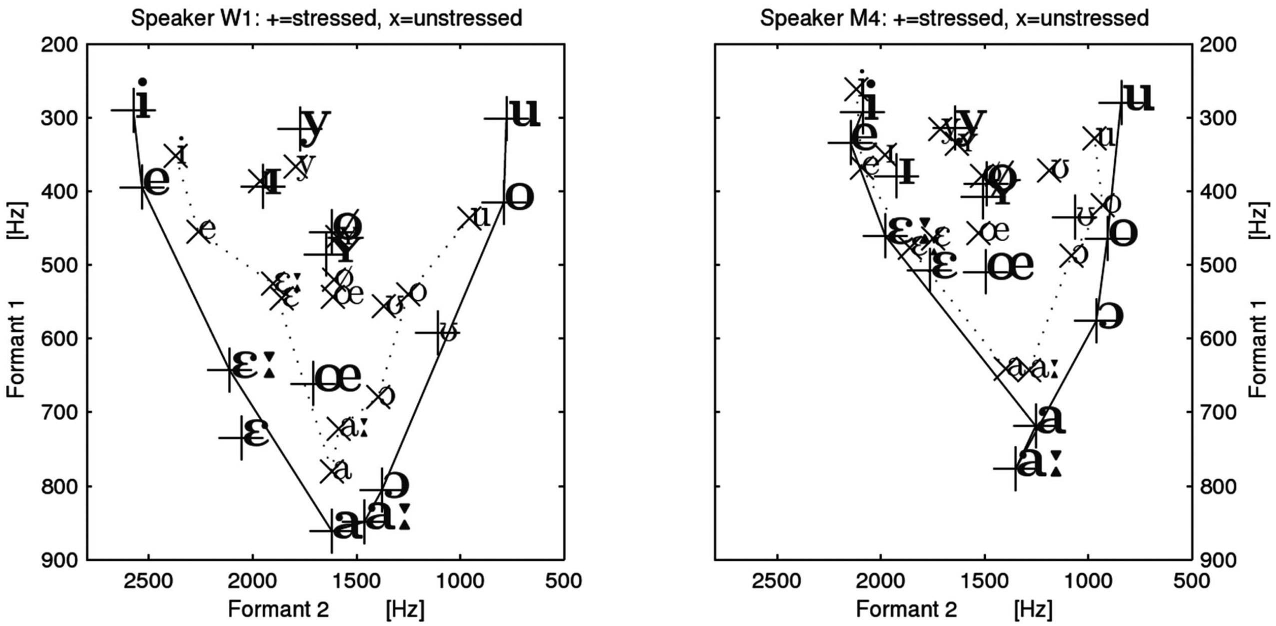


FIG. 1. Means of formant frequencies for stressed (bold and Large symbols with "+") and unstressed vowels [light and smaller symbols with "×") for a female (left) and a male speaker (right)].

procedures, with the LDA being one standard choice.[1] Having said that, it is obvious that it is necessary to be selective with regard to the classification procedures chosen. We compared two classification procedures, including (i) LDA—for comparability purposes with Adank *et al.* (2004) and (ii) an additional logistic discriminator (LOGDA) (Ripley, 1994). This logistic classifier has the advantage of making less assumptions than both QDA as well as LDA.

III. RESULTS

This section is concerned with the evaluation of the normalization procedure applied in this study. It will roughly be organized in analyses of the two modalities: acoustic and articulatory. The acoustic description of the data set involves comparing raw and normalized F1/F2 spaces. The capability of the approach to preserve phonemic identity, in particular with respect to the stress condition manipulated in the data set comprises both acoustic and articulatory modalities. Of special interest for the articulatory data is the question of how the deformations of the articulatory spaces are to be related to measurements of anatomical characteristics of the palate, as captured in terms of (a) results of the normalization procedure of palate outlines and (b) scalar doming indices (Johnson *et al.*, 1993). This also justifies the parsimonious selection of acoustic features: As we were primarily interested in vowel quality as manifested in tongue shapes, we have limited ourselves to the analysis of the first two formants in the acoustic domain. In other words, we were not interested in acoustic features such as F0 as it has stronger prosodic correlations—at least in our corpus—with stressed vowels carrying pitch accents. Similar arguments hold against F3, which mostly is associated with rounding information.

A. Formant spaces

In order to give a general impression of speaker-dependent differences for the production of German vowels, formant spaces of two speakers from Bavaria, one female (W1) and one male (M4), are presented in Fig. 1. The means of the frequencies of the first and second formant are indicated as bold symbols with a "+" sign for the stressed vowels. Unstressed vowels are indicated by "×." For reasons of clarity, the marginal vowels are connected by lines.

As can be seen, the speakers differ not so much in the relative location of vowels in the vowel system, with the exception of some minor relative changes in the location of the low vowels. However, the two speakers do differ in the way they realized the stressed-unstressed distinction: On the one hand, speaker W1 reduces the unstressed vowels consistently toward the center of the formant space. On the other hand, speaker M4's formant frequencies of unstressed vowels differ to a greater degree for the back and low vowels from their stressed counterparts, whereas the front vowels are only slightly affected. The direction of change for the back and low vowels suggests a more fronted and closer constriction for the unstressed vowels. Therefore, the acoustical results suggest that this speaker produces the stressed vowels with a greater contrast between the neighboring consonants and the vowels.

The consensus object, shown on the left in Fig. 2, gives more evidence for the latter strategy for reducing vowels; i.e., the back and the lower vowels are centralized when unstressed but not the high and mid palatal vowels /i, y, e, ø, ɪ, ʏ/, which change only very little. This finding correlates well with the observation that their constriction location is already quite close to the constriction location of the neighboring apical stops. The formant values of the remaining unstressed vowels change in the direction of front high vowels.

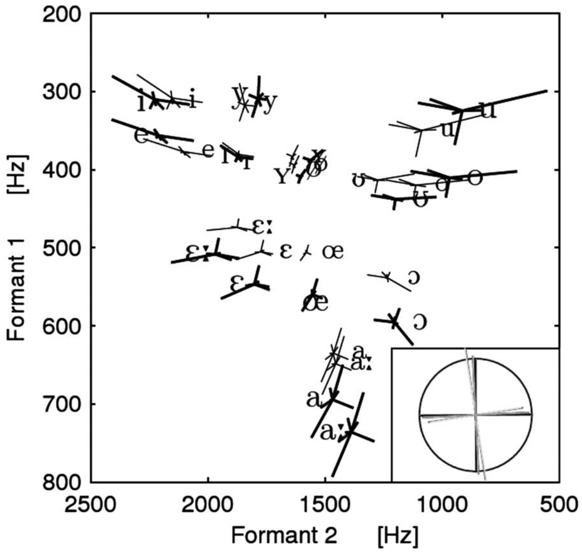


FIG. 2. Formant consensus object of stressed vowels (large symbols) and unstressed vowels (small symbols). Lines correspond to the distance between the speaker-dependent models for each of the speakers and the formant consensus object. Inset: amount of necessary speaker-dependent affine transformation compared to the consensus object, all speakers, displayed as strain cross.

TABLE I. Scatter remaining after normalization in percent of the original data.

	Model		Lobanov	
	Stressed	Unstressed	Stressed	Unstressed
i	34.63	29.56	61.73	26.92
ɪ	11.20	5.60	74.06	68.58
y	26.09	19.03	182.59	75.24
ʏ	1.47	3.43	17.88	44.16
e	19.92	8.98	23.91	12.58
ɛː	4.34	3.16	23.80	60.34
ɛ	3.00	1.67	33.41	32.97
ø	4.77	1.61	29.25	21.23
œ	5.17	1.97	48.68	59.73
aː	75.77	27.88	92.50	66.84
a	24.98	21.42	48.43	10.53
o	43.46	10.38	209.51	42.11
ɔ	19.36	11.94	30.73	64.01
u	151.72	40.43	206.62	37.73
ʊ	6.09	5.05	39.35	75.18
$\bar{x}$	28.80	12.81	74.83	46.54
sd	38.06	11.73	65.54	21.56

As was explained in Sec. I B 2, strain crosses display the amount of necessary affine transformation of individual speakers in order to fit to the consensus object. The more the crosses deviate from the cross in the unit circle in length and orientation, the more affine transformation was necessary for the corresponding speaker. Only for two speakers (M1 and W2) do the strain crosses deviate clearly from the unit circle. This implies that not much affine deformation is applied to single speakers' vowel spaces in order to achieve the best fit.

1. Cross-method comparisons

One way of evaluating the quality of a normalization procedure is to measure the amount of scatter remaining after normalization, which has been considered good practice as a tool for the evaluation of speaker normalizations since Disner (1980). Scatter remaining is computed as the percentage of ellipse areas of transformed data relative to the ellipse areas of the raw data. Table I summarized the results for our method and the Lobanov transform. In terms of scatter reduction, our method outperforms the Lobanov transform, which was confirmed by the results of a *t*-test ($t=3.67$, $p<0.01$, df=29). Another pattern which becomes evident from an inspection of Table I is that not all vowels are normalized to the same degree, an effect which appears to be common to both procedures applied. While there is a considerable amount of scatter reduction for most vowels, there are instances where there is even more variability after normalization. While technically possible, such a result runs counter the objective of normalization. This pattern is particularly prominent for stressed /uː/. Further, from visual inspection of the dispersion ellipses, it seemed evident that our procedure affects the orientations of the ellipses to a larger degree than the Lobanov transform, in particular for back vowels. This was evaluated by calculating the ellipse orientations of the raw data and the normalizations and then by bootstrapping the correlation coefficients of these angles. While the ellipse orientations of the Lobanov correlated well with the orientations of the original data ($r=0.94$**, 95% CI [0.86, 0.97]), this was not the case for the Procrustes-influenced approach ($r=0.29$, $p=0.11$, 95% CI [−0.08, 0.58]). The high CI range (0.67) for our method further evidences that only a subset of ellipse orientations is affected (see also Fig. 3).

Apart from scatter reduction, normalization procedures

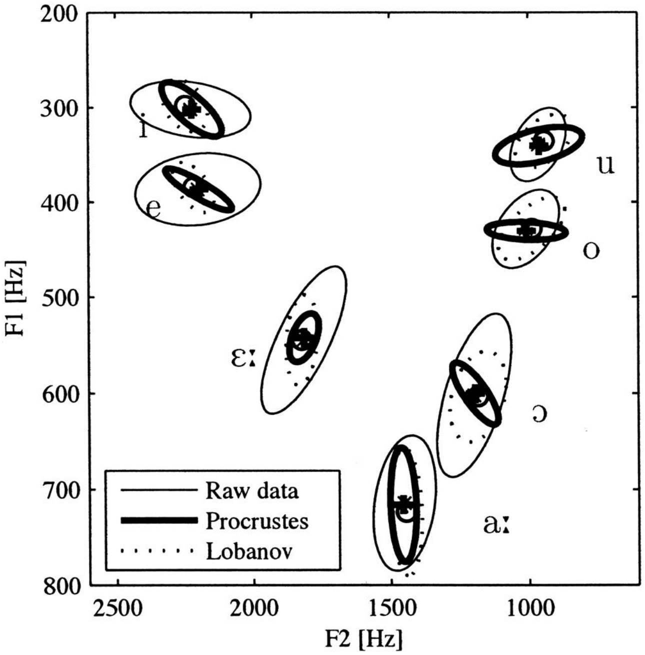


FIG. 3. $1-\sigma$ dispersion ellipses for raw formant data (thin lines), the speaker-dependent Procrustes data (bold lines), and the Lobanov transformed data (dashed line). Centroids are indicated by ○ (Lobanov), + (Procrustes), and × (raw data), respectively. For reasons of clarity, only the cardinal vowels are plotted here.

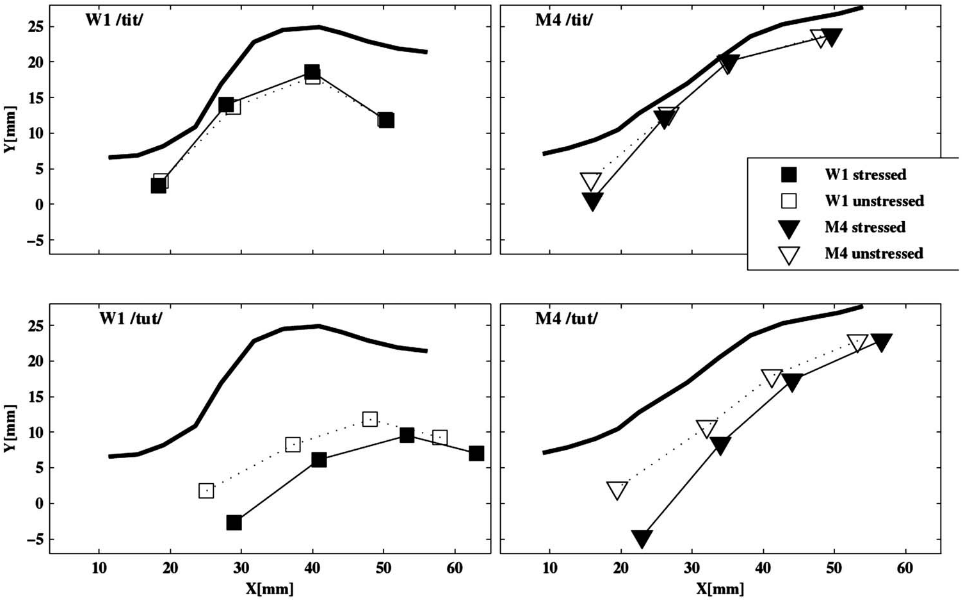


FIG. 4. Comparison between averaged articulatory configurations of the two speakers W1 (thin lines in left panels) and M4 (right panels). The upper two panels show tongue configurations for /tit/ (left) and the lower panels for /tut/. The upper curved lines show the outlines derived from EPG palates. Tongue configurations marked with filled markers were measured at the midpoints of stressed vowels and the empty markers for the corresponding unstressed vowels.

should also maintain the relative positions of items, e.g., in order to compare different languages or dialects. In order to ensure that this is not the case, we conducted a MANOVA to compare the formant means of the raw data, the Lobanov transformed data and the modeled Procrustes data. The differences were insignificant [Wilk's lambda $F(4,34)=0.00006$, $p=1$].

B. Articulatory spaces

In the second part of this section, lingual configurations for the 15 vowels of German in stressed and unstressed positions are used as input to the normalization procedure. In order to exemplify the consequences of speaker-dependent morphology and inconsistent sensor placements, we turn to a description of speaker-dependent strategies for the stress distinction. Figure 4 shows tongue configurations (X-and Y-coordinates) for the four tongue sensors *before* normalization. The upper two panels of this figure show the tongue configurations during the vowel /i/ in stressed (filled symbols) and unstressed positions (unfilled symbols). On the left side, data of the female speaker W1 are displayed together with her palate outline, and on the right side, those for the male speaker M4. Typically, for /i/ the tongue is braced against the sides of the palate. In the lower two panels of Fig. 4, the tongue configurations are presented for the back vowel /u/. The speakers W1 and M4 not only differ to a great degree in the way they produce the vowel /i/ but also in their palate shapes. Speaker W1 has an extremely steep and domed palate, whereas the palate of speaker M4 is rather flat. Accordingly, the tongue configuration for /i/ of speaker W1 is bunched toward the palate and speaker M4's is shaped straighter and oriented in parallel with the palate outline. Despite these immediately obvious differences between these two speakers, both show little differences between stressed and unstressed tongue configurations for /i/. For /u/ the tongue shapes again show a pronounced difference between the speakers. In contrast to /i/, however, stress affects the tongue configurations for both speakers, with the unstressed /u/ being produced with an elevated tongue tip and a generally more fronted tongue body. Based on these observations, it seems reasonable to conjecture (i) that interindividual differences in anatomy can to a large extent be made responsible for the patterns just described, (ii) that the speakers' vowel gestures are still functionally equivalent, i.e., forming a close palatal constriction with the front part of the tongue for /i/ and a uvular constriction with the rear part of the tongue for /u/, and (iii) that speakers apply a general strategy for producing the stress distinction.[2]

In order to test these assumptions, the normalization procedure was separately applied to both the articulatory data and the palate outlines. Speaker-dependent data were calculated according to Eq. (3). The results are shown in Fig. 5. Plotted in bold lines are the X-and Y-coordinates of palate and tongue configurations of the consensus objects, the thinner lines representing the speaker-dependent normalized data of speaker W1 (left) and speaker M4 (right). These two speakers are extreme in their configurations; the remaining speakers compromise between these two participants and therefore are closer to the consensus object. These speakers were displayed to illustrate the problem to be tackled in the articulatory domain. As can be seen in the upper two panels, speaker-dependent tongue configurations for the high front vowel /i/ deviate very little from the consensus object. This is mainly due to the fact that the applied transformations rigorously twist the tongue configuration of speaker M4 in order to fit in the consensus object. For the back vowel /u/ there is more speaker-dependent deviation from the consensus object. Concerning the stressed-unstressed distinction, tongue configurations of /i/ differ only very slightly with a somewhat lower tongue tip position for stressed /i/. For /u/, however, the speaker-dependent and the consensus configurations of stressed /u/ are clearly more retracted, and the tongue tip points downward. Even though the speakers vary in the extent of the difference, the direction is similar for the two speakers and the consensus object.

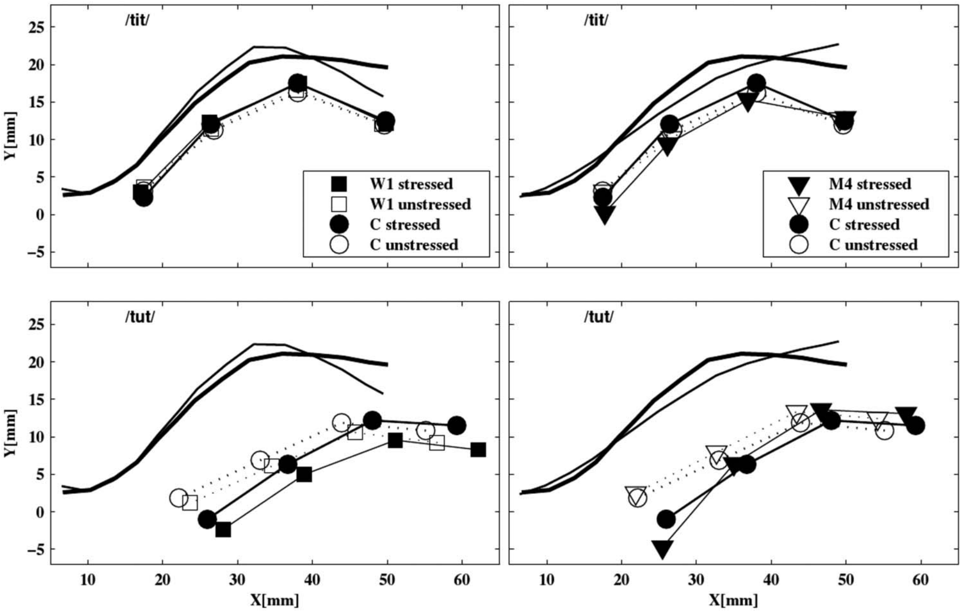


FIG. 5. Speaker-dependent modeled articulatory configurations (tongue contours with filled square symbols) and the consensus objects (contours with circles) for /tit/ (upper panels) and /tut/ (lower panels). Palate contours of the consensus object are printed as thin lines, the contours of speaker-dependent modeled data as bold lines. Modeled data of speaker W1 are presented on the left side and that of speaker M4 on the right side.

C. Relationship between tract morphology and speaker-dependent modeled data

By means of CCA, we aimed at summarizing the relations between the eigenvalues—which measure the amount of affine deformation—in the tongue and the palate data sets as analyzed by the normalization. Recall that the eigenvalues of the transformation matrix relating the consensus object to the individual articulatory spaces contain the amount of uniform affine deformation relating the individual speaker to the consensus. In the morphological data set, these directions have visually interpretable meanings in 2D-Euclidean space, which is not the case for the articulatory vowel spaces. Therefore, we truncated the matrix of eigenvalues in the articulatory data set by visual inspection of the scree plot. Its inspection suggested the use of four eigenvalues, and these were entered in the canonical correlation as predictors (7 speakers × 4 matrix) and both eigenvalues of the palate outline analysis served as criteria (7 × 2 matrix). Given the low explanatory power of the data set, the results of the following analyses are to be considered as exploratory descriptions. Further note that the division in predictors and criteria is meaningless in CCA; the number of extracted canonical correlations is equal to the minimum number of variables in either set. The expected behavior of the analysis is as follows: Apart from finding substantial canonical correlations between predictor and criterion variables, we expected higher redundancies for the morphological variates, given the tongue configurations. The canonical correlations in this analysis were substantial with values of 0.94 and 0.78, and the summed redundancies over both covariates amounted to 71% of the variance on the criterion side (palate outline data) but only to 52% on the predictor side. In other words, the correlation appears to be substantial and, more importantly, the normalization procedure captures similar directions in the tongue and palate data. Table II summarizes these results.

As a next step, selected intermediate results as obtained from the analysis of tongue shapes were related to the palate doming variable. Here, standard multiple regression was used in order to roughly describe the correlational structure between eigenvalue predictors as derived from palate outline and tongue configuration data sets. To our surprise, the multiple regression for the palate outline eigenvalues on the doming index was not substantial with only 11% of explained variance. In contrast, the regression of the first four eigenvalues of the tongue analysis was substantial, explaining 95% of the variance. Furthermore, it was again (as in the canonical analysis described) necessary to include more than two predictors to capture the anatomical variance. In summary—given the exploratory character of these analyses—a substantial proportion of the variance captured by the normalization procedure when applied to articulatory configurations seems to be shared with what is extracted from the palate outlines.

D. Prediction of phonemic identity

A further check of the success of normalization procedures is to make attempts at measuring the increment of predictability of phonemic identity caused by normalization procedures [see Adank *et al.* (2004) for a more detailed adoption of such a rationale]. This is possible by means of statistical

TABLE II. Summary of analyses relating tract morphology to tongue analyzes.

	CCA	
Can correlations		
	CC I	0.94
	CC II	0.78
Redundancies		
	R^2_{yx}	0.71
	R^2_{xy}	0.52
	Multiple regressions on doming index	
Predictor		
	Tongue eigenvalues	95%
	Palate eigenvalues	11%

TABLE III. Percentages of correctly classified vowels by LDA and LOGDA. The predictors were first and second formant values (acoustic data set) or EMMA coil positions of the four tongue sensors (articulatory data set). Vowel quality with 30 levels (15 German monophthongs is stressed and unstressed positions) served as dependent variables for both data sets. Percentages are given for the whole data sets and separately for stressed and unstressed subsets. In brackets: benefit from normalization (in %).

Data set	Method		Raw	Normalized	Lobanov
Acoustic					
	LDA	Whole	47	80(33)	63(16)
		Stressed	54	82(26)	67(13)
		Unstressed	39	77(38)	60(21)
	LOGDA	Whole	51	83(32)	68(17)
		Stressed	57	86(29)	73(16)
		Unstressed	44	80(36)	63(19)
Articulatory					
	LDA	Whole	41	74(33)	⋯
		Stressed	44	70(25)	⋯
		Unstressed	38	77(39)	⋯
	LOGDA	Whole	48	87(39)	⋯
		Stressed	59	90(31)	⋯
		Unstressed	37	83(46)	⋯

discrimination. The predictors used in the current context were the first and second formant frequency values in the acoustic data set and EMMA coil positions for four tongue sensors in the articulatory data set. The data were pooled over speaker and vowel identity in stressed and unstressed positions, yielding 210 (7 speakers × 15 vowels × 2 stress conditions) cases altogether. Vowel identity (15 German monophthongs in stressed and unstressed positions) was the dependent variable. In order to compare the normalization we used with an alternative normalization scheme, the acoustic data set was also Lobanov transformed and subjected to the same discrimination analysis. The calculation of an analog to the Lobanov-normalization in the articulatory case was not considered as meaningful and therefore not undertaken. Results were evaluated in terms of percentage correctly classified tokens, which can be calculated from confusion matrices between vowel quality as intended by the speaker and vowel quality as predicted by the classification procedures. These percentages were calculated both on the whole data set and separately for vowels in stressed and unstressed conditions in order to reveal potentially different effects of the normalization procedures on stressed and unstressed tokens. Classification results were compared by means of McNemar χ^2-tests. In a first step, LOGDA was compared with LDA. LOGDA had the tendency to perform better than LDA, although this effect depended on whether the classification procedures were applied to the raw data or to one of the normalized data sets. For example, LOGDA achieved significantly higher amounts of correctly classified items than LDA for the Lobanov data ($\chi^2=2.7$, $p=0.049$), but this effect did not reach the level of significance neither for the procedure proposed here ($\chi^2=1.33$, $p=0.12$) nor for the raw data ($\chi^2=2.23$, $p=0.067$). In the following we report the results of our analyses with both classification procedures. In contrast to the comparison of classification methods, the comparison of normalization procedures yielded the by far more substantial results (summarized in Table III): Regardless of the classification method used, the normalization procedure proposed performed substantially better than the Lobanov procedure (LDA: $\chi^2=12.38$, $p=0.0002$, LOGDA: $\chi^2=11.69$, $p=0.0003$). This result is in accordance with the amount of scatter reduction reported above.

IV. SUMMARY AND DISCUSSION

In this paper, we have described and applied a normalization procedure applicable to articulatory and acoustic vowel spaces. The procedure consists of constructing a so-called consensus object with the property that the normalized acoustic or articulatory spaces have equal variances in the main directions of affine deformation and performing multiple multivariate regression analysis of this consensus object on the raw configurations to yield speaker-specific normalized data. We delivered a qualitative description of raw and normalized configurations as the first empirical step. For the transformation from the consensus to the speaker-specific formant spaces, only little affine transformation appeared to be necessary (see Fig. 1), which is equivalent to the orthogonal and affine versions of the algorithm yielding very similar results. Still, a higher degree of scatter reduction of speaker-specific variation was achieved by our procedure as compared to Lobanov speaker normalization (see Table I). This benefit in terms of scatter reduction comes at the cost of partially more aggressively transforming the orientation of the data for some vowel categories. For the articulatory data, substantial affine transformation was necessary in order to map the speaker-specific data onto the consensus object. As a result, affine transformations yielded tongue configurations for which a large part of speaker-specific pellet placement and shape differences were removed (see Fig. 5).

We also tested the assumption that the normalization procedure for the articulatory spaces captures directions in the data which correspond to primary dimensions of uniform shape change. For this purpose we applied the normalization procedure to a control data set of anatomical characteristics, i.e., palate outlines. The substantial correlations obtained suggest that the procedure's success is at least partly related to the removal of uniform shape change differences between individual speakers.

Finally, in order to measure the procedure's capability to recover phonemic identity, we performed separate discrimination analyzes of original and normalized vowel spaces in both articulatory and acoustic domains on phonemic identity. Phoneme recovery probabilities for both articulatory and acoustic data increased substantially in terms of percentages of correctly classified tokens (see Table III). In particular, the classification rates for the unnormalized acoustic data are relatively low in comparison to results published in the literature. For example, Adank *et al.* (2004) reported correct classification rates for unnormalized data (i.e., their "Hz" condition) of about 80% and of more than 90% after Lobanov transformation. However, we do not consider this as alarming for the following reasons: First, they entered additional predictors ($F0$ and $F3$) in their discriminators; second, their data set contained more speakers than ours (160 vs 7), and, presumably most important, ours contain more than three times as many categories in the criterion (30 vs 9) due to the more crowded German vowel system and the additional word stress condition. A further point to consider for the articulatory data set is that the discrimination procedure had no access to lip rounding information, which presumably provides the most important information for the distinction between rounded and non-rounded front vowels than the tongue data presented to the discrimination procedure here.

Another striking aspect of this analysis is that unstressed vowels benefit more from the result of the normalization. This pattern in principle holds for both the articulatory and the acoustic data sets, although there is a large difference in the performance of the discrimination procedure already in the unnormalized baseline for the articulatory data. This observation is compatible with a scenario according to which unstressed vowels are more prone to coarticulatory influences of the consonantal environment, which is sensitive to the normalization (Mooshammer and Geng, 2008).

Still, as already mentioned in the Introduction, the present study was designed not only to propose an alternative normalization scheme but also to attempt to quantitatively relate aspects of vocal-tract anatomy to the functioning of the normalization. This second aim of the study resembles classical tenets of the universal articulatory phonetics hypothesis (Johnson *et al.*, 1993) according to which interindividual differences should be lawfully related to factors such as others vocal-tract geometry. This was achieved by validating the directions extracted by our normalization method against independently extracted models of palate shapes, which resulted in high correlations of these independently extracted directions. This suggests that the normalization procedure in the articulatory data set partially operates on interindividual differences related to aspects of tract morphology.

This virtue to some extent might at the same time be the largest drawback: Allowing orthogonal rotational or even affine transformations in addition to the Lobanov scalings clearly has a strong desirable effect on phoneme classification rates, but in some instances such transformations might be too aggressive and distort the categorical structure. In particular, ellipse orientations are affected more strongly by our procedure than it is the case for the more conservative Lobanov transform. This particular observation on ellipse orientation and the likewise less conservative nature of formalism and transformation applied might lead to speculations about detrimental effects of the procedure also in other situations. Still, as demonstrated, the affine version of the procedure produces output showing how the fit was achieved. However, affine transformation might be undesirable in some settings. Then, it might be indicated to revert to the orthogonal version—the generalized Procrustes analysis—or to the Lobanov transform. Still, even in such cases, the analysis provides relevant diagnostic information on the data set and therefore presents a useful additional tool for the phonetic practitioner.

ACKNOWLEDGMENTS

Special thanks to Dirk Fischer, Suse Fuchs, and Daniel Pape for their help with analyzing the data and to Jörg Dreyer for assisting in the EMMA recordings. We are also very grateful to Jonathan Harrington, Phil Hoole, and Michel Scheffers for invaluable comments on earlier drafts of this paper. Any remaining errors are our responsibility. This work was supported by the German Research Council (DFG) Grant No. GWZ 4/5-2(A1).

[1] But, as Hastie *et al.* (1995, p. 2)—quoting Michie *et al.* (1994)—note, often the "LDA produces the best classification results because of its simplicity […]. LDA was among the top 3 classifiers for 11 of 22 data sets studied in the STATLOG project." Further, they mention that the cases of shortcomings of LDA can be paraphrased as "saying that a single prototype per class is insufficient." In this line of reasoning, the use of arbitrary decision regions is prohibitive because more than one prototype per category runs against the concept of vowel normalization, which aims at augmenting the coherence between individual tokens and the phonetic class.

[2] As outlined in greater detail in Mooshammer and Geng (2008) and Mooshammer and Fuchs (2002), there is an interesting asymmetry in German: Tense vowels are substantially shortened in unstressed syllables, whereas lax vowels tend to maintain their duration in stressed and unstressed positions. Nevertheless, for both vowel series the tongue tip is elevated and fronted in unstressed position because of the adjacent alveolar stops. Therefore, we concluded that German unstressed vowels are generally produced with a greater degree of coarticulation, irrespectively of durational shortening.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Apostol, L., Perrier, P., and Bailly, G. (**2004**). "A model of acoustic interspeaker variability based on the concept of formant-cavity affiliation," J. Acoust. Soc. Am. **115**, 337–351.

Beckman, M. E., Jung, T., Lee, S., de Jong, K., Krishnamurthy, A., Ahalt, S., Cohen, B., and Collins, M. (**1995**). "Variability in the production of quantal vowels revisited," J. Acoust. Soc. Am. **97**, 471–490.

Clopper, C., Pisoni, D., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Disner, S. F. (**1980**). "Evaluation of vowel normalization procedures," J. Acoust. Soc. Am. **67**, 253–261.

Fitzpatrick, L., and Ni Chasaide, A. (**2002**). "Estimating lingual constriction location in high vowels: A comparison of EMA- and EPG-based measures," J. Phonetics **30**, 397–415.

Geng, C., and Mooshammer, C. (**2000**). "Modeling the German stress distinction," in Proceedings of the Fifth Speech Production Seminar, pp. 161–164.

Goodall, C. R., and Green, P. (**1986**). "Quantitative analysis of surface growth," Botanical Gazette **147**, 1–15.

Gower, J. C. (**1975**). "Generalized Procrustes analysis," Psychometrika **40**, 33–51.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

Harshman, R. A., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Hashi, M., and Westbury, J. (**1998**). "Vowel posture normalization," J. Acoust. Soc. Am. **104**, 2426–2437.

Hastie, T., Tibshirani, R., and Buja, A. (**1995**). "Flexible discriminant and mixture models," in *Neural Networks and Statistics*, edited by J. Kay and D. Titterington (Oxford University Press, New York).

Hoole, P. (**1999**). "On the lingual organization of the German vowel system," J. Acoust. Soc. Am. **106**, 1020–1032.

Jackson, M. T. T. (**1988**). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.

Johnson, K. (**2005**). "Speaker normalization in speech perception," in *The Handbook of Speech Perception*, edited by D. Pisoni and R. Remez (Blackwell, Oxford), pp. 363–389.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). "Individual differences in vowel production," J. Acoust. Soc. Am. **94**, 701–714.

Joos, M. (**1948**). "Acoustic phonetics," Language Monograph 23, Supplement to Language **24**, 1–136.

Kingston, J., and Diehl, R. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Lobanov, B. M. (**1971**). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**, 606–608.

Michie, D., Spiegelhalter, D., and Taylor, C. (**1994**). *Machine Learning, Neural and Statistical Classiffication* (Ellis Horwood, New York).

Mooshammer, C., and Fuchs, S. (**2002**). "Stress distinction in German: Simulating kinematic parameters of tongue tip gestures," J. Phonetics **30**, 337–355.

Mooshammer, C., and Geng, C. (**2008**). "Acoustic and articulatory manifestations of vowel reduction in German," J. Int. Phonetic Assoc. **38**, 117–136.

Nearey, T. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Perkell, J., and Nelson, W., (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1895.

Ripley, B. (**1994**). "Neural networks and related methods for classification," J. R. Stat. Soc. Ser. B (Methodol.) **56**, 409–456.

Rohlf, F. J., and Slice, D. (**1990**). "Extensions of the Procrustes method for the optimal superimposition of landmarks," Syst. Zool. **39**, 40–59.

Simpson, A. (**2002**). "Gender-specific articulatory-acoustic relations in vowel sequences," J. Phonetics **30**, 417–436.

Stevens, J. (**1986**). *Applied Multivariate Statistics for the Social Sciences* (Erlbaum, Hillsdale, NJ).

Westbury, J. R. (**1994**). *X-ray Microbeam Speech Production Database User's Handbook, Version 1.0* (Waisman Center on Mental Retardation & Human Development, Madison, WI).

Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering

Paavo Alku[a] and Carlo Magi[b]
Department of Signal Processing and Acoustics, Helsinki University of Technology, P.O. Box 3000, Fi-02015 TKK, Finland

Santeri Yrttiaho
Department of Signal Processing and Acoustics and Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, P.O. Box 3000, Fi-02015 TKK, Finland

Tom Bäckström
Department of Signal Processing and Acoustics, Helsinki University of Technology, P.O. Box 3000, Fi-02015 TKK, Finland

Brad Story
Speech Acoustics Laboratory, University of Arizona, Tuscon, Arizona 85721

(Received 23 November 2007; revised 5 February 2009; accepted 8 February 2009)

Closed phase (CP) covariance analysis is a widely used glottal inverse filtering method based on the estimation of the vocal tract during the glottal CP. Since the length of the CP is typically short, the vocal tract computation with linear prediction (LP) is vulnerable to the covariance frame position. The present study proposes modification of the CP algorithm based on two issues. First, and most importantly, the computation of the vocal tract model is changed from the one used in the conventional LP into a form where a constraint is imposed on the dc gain of the inverse filter in the filter optimization. With this constraint, LP analysis is more prone to give vocal tract models that are justified by the source-filter theory; that is, they show complex conjugate roots in the formant regions rather than unrealistic resonances at low frequencies. Second, the new CP method utilizes a minimum phase inverse filter. The method was evaluated using synthetic vowels produced by physical modeling and natural speech. The results show that the algorithm improves the performance of the CP-type inverse filtering and its robustness with respect to the covariance frame position. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3095801]

PACS number(s): 43.70.Gr, 43.70.Jt [CHS] Pages: 3289–3305

I. INTRODUCTION

All areas of speech science and technology rely, in one form or another, on understanding how speech is produced by the human voice production system. In the area of voice production research, glottal inverse filtering (IF) refers to methodologies that aim to estimate the source of voiced speech, the glottal volume velocity waveform. The basis for these techniques is provided by the classical source-filter theory, according to which the production of a voiced speech signal can be interpreted as a cascade of three separate processes: the excitation, that is, the glottal volume velocity waveform, the vocal tract filter, and the lip radiation effect (Fant, 1970). In order to compute the first of these processes, IF methodologies estimate the second and third processes typically in forms of linear, time-invariant digital systems and then cancel their contribution from the speech signal by filtering it through the inverse models of the vocal tract and lip radiation effect. Since the lip radiation effect can be estimated at low frequencies as a time-derivative of the flow (Flanagan, 1972), which is easily modeled digitally by a fixed first order finite impulse response (FIR) filter, the key problem in IF methods is the estimation of the vocal tract.

Among the main methodologies used to analyze human voice production, IF belongs to the category of acoustical methods. As alternatives to the acoustical methods, it is possible to investigate voice production with visual inspection of the vocal fold vibrations or with electrical (e.g., Lecluse *et al.*, 1975) or electromagnetic methods (Titze *et al.*, 2000). Visual analysis of the vibrating vocal folds is widely used especially in clinical investigation of voice production. Several techniques, such as video stroboscopy (e.g., Hirano, 1981), digital high-speed stroboscopy (e.g., Eysholdt *et al.*, 1996), and kymography (Švec and Schutte, 1996), have been developed, and many of them are currently used in daily practices in voice clinics. Acquiring visual information about voice production, however, always calls for invasive measurements in which the vocal folds are examined either with a solid endoscope inserted in the mouth or with a flexible fiberscope inserted in the nasal cavity. In contrast to these techniques, a benefit of glottal IF is that the analysis can be computed from the acoustic signal in a truly non-invasive manner. This feature is essential especially in such research areas in which vocal function needs to be investigated under as natural circumstances as possible, for instance, in under-

[a] Electronic mail: paavo.alku@tkk.fi
[b] Deceased in February 2008.

standing the role of the glottal source in the expression of vocal emotions (Cummings and Clements, 1995; Gobl and Ní Chasaide, 2003; Airas and Alku, 2006) or in studying occupational voice production (Vilkman, 2004; Lehto *et al.*, 2008). In addition to its non-invasive nature, glottal IF provides other favorable features. IF results in a temporal signal, the glottal volume velocity waveform, which is an estimate of a real acoustical waveform of the human voice production process. Due to its direct relationship to the acoustical production of speech, estimates of glottal excitations computed by IF can be modeled with their artificial counterparts to synthesize human voice in speech technology applications (Klatt and Klatt, 1990; Carlson *et al.*, 1991; Childers and Hu, 1994).

Since the introduction of the idea of IF by Miller (1959), many different IF methods have been developed. The methods can be categorized, for example, based on the input signal, which can be either the speech pressure waveform recorded in the free field outside the lips (e.g., Wong *et al.*, 1979; Alku, 1992) or the oral volume velocity captured by a specially designed pneumotachograph mask, also known as the Rothenberg mask (e.g., Rothenberg, 1973; Hertegård *et al.*, 1992). In addition, methods developed to do IF differ depending on whether they need user adjustments in defining the settings of the vocal tract resonances (e.g., Price, 1989; Sundberg *et al.*, 2005) or whether the analysis is completely automatic (e.g., Veeneman and BeMent, 1985). From the methodological point of view, the techniques developed can be categorized based on how the effect of the glottal source is taken into account in the estimation of the vocal tract in the underlying IF method. From this perspective, there are, firstly, methods (e.g., Alku, 1992) that are based on the gross estimation of the glottal contribution during both the closed and open phase of the glottal pulse using all-pole modeling. By canceling the glottal contribution from the speech signal, a model for the vocal tract is computed with linear prediction (LP) (Rabiner and Schafer, 1978) although other spectral envelope fitting techniques such as those based on the penalized likelihood approach (Campedel-Oudot *et al.*, 2001) or cepstrum analysis (Shiga and King, 2004) could, in principle, be used as well. Secondly, the use of a joint optimization of the glottal flow and vocal tract is possible based on synthetic, pre-defined models of the glottal flow (e.g., Milenkovic, 1986; Kasuya *et al.*, 1999; Fröhlich *et al.*, 2001; Fu and Murphy, 2006). Thirdly, it is possible to estimate the glottal flow using closed phase (CP) covariance analysis (Strube, 1974; Wong *et al.*, 1979). This is based on the assumption that there is no contribution from the glottal source to the vocal tract during the CP of the vocal fold vibration cycle. After identification of the CP, covariance analysis is used to compute a parametric all-pole model of the vocal tract using LP.

CP covariance analysis is among the most widely used glottal IF techniques. Since the original presentation of the method by Strube (1974), the CP method has been used as a means to estimate the glottal flow, for instance, in the analysis of the phonation type (Childers and Ahn, 1995), prosodic features of connected speech (Strik and Boves, 1992), vocal emotions (Cummings and Clements, 1995), source-tract interaction (Childers and Wong, 1994), singing (Arroabarren and Carlosena, 2004), and speaker identification (Plumpe *et al.*, 1999). In addition to these various applications, CP analysis has been a target of methodological development. The major focus of this methodological work has been the method of accurately determining the location of the covariance frame, the extraction of the CP of the glottal cycle. In order to determine this important time span from a speech waveform, an approach based on a series of sliding covariance analyses is typically used. In other words, the analysis frame is sequentially moved one sample at a time through the speech signal and the results of each covariance analysis are analyzed in order to determine the CP. Strube (1974) used this approach and identified the glottal closure as an instant when the frame was in a position which yielded the maximum determinant of the covariance matrix. Wong *et al.* (1979) instead defined the CP as the interval when the normalized squared prediction error was minimum, and this technique has been used by several authors since, although sometimes with slight modifications (e.g., Cummings and Clements, 1995). Plumpe *et al.* (1999), however, argued that the use of the prediction error energy in defining the frame position of the covariance analysis might be problematic for sounds which involve gradual closing or opening of the vocal folds. As a remedy, they proposed an idea in which sliding covariance analyses are computed and formant frequency modulations between the open and CP of the glottal cycle are used as a means to define the optimal frame position. Akande and Murphy (2005) suggested a new technique, adaptive estimation of the vocal tract transfer function. In their method, the estimation of the vocal tract is improved by first removing the influence of the glottal source by filtering the speech signal with a dynamic, multi-pole high-pass filter instead of the traditional single-pole pre-emphasis. The covariance analysis is then computed in an adaptive loop where the optimal filter order and frame position are searched for by using phase information of the filter candidates.

All the different CP methods referred to above are based on the identification of the glottal CP from a single source of information provided by the speech pressure waveform. Therefore, they typically involve an epoch detection block in which instants of glottal closure and opening are extracted based on algorithms such as DYPSA (Naylor *et al.*, 2007). Alternatively, if electroglottography (EGG) is available, it is possible to use two information channels so that the position and duration of the CP is estimated from EGG, and then the speech waveform is inverse filtered. This so-called two-channel analysis has been shown to yield reliable results in IF due to improved positioning of the covariance frame (Veeneman and BeMent, 1985; Krishnamurthy and Childers, 1986). In this technique, the CP analysis is typically computed by estimating the CP of the glottal cycle as the time interval between the minimum and maximum peaks of the first time-derivative of the EGG waveform (Childers and Ahn, 1995). It is important to notice that even though there have been many modifications to CP analysis since the work by Strube (1974), all the methods developed are based on the same principle in the mathematical modeling of the vocal tract, namely, the use of conventional LP with the covariance criterion described in Rabiner and Schafer (1978).

Even though different variants of CP covariance analysis have been shown to yield successful estimates of the glottal flow by using simple synthesized vowels, this IF methodology has certain shortcomings. Several previous studies have in particular indicated that glottal flow estimates computed by the CP analysis vary greatly depending on the position of the covariance frame (e.g., Larar *et al.*, 1985; Veeneman and BeMent, 1985; Yegnanarayana and Veldhuis, 1998; Riegelsberger and Krishnamurthy, 1993). Given the fundamental assumption of the method, that is the computation of the vocal tract model during an excitation-free time span, this undesirable feature of the CP analysis is understandable. The true length of the glottal CP is typically short, which implies that the amount of data used to define the parametric model of the vocal tract with the covariance analysis is sparse. If the position of this kind of a short data frame is misaligned, the resulting linear predictive filter typically fails to model the vocal tract resonances, which might result in severe distortion of the glottal flow estimates. This problem is particularly severe in voices of high fundamental frequency (F0) because they are produced by using very short lengths in the glottal CP. In order to cope with this problem, previous CP methods typically exploit techniques to improve the extraction of the covariance frame position. In the present work, however, a different approach is suggested based on setting a mathematical constraint in the computation of the inverse model of the vocal tract with LP. The constraint imposes a predefined value for the direct current (dc) gain of the inverse filter as a part of the optimization of the filter coefficients. This results in vocal tract filters whose transfer functions, in comparison to those defined by the conventional covariance analysis, are less prone to include poles in positions in the z-domain that are difficult to interpret from the point of view of the classical source-filter theory of vowel production (e.g., on the positive real axis). This new dc-constrained vocal tract model is combined in the present study with an additional procedure, checking of the minimum phase property of the inverse filter, to yield a new CP algorithm.

In the following, typical artifacts caused by the CP analysis are first described using representative examples computed from natural vowels. These examples are then used to motivate the proposed new method to compute LP in vocal tract modeling of the CP analysis. The new method is then tested with both synthetic vowels produced by physical modeling of the human voice production mechanism and with natural speech of both female and male subjects.

II. METHODS

A. Sources of distortion in the conventional CP analysis

In this section, two major sources of error in the conventional CP analysis are described with the help of examples. The word "conventional" refers here to the CP analysis in which the vocal tract is modeled with a pth order all-pole filter computed by the basic form of the covariance analysis described by Rabiner and Schafer (1978), and the lip radiation effect is modeled with a fixed first order FIR filter. All the analyses described were computed using the sampling frequency of 8 kHz and the order of the vocal tract filter set to $p=12$. The length of the covariance frame was 30 samples (3.75 ms). The instant of glottal closure was extracted, when needed, as the instant of the negative peak of the EGG derivative.

First, the sensitivity of the glottal flow estimate about the position of the covariance frame is demonstrated. Figure 1 shows three glottal flow estimates, which were inverse-filtered from the same token of a male subject uttering the vowel [a] by using a minor change in the position of the covariance frame position: the beginning of the covariance frame in Figs. 1(b) and 1(c) was moved earlier in the signal by two and four samples, respectively, in comparison to the beginning of the covariance frame used in Fig. 1(a). The inverse filters obtained are shown in the z-domain in the left panels of Fig. 2, and the amplitude spectra of the corresponding vocal tract filters are depicted in the right panels of the same figure. The example indicates how a minor change in the position of the covariance frame has resulted in a substantial change in the estimated glottal flows. It is worth noticing that the covariance analyses illustrated in Figs. 2(a) and 2(b) have resulted in two inverse filters both of which have one root on the positive real axis in the z-domain. In Fig. 2(b), the position of this root is slightly closer to the unit circle than in Fig. 2(a). The CP analysis shown in Fig. 2(c) has, in turn, resulted in an inverse filter with a complex conjugate pair of roots at low frequencies. The effect of an inverse filter root which is located on the positive real axis approaches that of a first order differentiator [i.e., $H(z)=1-z^{-1}$] when the root approaches the unit circle, and a similar effect is also produced by a complex conjugate pair of roots at low frequencies. Consequently, the resulting glottal flow estimate, as shown in Figs. 1(b) and 1(c), becomes similar to a time-derivative of the flow candidate given by an inverse filter with no such roots or when these roots are located in a more neutral position close to the origin of the z-plane. This severe distortion of the glottal flow estimate caused by the occurrence of inverse filter roots, both real and complex conjugate pairs, at low frequencies is greatest at time instants when the flow changes most rapidly, that is, near glottal closure. As shown in Figs. 1(b) and 1(c), this distortion[1] is typically seen as sharp negative peaks, called "jags" by Wong *et al.* (1979), of the glottal flow pulses at the instants of closure.

The undesirable distortion of the glottal flow estimates by the occurrence of jags implies that the corresponding all-pole vocal tract model has roots on the positive real axis or at low frequencies, and, consequently, its amplitude spectrum shows boosting of low frequencies. This effect is clearly shown in the example by comparing the right panel of Fig. 2(a) to the corresponding panels in Figs. 2(b) and 2(c). It is worth emphasizing that the source-filter theory of voice production by Fant (1970) assumes that poles of the vocal tract for non-nasalized voiced sounds occur as complex conjugate pairs and the low-frequency emphasis of the vowel spectrum results from the glottal source. Therefore, it can be argued

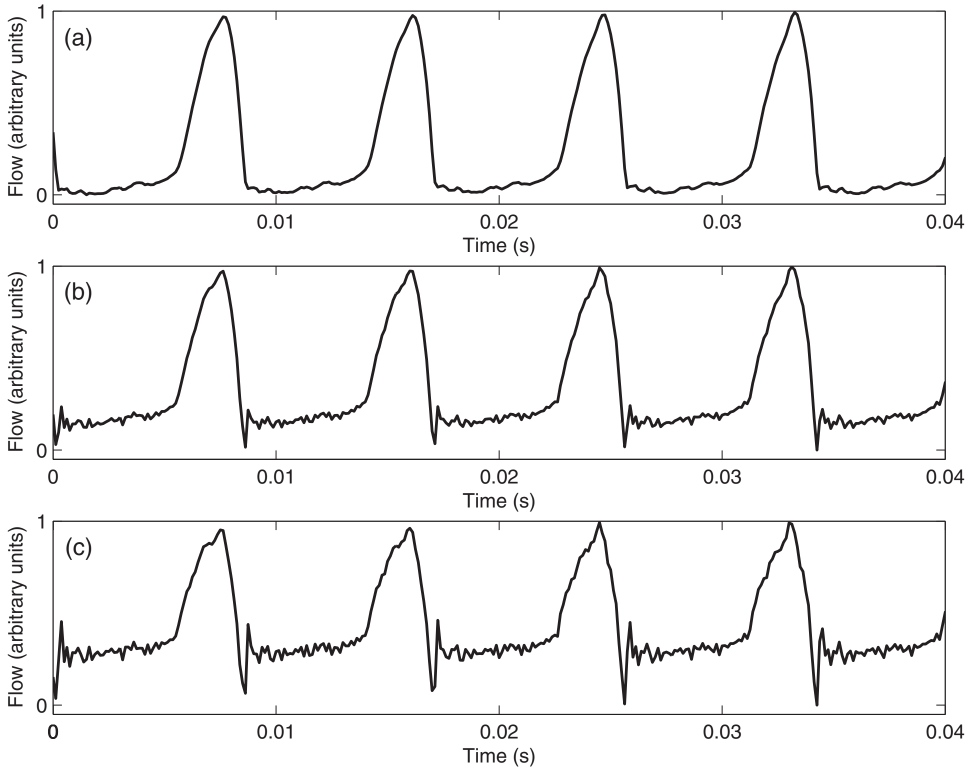


FIG. 1. Glottal flows estimated by IF the vowel [a] uttered by a male speaker by varying the position of the covariance frame in the CP analysis. The covariance frame was placed in the beginning of the CP using the differentiated EGG in panel (a), and its position was moved earlier by two samples in panel (b) and by four samples in panel (c).

that among the three vocal tract models computed by the CP analysis, the one depicted in Fig. 2(a) is the most plausible to represent an amplitude spectrum of an all-pole vocal tract of a vowel sound.

Quality of glottal flows computed by the CP analysis can be made less dependent on the position of the covariance frame by removing the roots of the vocal tract model located on the real axis (Wong *et al.*, 1979; Childers and Ahn, 1995). This is typically done by first solving the roots of the vocal tract model given by LP and then by removing those roots that are located on the positive real axis while preserving the roots on the negative real axis. This procedure was used for the example described in Figs. 1 and 2, and the results are shown in the time domain in Fig. 3 and in the frequency domain in Fig. 4. It can be seen that this standard procedure indeed decreased the distortion caused by the jags, as shown in Fig. 3(b). It is, however, worth noticing that this procedure is blind to complex roots located at low frequencies, which

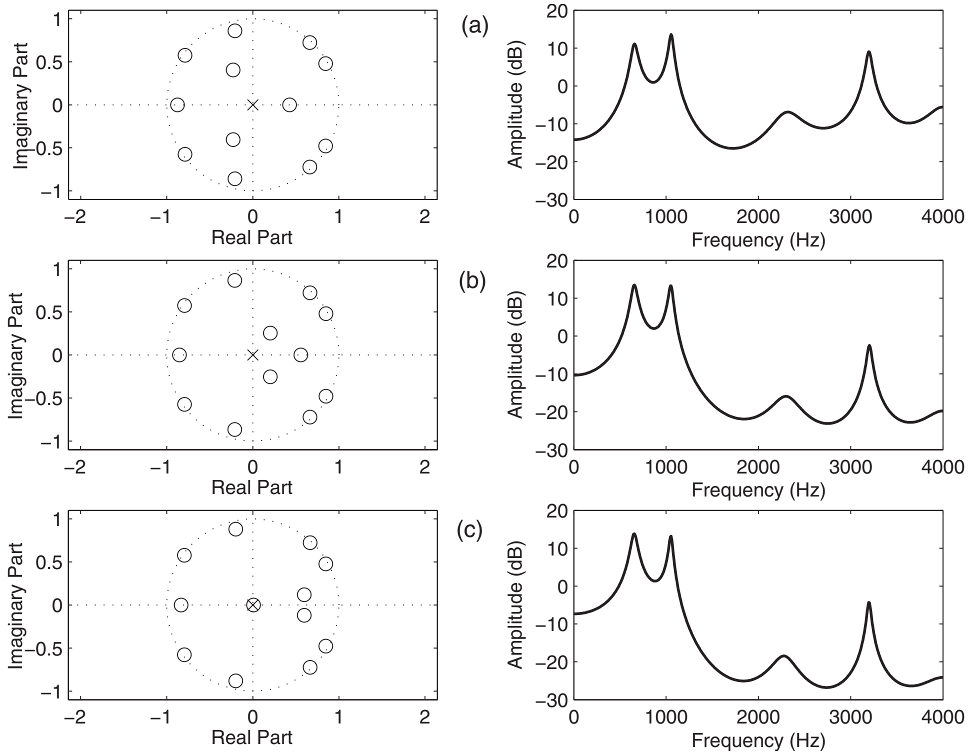


FIG. 2. Transfer functions of inverse filters in the *z*-domain (left panels) and the corresponding amplitude spectra of the all-pole vocal tract models (right panels) used in the CP analyses shown in Fig. 1.

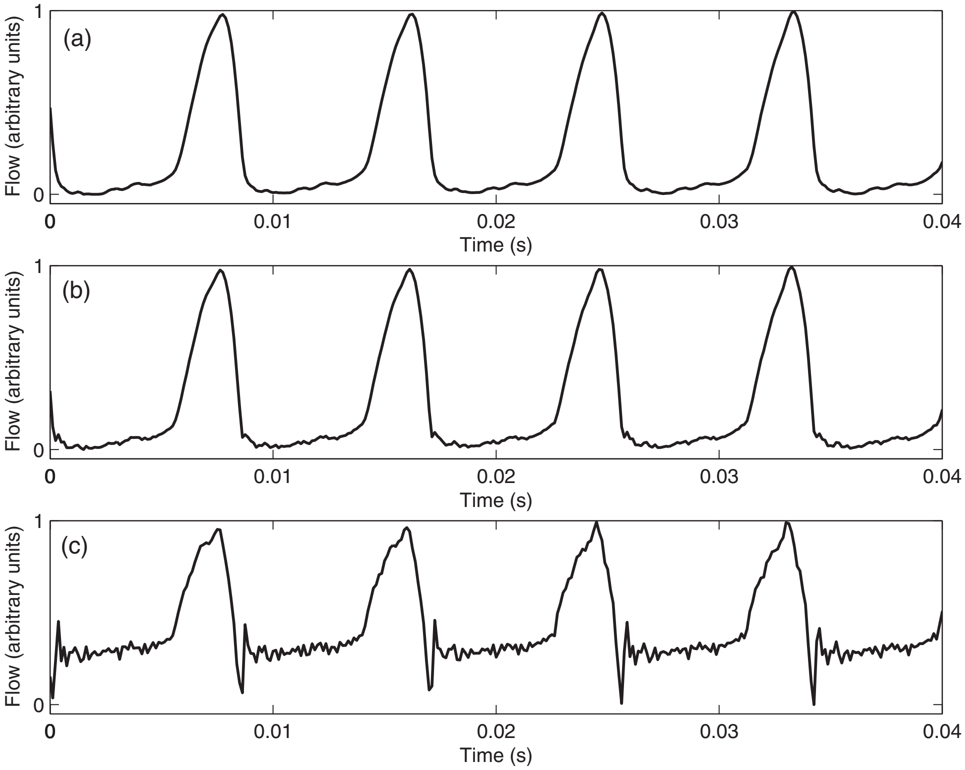


FIG. 3. Glottal flows estimated by IF the same [a] vowel used in Fig. 1. Roots located on the positive real axis were removed before IF. The covariance frame was placed in the beginning of the CP with the help of the differentiated EGG in panel (a), and its position was moved earlier by two samples in panel (b) and by four samples in panel (c).

cause distortion, described in Figs. 1(c) and 3(c), that might be even more severe than that resulting from the roots on the positive real axis.

In addition to the distortion caused by the occurrence of inverse filter real and complex roots at low frequencies as described above, the estimation of the glottal flow with the CP analysis might be affected by another issue. Namely, the computation of the linear predictive analysis with the covariance analysis might yield an inverse filter that is not minimum phase; that is, the filter has roots outside the unit circle in the *z*-domain. Although this property of the covariance analysis is well-known in the theory of LP (Rabiner and Schafer, 1978), it is, unfortunately, typically ignored in most glottal IF studies (exceptions are Akande and Murphy, 2005; Bozkurt *et al.*, 2005; Bäckström and Alku, 2006). A possible explanation of why the occurrence of non-minimum phase filters gets so little attention in glottal wave analysis is the fact that IF is always computed via FIR filtering. Hence,

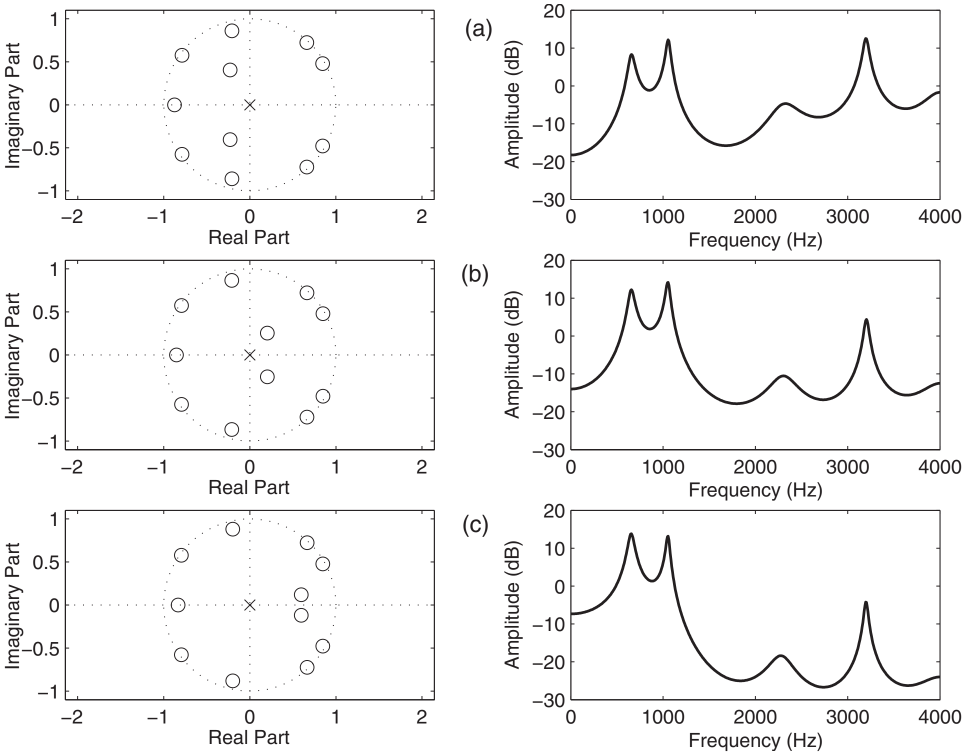


FIG. 4. Transfer functions of inverse filters in the *z*-domain (left panels) and the corresponding amplitude spectra of the all-pole vocal tract models (right panels) used in the CP analyses shown in Fig. 3.

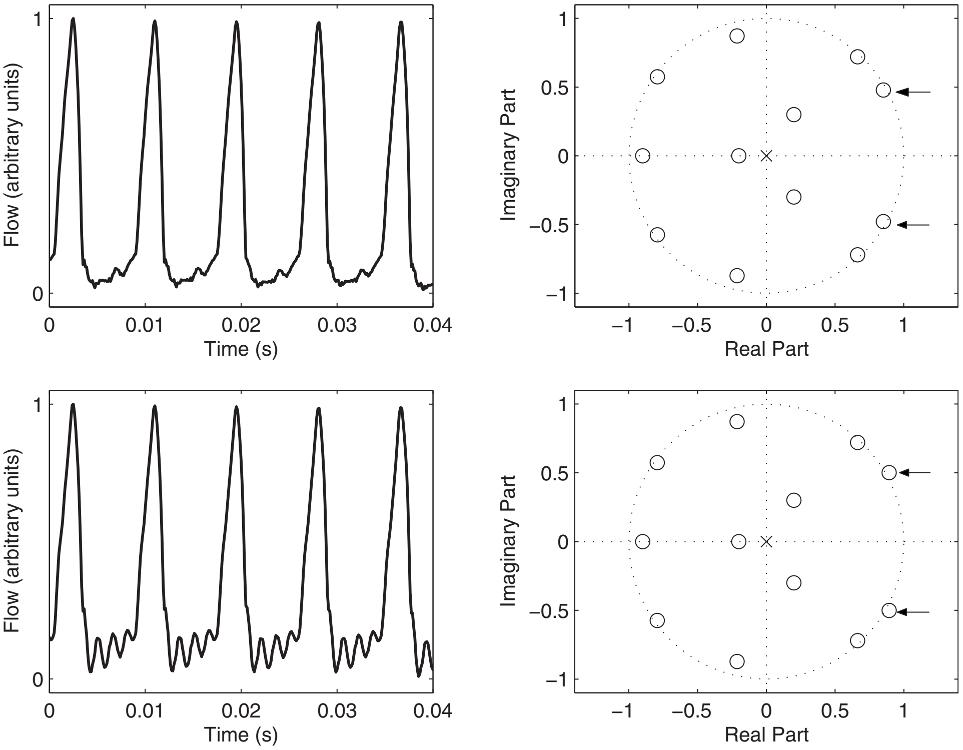


FIG. 5. Glottal flows estimated by the CP analysis (left panels) and inverse filter transfer functions in the z-domain (right panels) in the case of (a) minimum phase and (b) non-minimum phase IF. Radii of all roots in minimum phase filtering are less than unity. In non-minimum phase filtering, the complex conjugate root pair indicated by arrows in panel (a) is replaced by its mirror image pair outside the unit circle. The root radius of the indicated complex conjugate pair is 0.98 in panel (a) and 1.02 in panel (b).

non-minimum phase filters do not cause stability problems, which, of course, would be the case if non-minimum phase filters were used in all-pole synthesis, such as in speech coding or synthesis. Even though stability problems are not met in glottal in IF, the use of non-minimum phase inverse filters does cause other kinds of artifacts, as demonstrated below.

According to the source-filter theory of speech production, the glottal flow is filtered by a physiological filter, the vocal tract, which is regarded as a stable all-pole system for vowels and liquids. In the z-domain, this kind of system must have all its poles inside the unit circle (Oppenheim and Schafer, 1989). An optimal inverse filter cancels the effects of the vocal tract by matching each pole inside the unit circle with a zero of a FIR filter. However, it is well-known in the theory of digital signal processing that zeros of a FIR filter can be replaced by their mirror image partners; that is, a zero at $z = z_1$ is replaced by $z = 1/z_1^*$, without changing the shape of the amplitude spectrum of the filter (Oppenheim and Schafer, 1989). In other words, an inverse filter that is minimum phase can be replaced with a non-minimum phase FIR by replacing any of its roots with a corresponding mirror image outside the unit circle without changing the shape of inverse filter's amplitude response. Therefore, from the point of view of canceling the amplitude response of the all-pole vocal tract, there are several inverse filters, of which one is minimum phase and others are non-minimum phase, that can be considered equal. These candidates are, however, different in terms of their phase characteristics, and canceling the effects of an all-pole vocal tract with a non-minimum phase inverse filter produces phase distortion, which might severely affect the shape of the glottal flow estimate. This distortion is especially strong in cases where zeros in the inverse filter located in the vicinity of the lowest two formants are moved from inside the unit circle to the outside. Figure 5 shows an example of this effect. In Fig. 5(a), a glottal flow estimated with a minimum phase inverse filter is shown in the left panel, and the z-plane representation of the corresponding inverse filter is shown on the right. This inverse filter was deliberately modified by replacing one complex conjugate root pair located inside the unit circle by its corresponding mirror image pair located outside the circle. The root pair selected corresponds to the inverse model of the first formant and is represented in the z-plane graph of Fig. 5(a) by the complex conjugate pair having the lowest angular frequency (indicated by arrows). Even though the modification caused only a minor change in the root radius (original radius: 0.98, modified radius: 1.02), the change from the minimum phase structure into the non-minimum phase form is manifested as increased ripple during the CP of the glottal cycle, as shown in the left panel of Fig. 5(b).

B. The improved CP analysis

A new approach is proposed in the present study to compute IF with the CP analysis. The proposed technique aims to reduce the effects of the two major artifacts, occurrence of low-frequency roots of the inverse filter and occurrence of inverse filter roots outside the unit circle, described in the previous section. The main part of the method, to be described in Sec. II B 1, is represented by a new mathematical algorithm to define a linear predictive inverse filter. The novel way to compute vocal tract inverse filters is then combined, as described in Sec. II B 2, with an additional processing stage to yield the new glottal IF algorithm.

1. Computation of the vocal tract inverse filter with constrained linear prediction

The conventional CP analysis involves modeling the vocal tract with an all-pole filter defined according to the classical LP based on the covariance criterion (Rabiner and Schafer, 1978). The filter coefficients of pth order inverse filter are searched for by using a straightforward optimization where the energy of the prediction error is minimized over the covariance frame. In principle, this kind of optimization based on the mean square error (MSE) criterion treats all the frequencies equally, and the filter coefficients are mathematically adjusted so that the resulting all-pole spectrum accurately matches the high-energy formant regions of the speech spectrum (Makhoul, 1975). However, it is worth emphasizing that the conventional covariance analysis does not use any additional information in the optimization process, for example, to bias the location of roots of the resulting all-pole filter. This inherent feature of the conventional covariance analysis implies that roots of the resulting all-pole model of the vocal tract might be located in such a position in the z-domain (e.g., on the positive real axis) that is correct from the point of view of MSE-based optimization but unrealistic from the point of view of the source-filter theory of vowel production and its underlying theory of tube modeling of the vocal tract acoustics. In his fundamental work, Fant (1970) related vocal tract shapes derived from x-rays to the acoustic theory of different tube shapes and developed the source-filter theory of speech production. According to this theory, the transfer function of voiced speech, defined as the ratio of the Laplace transforms of the sound pressure at the lips to the glottal volume velocity, includes only complex poles in the s-domain. According to the discrete time version of this theory (e.g., Markel and Gray, 1976), the z-domain transfer function of the vocal tract is expressed for vowel sounds as an all-pole filter of order $2K$, which models K formants as a cascade of K second order blocks, each representing an individual resonance of a certain center frequency and bandwidth. In other words, there might be a mismatch in root locations of vocal tract filters between those optimized by the conventional covariance analysis and those assumed both in the source-filter theory and its underlying acoustical theory of tube shapes. It is likely that this mismatch becomes prevalent especially in cases when the covariance frame consists of a small number of data samples. Hence, the phenomenon discussed is related to the sensitivity of the CP analysis about the position of the covariance frame, a drawback discussed in several previous studies (e.g., Larar *et al.*, 1985; Veeneman and BeMent, 1985; Yegnanarayana and Veldhuis, 1998; Riegelsberger and Krishnamurthy, 1993).

Based on the concept of *constrained* LP, the computation of the conventional covariance analysis, however, can be modified in order to reduce the distortion that originates from such vocal tract model roots that are located in unrealistic positions in the z-domain. The key idea is to impose such restrictions on the linear predictive polynomials *prior* to the optimization that can be justified by the source-filter theory of voice production. Intuitively, this means that instead of allowing the linear predictive model to locate its roots freely in the z-domain based solely on the MSE criterion, the optimization is given certain restrictions in the predictor structure, which then result in more realistic root locations. In order to implement restrictions that end up in equations which can be solved in closed form, one has to first find a method to express the constraint in a form of a concise mathematical equation and then use the selected equation in the minimization problem. One such convenient constraint can be expressed with the help of the dc gain of the linear predictive inverse filter. The rationales to apply this quantity are as follows. First, the dc gain of a digital FIR filter can be expressed in a very compressed and mathematically straightforward manner as a linear sum of the predictor coefficients [see Eq. (4) below]. Consequently, the optimization of the constrained linear predictive filter is mathematically straightforward, ending up with a matrix equation [see Eq. (9)] that can be solved noniteratively in a similar manner as the corresponding normal equations of the conventional LP. Second, it is known from the classical source-filter theory of voice production that the vocal tract transfer function of nonnasalized sounds approaches unity at zero frequency provided that the losses through vibration of the cavity walls are small (Fant, 1970, pp. 42–44). In conventional LP, the dc gain of the inverse filter is not constrained, and, consequently, it is possible that the amplitude response of the vocal tract model computed by the covariance analysis shows excessive boost at zero frequency. If the covariance frame is short and placed incorrectly, it might even happen that the amplitude response of the obtained vocal tract model shows larger gain at zero frequency than at formants, which violates the assumptions of the source-filter theory and its underlying acoustical theory of tube shapes. Hence, by imposing a predefined constraint on the dc gain of the linear predictive inverse filter, one might expect to get such linear predictive vocal tract models whose amplitude response shows better correspondence with Fant's source-filter theory; that is, the transfer function indicates peaks at formant frequencies, while the gain at zero frequency is clearly smaller and approaches unity. It must be emphasized, however, that even though the proposed idea to assign the dc gain of the inverse filter into a pre-defined value is undoubtedly mathematically straightforward, this technique does not involve imposing explicit constraints on the root positions *per se* prior to the optimization. In other words, the exact z-domain root locations of the vocal tract model are still determined by the MSE-type optimization, yet the likelihood for these roots to become located in such positions that they create an excessive boost at low frequency is less than in the case of the conventional LP. Mathematical derivations to optimize the proposed idea of the dc-constrained LP will be described below.

In the conventional LP, the error signal, known as the residual, can be expressed in matrix form as follows:

$$e_n = x_n + \sum_{k=1}^{p} a_k x_{n-k} = \sum_{k=0}^{p} a_k x_{n-k} = \mathbf{a}^T \mathbf{x}_n, \qquad (1)$$

where $\mathbf{a}=[a_0,\ldots,a_p]^T$, with $a_0=1$, and the signal vector is $\mathbf{x}_n=[x_n \ldots x_{n-p}]^T$. The coefficient vector $\mathbf{a}$ is optimized according to the MSE criterion by searching for such parameters that minimize the square of the residual. In the covariance method, this minimization of the residual energy is computed over a finite time span (Rabiner and Schafer, 1978). By denoting this time span with $0 \leqslant n \leqslant N-1$, the prediction error energy $E(\mathbf{a})$ can be written as

$$E(\mathbf{a}) = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} \mathbf{a}^T\mathbf{x}_n\mathbf{x}_n^T\mathbf{a} = \mathbf{a}^T\left[\sum_{n=0}^{N-1} \mathbf{x}_n\mathbf{x}_n^T\right]\mathbf{a} = \mathbf{a}^T\mathbf{\Phi}\mathbf{a}, \tag{2}$$

where matrix $\mathbf{\Phi}$ is the covariance matrix defined from speech samples as

$$\mathbf{\Phi} = \sum_{n=0}^{N-1} \mathbf{x}_n\mathbf{x}_n^T \in R^{(p+1)\times(p+1)}. \tag{3}$$

It is worth noticing that the computation of matrix $\mathbf{\Phi}$ requires speech samples located inside the energy minimization frame, that is, x_n, where $0 \leqslant n \leqslant N-1$, plus p samples occurring before this frame, that is, x_n, where $-p \leqslant n < 0$. The optimal filter coefficients can be computed easily by minimizing the prediction error energy $E(\mathbf{a})$ with respect to the coefficient vector $\mathbf{a}$. This yields $\mathbf{a}=\sigma^2\mathbf{\Phi}^{-1}\mathbf{u}$, where $\boldsymbol{\sigma}^2 = (\mathbf{u}^T\mathbf{\Phi}^{-1}\mathbf{u})^{-1}$ is the residual energy given by the optimized predictor and $\mathbf{u}=[10\cdots0]^T$.

The conventional LP can be modified by imposing constraints on the minimization problem presented above. A mathematically straightforward way to define one such constraint is to set a certain pre-defined value for the frequency response of the linear predictive inverse filter at zero frequency. By denoting the transfer function of a pth order constrained inverse filter $C(z)$, the following equation can be written:

$$C(z) = \sum_{k=0}^{p} c_k z^{-k} \Rightarrow C(e^{j0}) = C(1) = \sum_{k=0}^{p} c_k = l_{\mathrm{dc}}, \tag{4}$$

where c_k, $0 \leqslant k \leqslant p$, are the filter coefficients of the constrained inverse filter and l_{dc} is a pre-defined real value for the gain of the filter at dc. Using matrix notation, the dc-constrained minimization problem can now be formulated as follows: minimize $\mathbf{c}^T\mathbf{\Phi}\mathbf{c}$ subject to $\mathbf{\Gamma}^T\mathbf{c}=\mathbf{b}$, where $\mathbf{c} = [c_0\cdots c_p]^T$ is the filter coefficient vector with $c_0=1$, $\mathbf{b} = [1 l_{\mathrm{dc}}]^T$, and $\mathbf{\Gamma}$ is a $(p+1)\times 2$ constraint matrix defined as

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ . & . \\ . & . \\ . & . \\ 0 & 1 \end{bmatrix}. \tag{5}$$

The covariance matrix defined in Eq. (3) is positive definite. Therefore, the quadratic function to be minimized in the dc-constrained problem is convex. Thus, in order to solve the minimization problem, the Lagrange multiplier method (Bazaraa *et al.*, 1993) can be used. This procedure begins with the definition of a new objective function,

$$\eta(\mathbf{c},\mathbf{g}) = \mathbf{c}^T\mathbf{\Phi}\mathbf{c} - 2\mathbf{g}^T(\mathbf{\Gamma}^T\mathbf{c} - \mathbf{b}), \tag{6}$$

where $\mathbf{g}=[g_1g_2]^T > \mathbf{0}$ is the Lagrange multiplier vector. The objective function of Eq. (6) can be minimized by setting its derivative with respect to vector $\mathbf{c}$ to zero. By taking into account that matrix $\mathbf{\Phi}$ is symmetric (i.e., $\mathbf{\Phi}=\mathbf{\Phi}^T$), this results in the following equation:

$$\nabla_c\,\eta(\mathbf{c},\mathbf{g}) = \mathbf{c}^T(\mathbf{\Phi}^T + \mathbf{\Phi}) - 2\mathbf{g}^T\mathbf{\Gamma}^T = 2\mathbf{c}^T\mathbf{\Phi} - 2\mathbf{g}^T\mathbf{\Gamma}^T$$
$$= 2(\mathbf{\Phi}\mathbf{c} - \mathbf{\Gamma}\mathbf{g}) = 0. \tag{7}$$

By combining Eq. (7) with the equation of the constraint (i.e., $\mathbf{\Gamma}^T\mathbf{c}-\mathbf{b}=0$), vector $\mathbf{c}$ can be solved from the group of equations

$$\mathbf{\Phi}\mathbf{c} - \mathbf{\Gamma}\mathbf{g} = 0,$$
$$\mathbf{\Gamma}^T\mathbf{c} - \mathbf{b} = 0, \tag{8}$$

which yields the optimal coefficients of the constrained inverse filter:

$$\mathbf{c} = \mathbf{\Phi}^{-1}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{\Phi}^{-1}\mathbf{\Gamma})^{-1}\mathbf{b}. \tag{9}$$

In summary, the optimal dc-constrained inverse filter, a FIR filter of order p given in Eq. (4) is obtained by solving for the vector $\mathbf{c}$ according to Eq. (9), in which the covariance matrix $\mathbf{\Phi}$ is defined by Eq. (3) from the speech signal x_n, matrix $\mathbf{\Gamma}$ is defined by Eq. (5), and matrix $\mathbf{b}=[1 l_{\mathrm{dc}}]^T$, where l_{dc} is the desired inverse filter gain at dc.

2. Checking the minimum phase property

In order to eliminate the occurrence of non-minimum phase filters, the roots of the inverse filter are solved, and if the filter is not minimum phase, those roots that are located outside the unit circle are replaced by their mirror image partners inside the circle. In principle, it is possible that the constrained LP computed according to Eq. (9) yields an inverse filter that has roots on the positive real axis. Due to the use of the dc constraint, the risk for this to happen is, however, clearly smaller than in the case of the conventional covariance analysis. Because the roots of $C(z)$ are solved for in order to eliminate the occurrence of non-minimum phase filters, it is trivial also to check simultaneously whether there are any roots on the positive real axis inside the unit circle. If so, these roots are simply removed, in a procedure similar to that used in the conventional CP analysis (Wong *et al.*, 1979).

3. Summary of the new algorithm

In summary, the new glottal IF algorithm can be presented by combining the procedures described in Secs. II B 1 and II B 2. The estimation of the glottal flow with this new CP-based IF algorithm consists of the following stages.

(1) Prior to the analysis, the speech pressure waveform is filtered through a linear-phase high-pass FIR with its cutoff frequency adjusted to 70 Hz. The purpose of this filter is to remove annoying low-frequency components picked up by the microphone during the recordings of the speech signals. The output of this stage, the high-pass filtered speech sound, is denoted by $S_{\mathrm{hp}}(n)$ below.

(2) The position of the covariance frame is computed using any of the previously developed methods based on, for

example, the maximum determinant of the covariance matrix (Wong *et al.*, 1979) or the EGG (Krishnamurthy and Childers, 1986).

(3) Vocal tract transfer function $C(z)$ is computed according to Eq. (9) by defining the elements of the covariance matrix in Eq. (3) from $S_{\mathrm{hp}}(n)$ by using the covariance frame defined in stage (2).

(4) Roots of $C(z)$ defined in stage (3) are solved. Those roots of $C(z)$ that are located outside the unit circle are replaced by their corresponding mirror image partner inside the unit circle. Any real roots located on the positive real axis are removed.

(5) Finally, the estimate of the glottal volume velocity waveform is obtained by filtering $S_{\mathrm{hp}}(n)$ through $C(z)$ defined in stage (4) and by canceling the lip radiation effect with a first order infinite impulse response filter, with its pole close to the unit circle (e.g., at $z=0.99$).

The algorithm runs in a frame-based manner, and the adjustable parameters are recommended to be set to values typically used in CP analysis: frame length: 50 ms; order of the vocal tract model: 12 (with sampling frequency of 8 kHz); the length of the covariance frame: 30 samples (a value that equals the order of the vocal tract model multiplied by 2.5). In the experiments conducted in the present study, the parameter l_{dc} used in the computation of the dc-constrained vocal tract inverse filters was adjusted so that the amplitude response of the vocal tract filter at dc was always equal to unity.[2]

III. MATERIALS AND EXPERIMENTS

In order to evaluate the performance of the new CP analysis technique, experiments were conducted using both natural and synthetic speech. The purpose of these experiments was to investigate whether the new modified covariance analysis based on the concept of constrained LP, when supplemented with the minimum phase requirement of the inverse filter, would make IF with the CP analysis less vulnerable to the position of the covariance frame.

A. Speech and EGG recordings

Simultaneous speech pressure waveform and EGG signals were recorded from 13 subjects (six females). The ages of the subjects varied between 29 and 43 (mean of 32), and none of them had experienced voice disorders. The speaking task was to produce the vowel [a] five times by using sustained phonation. Vowel [a] was used because it has a high first formant (F1).[3] Subjects were allowed to use the fundamental frequency of their own choice, but they were encouraged not to use a pitch that is noticeably higher than in their normal speech. The duration of each phonation was at least 1 s. The production was done by two types of phonation: normal and pressed. These two phonation types were selected because they are more likely to involve a CP in the vocal fold vibration, which would not be the case in, for example, breathy phonation (Alku and Vilkman, 1996). This, in turn, implies that the basic assumption of the CP analysis, that is, the existence of a distinct CP within the glottal cycle, should be valid. Consequently, using these two modes, one would expect to be able to demonstrate effectively the dependency of the CP analysis on the position of the covariance frame. The recordings were perceptually monitored by an experienced phonetician who trained the subjects to create the two registers properly. Phonations were repeated until the phonation type was satisfactory.

Speech pressure waves were captured by a condenser microphone (Brüel & Kjær 4188) that was attached to a sound level meter (Brüel & Kjær Mediator 2238) serving also as a microphone amplifier, and the EGG was recorded simultaneously (Glottal Enterprise MC2-1). The mouth-to-microphone distance was 40 cm. In order to avoid inconsistency in the synchronization of speech and EGG, the microphone distance was carefully monitored in the recordings, and its value was checked prior to each phonation. Speech and EGG waveforms were digitized using a (DAT) digital audio tape recorder (Sony DTC-690) by adopting the sampling rate of 48 kHz and the resolution of 16 bits.

The speech and EGG signals were digitally transferred from the DAT tape into a computer. Before conducting the IF analysis, the sampling frequency of both signals was down-sampled to 8 kHz. The propagation delay of the acoustic signal from the glottis to the microphone was estimated by using the vocal tract length of 15 and 17 cm for females and males, respectively, the mouth-to-microphone distance of 40 cm, and the speed of sound value of 350 m/s. These values yielded the propagation delay of 1.57 and 1.63 ms for female and male speakers, respectively. The fundamental frequency of each vowel sound was computed by searching for the peak of the autocorrelation function from the differentiated EGG signal. For female speakers, the mean F0 was 195 Hz (min: 178 Hz, max: 211 Hz) and 199 Hz (min: 182 Hz, max: 216 Hz) in normal and pressed phonation, respectively. For males, the mean F0 was 104 Hz (min: 90 Hz, max: 119 Hz) and 114 Hz (min: 95 Hz, max: 148 Hz) in normal and pressed phonation, respectively.

B. Synthetic vowels

A fundamental problem present both in developing new IF algorithms and in comparing existing methods is the fact that assessing the performance of an IF technique is complicated. When IF is used to estimate the glottal flow of natural speech, it is actually never possible to assess in detail how closely the obtained waveform corresponds to the true glottal flow generated by the vibrating vocal folds. It is, however, possible to assess the accuracy of IF by using synthetic speech that has been created using artificial glottal waveform. This kind of evaluation, however, is not truly objective because speech synthesis and IF analysis are typically based on similar models of the human voice production apparatus, for example, the traditional linear source-filter model (Fant, 1970).

In the current study, a different strategy was used in order to evaluate the performance of different CP analysis methods in the estimation of the glottal flow. The idea is to use *physical modeling* of the vocal folds and the vocal tract in order to simulate time-varying waveforms of the glottal

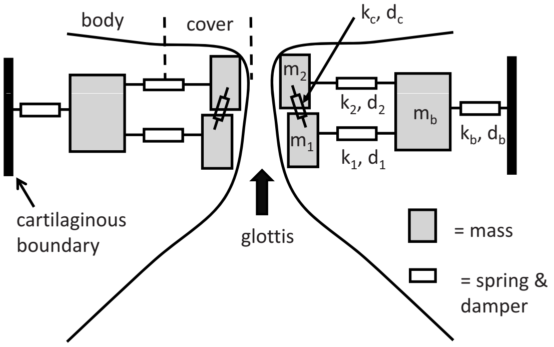


FIG. 6. Schematic diagram of the lumped-element vocal fold model. The cover-body structure of each vocal fold is represented by three masses that are coupled to each other by spring and damping elements. Bilateral symmetry was assumed for all simulations.

flow and radiated acoustic pressure. By using the simulated pressure waveform as an input to an IF method, it is possible to determine how closely the obtained estimate of the voice source matches the simulated glottal flow. This approach is different from using synthetic speech excited by an artificial form of the glottal excitation because the glottal flow waveform results from the interaction of the self-sustained oscillation of the vocal folds with subglottal and supraglottal pressures, as would occur during real speech production. Hence, the glottal flow waveform generated by this model is expected to provide a more stringent and realistic test of the IF method than would be permitted by a parametric flow waveform model where no source-tract interaction is incorporated.[4]

The sound pressure and glottal flow waveforms used to test the new IF technique were generated with a computational model of the vocal folds and acoustic wave propagation. Specifically, self-sustained vocal fold vibration was simulated with three masses coupled to one another through stiffness and damping elements (Story and Titze, 1995). A schematic diagram of the model is shown in Fig. 6, where the arrangement of the masses was designed to emulate the body-cover structure of the vocal folds (Hirano, 1974). The input parameters consisted of lung pressure, prephonatory glottal half-width (adduction), resting vocal fold length and thickness, and normalized activation levels of the cricothyroid (CT) and thyroarytenoid (TA) muscles. These values were transformed to mechanical parameters of the model, such as mass, stiffness, and damping, according to the "rules" proposed by Titze and Story (2002). The vocal fold model was coupled to the pressures and air flows in the trachea and vocal tract through aerodynamic and acoustic considerations as specified by Titze (2002), thus allowing for self-sustained oscillation. Bilateral symmetry was assumed for all simulations such that identical vibrations occur within both the left and right folds. Nine different fundamental frequency values (105, 115, 130, 145, 205, 210, 230, 255, and 310 Hz), which roughly approximate the ranges typical of adult male and female speech (e.g., Hollien *et al.*, 1971; Hollien and Shipp, 1972; Stoicheff, 1981), were generated by modifying the resting vocal fold length and activation levels of the CT and TA muscles; all other input parameters were held constant. The input parameters for all nine cases are shown in Table I. Those cases with the resting length (L_o) equal to 1.6 cm were intended to be representative of the male F0 range, whereas those with $L_o=0.9$ cm were intended to be in the female F0 range.

Acoustic wave propagation in both the trachea and vocal tract was computed in time synchrony with the vocal fold model. This was performed with a wave-reflection approach (e.g., Strube, 1982; Liljencrants, 1985) where the area functions of the vocal tract and trachea were discretized into short cylindrical sections or tubelets. Reflection and transmission coefficients were calculated at the junctions of consecutive tubelets, at each time sample. From these, pressure and volume velocity were then computed to propagate the acoustic waves through the system. The glottal flow was determined by the interaction of the glottal area with the time-varying pressures present just inferior and superior to the glottis as specified by Titze (2002). At the lip termination, the forward and backward traveling pressure wave components were subjected to a radiation load modeled as a resistance in parallel with an inductance (Flanagan, 1972), intended to approximate a piston in an infinite plane baffle. The output pressure is assumed to be representative of the pressure radiated at the lips. To the extent that the piston-in-a-baffle reasonably approximates the radiation load, the calculated output pressure can also be assumed to be representative of the pressure that would be transduced by a microphone in a non-reflective environment. The specific implementation of the vocal tract

TABLE I. Input parameters for the vocal fold model used to generate the nine different fundamental frequencies. Notation is identical to that used in Titze and Story (2002). The a_{CT} and a_{TA} are normalized activation levels (can range from 0 to 1) of the CT and TA muscles, respectively. L_o and T_o are the resting length and thickness of the vocal folds, respectively. ξ_{01} and ξ_{02} are the prephonatory glottal half-widths at the inferior and superior edges of vocal folds, respectively, and P_L is the respiratory pressure applied at the entrance of the trachea (see Fig. 7). The value of P_L shown in the table is equivalent to a pressure of 8 cm H_2O.

	Fundamental frequency (Hz)								
Parameter value	105	115	130	145	205	210	230	255	310
a_{CT}	0.1	0.4	0.1	0.4	0.2	0.3	0.3	0.4	0.7
a_{TA}	0.1	0.1	0.4	0.4	0.2	0.2	0.3	0.4	0.5
L_o (cm)	1.6	1.6	1.6	1.6	0.9	0.9	0.9	0.9	0.9
T_o (cm)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
ξ_{01} (cm)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
ξ_{02} (cm)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
P_L (dyn/cm^2)	7840	7840	7840	7840	7840	7840	7840	7840	7840

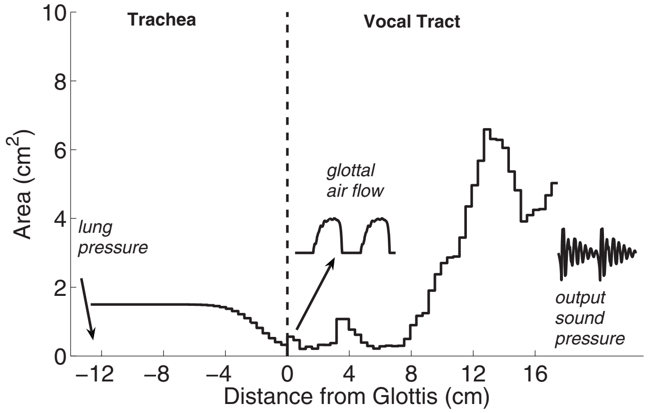


FIG. 7. Area function representation of the trachea and vocal tract used to simulate the male [a] vowel. The vocal fold model of Fig. 6 would be located at the 0 cm point indicated by the dashed vertical line. Examples of the glottal flow and output pressure waveforms are shown near the locations at which they would be generated.

model used for this study was presented in Story (1995) and included energy losses due to viscosity, yielding walls, heat conduction, as well as radiation at the lips.

In the model, a specific vocal tract shape is represented as an area function. For this study, glottal flow and output pressure waveforms were generated based on the area function for the [a] vowel reported by Story *et al.* (1996). For simulations of this vowel with the four lowest fundamental frequencies (105, 115, 130, and 145 Hz), the vocal tract length was set to 17.46 cm. For the five higher F0 speech simulations, exactly the same [a] vowel area function was used, but the length was non-uniformly scaled to 14.28 cm with scaling factors based on those reported by Fitch and Giedd (1999). The purpose of the shortened tract length was to provide an approximation of a possible female-like vocal tract to coincide with the higher F0 simulations. Although a measured female area function could have been used (e.g., Story, 2005), scaling the length of the male [a] vowel was done so that all cases resulted from fairly simple modifications of the same basic model.

A conceptualization of the complete model is given in Fig. 7, where the vocal fold model is shown to be located between the trachea and the vocal tract. The vocal tract is shown configured with the shape and length of the adult male [a] vowel, and the trachea is a uniform tube with a cross-sectional area of 1.5 cm^2 but tapered to 0.3 cm^2 near the glottis. An example glottal flow waveform is indicated near the middle of the figure. Note that the ripples in the waveform are largely due to interaction of the flow with the formant oscillations in the vocal tract. The coupling of the trachea to the vocal tract (via glottal area), however, will slightly alter the overall resonant structure of the system and, hence, will also contribute to glottal waveform shape. The sound pressure waveform radiated at the lips is also shown at the lip end of the area function and, as mentioned previously, can be considered analogous to a microphone signal recorded for a speaker.

In summary, the model is a simplified but physically-motivated representation of a speaker in which glottal airflow and output pressure waveforms result from self-sustained oscillation of the vocal folds and their interaction with propagating pressure waves within the trachea and vocal tract. The model generates both the signal on which IF is typically performed (microphone signal) and the signal that it seeks to determine (glottal flow), thus providing a reasonably realistic test case for IF algorithms.

C. Experiments

Four representative examples of glottal flow pulse forms computed by the proposed CP algorithm are shown in Fig. 8.

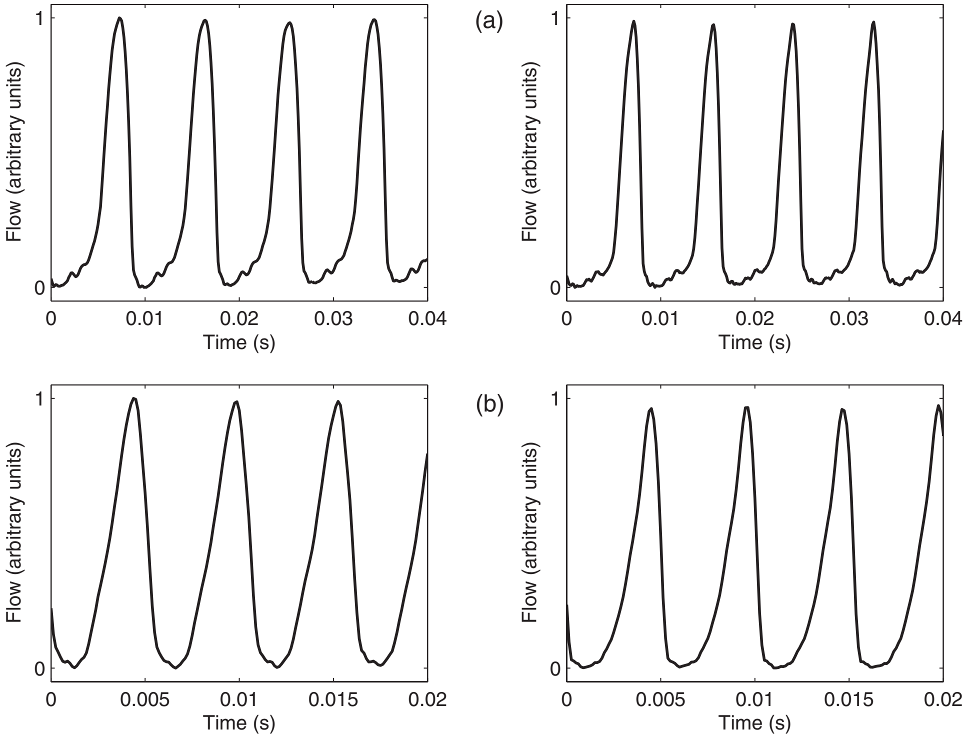


FIG. 8. Examples of glottal flows estimated by the proposed CP$_{con}$ method. IF was computed from [a] vowels produced by a male (panels a) and a female (panels b) speaker using normal (left panels) and pressed (right panels) phonation.

The examples shown in Figs. 8(a) and 8(b) were computed from a male and female speaker, respectively, by using both normal and pressed phonations of the vowel [a]. All these estimates of the glottal excitation were computed by using parameter values given at the end of Sec. II B 3. The beginning of the covariance frame was adjusted to a time instant three samples after the negative peak of the EGG derivative. It can be seen in Fig. 8 that none of the estimated glottal pulse forms show abrupt high amplitude peaks at the end of the closing phase, indicating that inverse filter roots are most likely located correctly in the formant region rather than in unrealistic positions at low frequency. CP can be identified rather easily from all the examples shown. However, the waveforms estimated from utterances spoken by the male speaker show a small ripple component. This ripple might be due to incomplete canceling of some of the higher formants by the inverse filter. Alternatively, this component might be explained by the existence of nonlinear coupling between the source and the tract, which cannot be taken into account in CP analysis because it is based on linear modeling of the voice production system.

The performance of the proposed CP analysis algorithm was tested by conducting two major experiments, one of which used synthetic vowels and the other natural speech. Both experiments involved estimating the glottal flow with three CP analysis types. The first one, denoted by $\mathrm{CP}_{\mathrm{bas}}$ in the rest of the paper, is represented by the basic CP analysis in which the vocal tract model computed by the covariance analysis is used as such in IF. The second one, denoted by $\mathrm{CP}_{\mathrm{rem}}$, is the most widely used form of the CP analysis in which the roots of the inverse filter polynomial computed by the covariance analysis are solved, and those located on the positive real axis are removed before IF. The third type, denoted by $\mathrm{CP}_{\mathrm{con}}$, is the proposed method based on the constrained LP described in Sec. II B.

In both experiments, the robustness of each CP analysis to the position of the covariance frame was evaluated by varying the beginning of the frame position near its optimal value, n_{opt}, the instant of glottal closure. For synthetic vowels, n_{opt} was first adjusted by using the derivative of the flow pulse generated by the physical vocal fold model. In this procedure, the optimal beginning of the covariance frame was set to the time instant after the negative peak of the flow derivative when the waveform returns to the zero level. For each synthetic vowel, the beginning of the covariance frame was then varied in 11 steps by defining the start index as $n = n_{\mathrm{opt}} + i$, where $i = -5$ to $+5$. (In other words, the optimal frame position corresponds to index value $i = 0$.) For natural vowels, the position of the covariance frame was varied by first extracting the glottal closure as the time instant when the EGG derivative reached a negative peak within a glottal cycle. Again, 11 frame positions were analyzed around this instant of glottal closure.

For synthetic sounds, there is no variation between periods, and, therefore, only a single cycle was analyzed. The total number of CP analyses conducted for synthetic speech was 297 (3 CP methods × 9 F0 values × 11 frame positions per cycle). For natural vowels, the analysis was repeated for six consecutive glottal cycles. Hence, the total number of CP analyses conducted for natural speech was 5148 (3 CP methods × 2 phonation types × 13 speakers × 11 frame positions per cycle × 6 cycles). The estimated glottal flows were parametrized using two frequency-domain measures. The first of these, H1H2, is defined as the difference in decibel between the amplitudes of the fundamental and the second harmonic of the source spectrum (Titze and Sundberg, 1992). The second parameter, the harmonic richness factor (HRF), is defined from the spectrum of the glottal flow as the difference in decibel between the sum of the harmonic amplitudes above the fundamental and the amplitude of the fundamental (Childers and Lee, 1991). (Notice the difference in the computation of the spectral ratio between the two parameters: if only the second harmonic is included in HRF, then its value becomes equal to H1H2 multiplied by −1.) These parameters were selected for two reasons. First, both of them can be computed automatically without any user adjustments. In CP analysis with a varying frame position, this is highly justified because the glottal flow waveforms, especially those computed with $\mathrm{CP}_{\mathrm{bas}}$, are sometimes so severely distorted that their reliable parametrization with, for example, time-based glottal flow quotients is not possible. Second, both H1H2 and HRF are known to reflect the spectral decay of the glottal excitation: a slowly decaying source spectrum is reflected by a small H1H2 and a large HRF value. Hence, if the glottal flow estimate is severely distorted by artifacts seen as jags in the closing phase, as shown in Figs. 1(b), 1(c), and 3(c), one is expected to get a decreased H1H2 value and an increased HRF value because the spectrum of the distorted glottal flow approaches that of the impulse train, that is, a flat spectral envelope. Since HRF takes into account a larger number of spectral harmonics, one can argue that its value reflects more reliable changes in the glottal flow than H1H2. Therefore, HRF alone might represent a sufficient spectral parameter to be used from the point of view of the present study. H1H2 is, however, a more widely used parameter in voice production studies, which justifies its selection as an additional voice source parameter in the present investigation.

IV. RESULTS

A. Experiment 1: Synthetic vowels

Robustness of the different CP analyses to the covariance frame position is demonstrated for the synthetic vowels by the data given in Table II. H1H2 and HRF values were first computed in each covariance frame position with each of the three CP techniques. For both H1H2 and HRF, the difference between the parameters extracted from the original flow and the estimated flow was computed. The data in Table II show the absolute value of this difference computed as an average pooled over 11 frame positions. The obtained results indicate that the error in both H1H2 and HRF due to the variation of the CP frame position is smallest for all vowels with F0 less than 310 Hz when IF is computed with the proposed new method. The average value of H1H2, when pooled over all vowels with F0 less than 310 Hz, equaled to 2.6, 0.9, and 0.5 dB for $\mathrm{CP}_{\mathrm{bas}}$, $\mathrm{CP}_{\mathrm{rem}}$, and $\mathrm{CP}_{\mathrm{con}}$, respectively. For HRF, the average value equaled to 7.8, 3.8, and 2.4 dB for $\mathrm{CP}_{\mathrm{bas}}$, $\mathrm{CP}_{\mathrm{rem}}$, and $\mathrm{CP}_{\mathrm{con}}$, respectively. For the synthetic

TABLE II. Effect of the covariance frame position on H1H2 and HRF using vowels synthesized by physical modeling. Absolute value of the difference (in dB) was computed between parameter values extracted from the original flows and from the glottal flows estimated by IF. Inverse filtering was computed by three CP algorithms: CP_{bas}, CP_{rem}, and CP_{con}. Data were averaged over 11 different frame positions starting around the instant of glottal closure.

F0 (Hz)	Diff in H1H2 (dB)			Diff in HRF (dB)		
	CP_{bas}	CP_{rem}	CP_{con}	CP_{bas}	CP_{rem}	CP_{con}
105	1.36	0.06	0.03	5.08	2.30	1.87
115	2.93	0.14	0.08	9.15	2.37	1.75
130	1.81	0.13	0.06	6.14	2.23	1.63
145	3.42	0.10	0.07	10.36	1.74	1.59
205	2.98	1.66	0.83	8.80	6.13	2.72
210	2.67	1.40	0.82	8.06	5.91	2.91
230	3.17	1.28	0.90	8.31	4.00	3.21
255	2.40	2.13	1.24	6.35	5.44	3.40
310	0.69	0.69	3.38	6.24	6.24	5.02

vowel with the largest F0 value, the best result was also given by CP_{con} when the parametrization was performed with HRF. However, H1H2 indicated a surprisingly small error for this high-pitch vowel when IF was conducted with CP_{bas} and CP_{rem}. The waveforms, however, were greatly distorted, but the levels of the fundamental and the second harmonic, that is, those sole spectral components used in the computation of H1H2, were only marginally affected. It is, though, worth emphasizing that the length of the glottal CP for this high-pitch vowel with $F0=310$ Hz is only ten samples (1.25 ms). This implies that the underlying assumption underlying all the three assessed IF techniques, that is, the existence of sufficiently long CP, is greatly violated. Hence, the surprisingly small value of H1H2 difference for this signal is explained mainly by the shortcomings of the simple spectral parameter rather than by the successful voice source estimation. In summary, the experiments conducted with synthetic vowels indicate that the proposed CP algorithm was the least vulnerable to the covariance frame position among the three techniques when voices of different F0 were compared.

B. Experiment 2: Natural vowels

The standard deviations (std) and means of the H1H2 and HRF values extracted from the glottal flows computed from natural vowels of varying covariance frame positions were compared with repeated measures analyses of variance (ANOVAs). The data were analyzed with sex × method × phonation ANOVAs where "sex" included male and female sexes, factor "method" included three different CP algorithms, CP_{bas}, CP_{rem}, and CP_{con}, and factor "phonation" included phonation types normal and pressed. H1H2 and HRF data were analyzed with separate ANOVAs, and Newman–Keuls tests were used as a means of *post hoc* analysis for pairwise differences in the data. The standard deviations and mean values of H1H2 and HRF obtained from the 66 window positions (11 frame positions of 6 cycles) are shown in Fig. 9. The main and interaction effects of the corresponding ANOVA results are given in Table III.

The standard deviation of both H1H2 and HRF differed significantly between the IF methods. *Post hoc* analyses showed that the standard deviations of H1H2 and HRF were, on the average, smaller when the new CP method

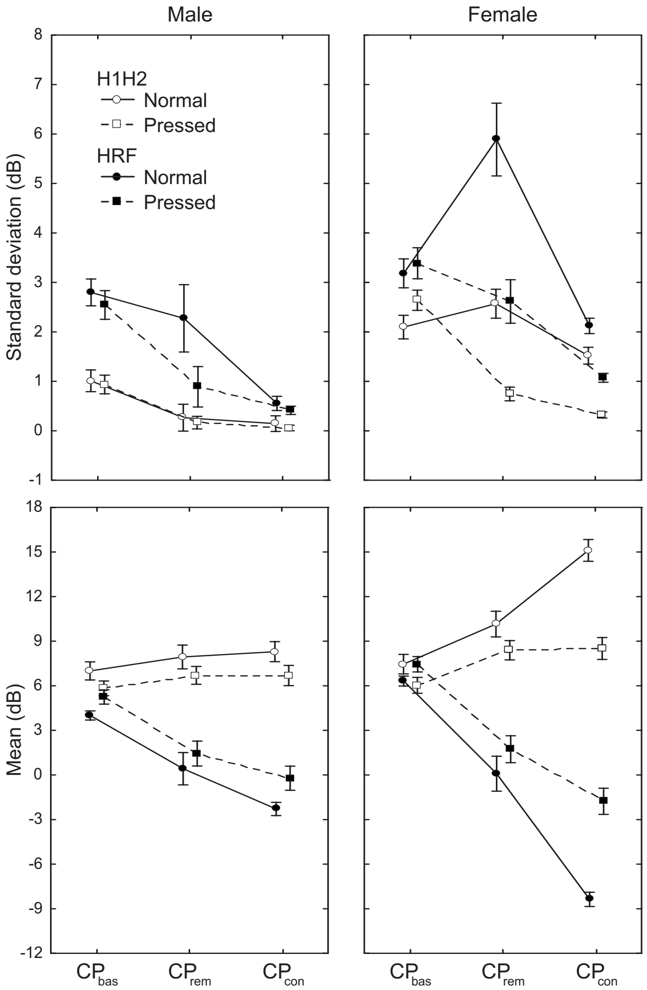


FIG. 9. Standard deviations (top panels) and means (bottom panels) of H1H2 and HRF according to the speaker sex and the type of phonation for CP analyses computed by CP_{bas}, CP_{rem}, and CP_{con}. Error bars represent standard error of the mean.

TABLE III. ANOVA results for standard deviations and means of H1H2 (upper table) and HRF (lower table). The degrees of freedom (DF), Greenhouse–Geisser epsilons (ε), F-values, and the associated probability (p) values are shown for each ANOVA effect. Analyses were conducted for utterances produced by 13 speakers using 11 covariance frame positions per glottal cycle and 6 successive periods.

		Standard deviations			Means		
H1H2 Effects and degrees of freedom (Df1, Df2)		ε	F	p	ε	F	p
Sex	1, 11		83.86	<0.001		9.38	<0.05
Method	2, 22	0.73	47.22	<0.001	0.69	88.85	<0.001
Method × sex	2, 22	0.73	4.12	<0.05	0.69	37.70	<0.001
Phonation	1, 11	1.00	10.42	<0.01	1.00	20.43	<0.001
Phonation × sex	1, 11	1.00	6.80	<0.05	1.00	3.58	ns
Method × phonation	2, 22	0.95	16.26	<0.001	0.68	23.22	<0.001
Method × phonation × sex	2, 22	0.95	15.57	<0.001	0.68	16.81	<0.001
		Standard deviations			Means		
HRF Effects and degrees of freedom (Df1, Df2)		ε	F	p	ε	F	p
Sex	1, 11		49.33	<0.001		0.83	ns
Method	2, 22	0.61	26.43	<0.001	0.87	262.27	<0.001
Method × sex	2, 22	0.61	6.26	<0.05	0.87	30.99	<0.001
Phonation	1, 11	1.00	18.12	<0.01	1.00	15.22	<0.01
Phonation × sex	1, 11	1.00	2.86	ns	1.00	2.06	ns
Method × phonation	2, 22	0.77	12.83	<0.001	0.67	9.29	<0.01
Method × phonation × sex	2, 22	0.77	3.10	ns	0.67	4.58	<0.05

ns = not significant

(H1H2-std=0.5, HRF-std=1.0) was used than when either CP_{bas} (H1H2-std=1.6, HRF-std=3.0) or CP_{rem} (H1H2-std =0.9, HRF-std=2.8) was used.

For H1H2, the difference between CP_{bas} and CP_{rem} was also significant. Additional effects on H1H2 and HRF variability were observed for sex and phonation type. The H1H2 and HRF standard deviations were larger for female (H1H2-std=1.6, HRF-std=3.0) than for male (H1H2-std =0.4, HRF-std=1.6) speakers. Further, the variability of H1H2 and HRF was larger for the normal (H1H2-std=1.2, HRF-std=2.7) than for the pressed (H1H2-std=0.8, HRF-std=1.8) type of phonation. Finally, significant method × sex, method × phonation, and method × phonation × sex interactions were found for both H1H2 and HRF, and a phonation × sex interaction was additionally significant for the H1H2.

The results indicated a statistically significant effect of CP method on the mean H1H2 and HRF values. The mean H1H2 and HRF values increased and decreased, respectively, when the IF algorithm CP_{bas} (H1H2=6.6 and HRF=5.7) was changed to CP_{rem} (H1H2=8.2 and HRF=0.9) and, then, further to the new CP_{con} algorithm (H1H2=9.5 and HRF= −3.0). While HRF mean values were similar for both sexes, the average H1H2 values were larger for female (9.3) than for male (7.1) speakers. Additionally, a smaller mean H1H2 and a larger mean HRF value was observed for the pressed phonation (H1H2=7.0 and HRF=2.3) than for the normal phonation (H1H2=9.2 and HRF=0.1). Finally, significant method × sex, method × phonation, and method × phonation × sex interactions were found for both H1H2 and HRF data.

V. CONCLUSIONS

CP covariance analysis, a widely used glottal IF method, computes a parametric model of the vocal tract by conducting linear predictive analysis over a frame that is located in the CP of the glottal cycle. Since the length of the CP is typically short, the resulting all-pole model is highly vulnerable with respect to the extraction of the frame position. Even a minor change in the frame position might greatly affect the z-domain locations of the roots of the all-pole model given by LP. This undesirable feature of the conventional CP analysis typically results in vocal tract models, which have roots, both real and complex, at low frequencies or roots that are located outside of the unit circle. These kinds of false root locations, in turn, result in distortion of the glottal flow estimates, which is typically seen as unnatural peaks at the instant of glottal closure, the so-called jags, or as increased formant ripple during the CP.

The present study proposed an improved version of the CP analysis based on a combination of two algorithmic issues. First, and most importantly, a constraint is imposed on the dc gain of the inverse filter prior to the optimization of the coefficients. With this constraint, linear predictive analysis is more prone to give vocal tract models that can be justified from the point of view of the source-filter theory of vowel production; that is, they show complex conjugate roots in the vicinity of formant regions rather than unrealistic resonances at low frequencies. Second, the new CP method utilizes an inverse filter that is minimum phase, a property that is not typically used in glottal IF.

The new glottal IF method, $\mathrm{CP_{con}}$, was compared to two CP analysis techniques by using both synthetic vowels produced by physical modeling of the voice production apparatus and natural vowels produced by male and female speakers. In summary, the experiments conducted with synthetic vowels having F0 from 105 to 310 Hz indicate that the proposed CP method gave glottal flow estimates with better robustness to the covariance frame position than the conventional CP methods. The result suggests that the parametric model of the vocal tract computed with the dc-constrained linear predictive analysis is less prone to distortion by the problem typically met in the CP analysis, namely, the involvement of samples outside the CP in the computation of the vocal tract. This problem violates the basic assumption of the CP analysis that the estimation of the vocal tract transfer function is made during the excitation-free time span. It can be argued that this violation is larger for voices of high pitch because they typically show short CPs in the glottal excitation. Violation results in the occurrence of unjustified inverse filter roots at low frequencies, which, in turn, distorts the resulting glottal flow estimates. Based on the results achieved with synthetic speech, the involvement of the dc constraint in the optimization process of the vocal tract model, however, seems to reduce this distortion and hence improve the estimation robustness with respect to the CP frame position. It must be emphasized, though, that if the amount of data samples during the glottal CP becomes extremely small, which was the case in analyzing the vowel with $\mathrm{F0}=310$ Hz in the present investigation, distortion of the glottal flow estimates becomes large with all CP techniques.

The experiments conducted with natural speech indicate that the deviation of H1H2 and HRF due to the varying of the covariance frame position inside the glottal cycle was larger for female speech than for male vowels and the deviation was also larger in normal than in pressed phonation. These results are in line with findings reported in previous studies (e.g., Veeneman and BeMent, 1985) as well as with experiments conducted in the present investigation with synthetic speech, indicating that the robustness of the CP analysis with respect to the frame position tends to decrease for shorter CP intervals, as in higher F0 speech or in normal as opposed to pressed phonation. The proposed new CP method, importantly, gave the smallest deviation of H1H2 and HRF, suggesting that the involvement of the dc constraint reduces the sensitivity of the CP analysis to the covariance frame position and that this holds true also for natural vowels. This finding is also supported by the fact that the mean levels of H1H2 and HRF were found to be largest and smallest, respectively, when IF was computed with $\mathrm{CP_{con}}$. In other words, the average spectral decay of the glottal flow pulse forms computed by varying the frame position was steeper with $\mathrm{CP_{con}}$ than with the other two CP methods. This is explained by the frequency-domain effect produced by distortion represented by impulse-like jags: the larger their contribution, the flatter the spectrum.

In summary, the proposed IF method constitutes a potential means to compute the CP covariance analysis to estimate the glottal flow from speech pressure signals. It reduces distortion caused by one of the major drawbacks of the conventional CP analysis, the sensitivity of the analysis to the position of the covariance frame. The computational load of the new method is only slightly larger than that of the conventional CP method. In addition, the method can be implemented in a manner similar to the conventional one, that is, either based solely on the speech pressure signal or in a two-channel mode where an EGG signal is used to help extract the covariance frame position. Therefore, there are no obstacles in principle for the implementation of the proposed method in environments where the conventional analysis is used. One has to keep in mind, though, that the new method does not change the basic assumptions of the CP analysis, namely, that the voice source and vocal tract are linearly separable, and there is a CP of finite duration during which there is no excitation by the source of the tract.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (Project No. 111848) and by the COST action 2103, “Advanced Voice Function Assessment.”

[1] It is worth emphasizing that glottal pulses estimated from natural speech sometimes show fluctuation, typically referred to as “ripple,” after the instant of glottal closure. This component might correspond to actual phenomena or it may result from incorrect inverse filter settings. If the pulse waveform is fluctuating after the instant of the glottal closure, it is, though, difficult, if not impossible, to define accurately which part of the fluctuation corresponds to real phenomena and which part results from incorrect IF. If, however, the flow waveform shows an abrupt peak at the end of the closing phase, such as in Fig. 1(c), and if this component is removed by, for example, a minor change in the position of the analysis frame, it is more likely that the component represents an artifact than a real phenomenon.

[2] By using Eq. (4), the gain of the vocal tract filter at dc, denoted by G_{dc}, is defined as the absolute value of the inverse of the frequency response of the constrained predictor at $\omega=0$: $G_{dc}=|1/C(e^{j0})|=|1/l_{dc}|$. In principle, the requirement $G_{dc}=1$ can be satisfied by assigning either $l_{dc}=1$ or $l_{dc}=-1$. Although both of these values result in vocal tract filters of equal gain at dc, they end up as different constrained transfer functions. In order to test the difference between the two values of l_{dc}, the glottal flows were estimated from the synthetic vowels described in Sec. III B by using H1H2 and HRF parameters described in Sec. III C and by conducting the constrained IF analysis by assigning both $l_{dc}=1$ and $l_{dc}=-1$. The results indicated clearly that the choice $l_{dc}=-1$ yielded glottal flow estimates that were closer to the original flows generated by the physical modeling approach.

[3] In the area of glottal IF, most studies analyze vowels with high first formant such as [a] or [ae]. The reason for this is the fact that the separation of the source and the tract becomes increasingly difficult from a mathematical point of view if the first formant is low. This is due to the fact that the strong harmonics at low frequencies bias the estimation of the first formant in all-pole modeling (El-Jaroudi and Makhoul, 1991). This, in turn, results in severe distortion of the glottal flow estimates.

[4] It should be noted that while synthetic vowels produced by the physical modeling approach mimic real speech production by involving source-tract interaction, this effect is not taken into account in CP analysis, which simply assumes that the source and tract are linearly separable (Strube, 1974; Wong *et al.*, 1979). The proposed dc-constrained LP is a new mathematical method to compute the vocal tract model of CP analysis, but it does not in any way change the underlying assumption of the linear coupling between the source and the tract. Therefore, the use of physically-motivated synthetic speech was justified by a need to have more realistic artificial vowels as test material, not by a goal to analyze how source-tract interaction affects different versions of the CP technique, all of which are based on the linear source-filter theory and are therefore unable to take into account the coupling between the source and the tract.

Airas, M., and Alku, P. (**2006**). “Emotions in vowel segments of continuous speech: Analysis of the glottal flow using the normalized amplitude quotient,” Phonetica **63**, 26–46.

Akande, O., and Murphy, P. (**2005**). “Estimation of the vocal tract transfer function with application to glottal wave analysis,” Speech Commun. **46**, 15–36.

Alku, P. (**1992**). “Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering,” Speech Commun. **11**, 109–118.

Alku, P., and Vilkman, E. (**1996**). “A comparison of glottal voice source quantification parameters in breathy, normal, and pressed phonation of female and male speakers,” Folia Phoniatr Logop **48**, 240–254.

Arroabarren, I., and Carlosena, A. (**2004**). “Vibrato in singing voice: The link between source-filter and sinusoidal models,” EURASIP J. Appl. Signal Process. **7**, 1007–1020.

Bäckström, T., and Alku, P. (**2006**). “Harmonic all-pole modelling for glottal inverse filtering,” in CD Proceedings of the seventh Nordic Signal Processing Symposium, Reykjavik, Iceland.

Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (**1993**). *Nonlinear Programming: Theory and Algorithms* (Wiley, New York).

Bozkurt, B., Doval, B., D’Alessandro, C., and Dutoit, T. (**2005**). “Zeros of z-transform representation with application to source-filter separation of speech,” IEEE Signal Process. Lett. **12**, 344–347.

Campedel-Oudot, M., Cappe, O., and Moulines, E. (**2001**). “Estimation of the spectral envelope of voiced sounds using a penalized likelihood approach,” IEEE Trans. Speech Audio Process. **9**, 469–481.

Carlson, R., Granström, B., and Karlsson, I. (**1991**). “Experiments with voice modelling in speech synthesis,” Speech Commun. **10**, 481–489.

Childers, D., and Ahn, C. (**1995**). “Modeling the glottal volume-velocity waveform for three voice types,” J. Acoust. Soc. Am. **97**, 505–519.

Childers, D., and Hu, H. (**1994**). “Speech synthesis by glottal excited linear prediction,” J. Acoust. Soc. Am. **96**, 2026–2036.

Childers, D., and Lee, C. (**1991**). “Vocal quality factors: Analysis, synthesis, and perception,” J. Acoust. Soc. Am. **90**, 2394–2410.

Childers, D., and Wong, C.-F. (**1994**). “Measuring and modeling vocal source-tract interaction,” IEEE Trans. Biomed. Eng. **41**, 663–671.

Cummings, K. E., and Clements, M. A. (**1995**). “Analysis of the glottal excitation of emotionally styled and stressed speech,” J. Acoust. Soc. Am. **98**, 88–98.

El-Jaroudi, A., and Makhoul, J. (**1991**). “Discrete all-pole modeling,” IEEE Trans. Signal Process. **39**, 411–423.

Eysholdt, U., Tigges, M., Wittenberg, T., and Pröschel, U. (**1996**). “Direct evaluation of high-speed recordings of vocal fold vibrations,” Folia Phoniatr Logop **48**, 163–170.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fitch, T., and Giedd, J. (**1999**). “Morphology and development of the human vocal tract: A study using magnetic resonance imaging,” J. Acoust. Soc. Am. **106**, 1511–1522.

Flanagan, J. (**1972**). *Speech Analysis, Synthesis and Perception* (Springer, New York).

Fröhlich, M., Michaelis, D., and Strube, H. (**2001**). “SIM—Simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals,” J. Acoust. Soc. Am. **110**, 479–488.

Fu, Q., and Murphy, P. (**2006**). “Robust glottal source estimation based on joint source-filter model optimization,” IEEE Trans. Audio, Speech, Lang. Process. **14**, 492–501.

Gobl, C., and Ní Chasaide, A. (**2003**). “The role of voice quality in communicating emotion, mood and attitude,” Speech Commun. **40**, 189–212.

Hertegård, S., Gauffin, J., and Karlsson, I. (**1992**). “Physiological correlates of the inverse filtered flow waveform,” J. Voice **6**, 224–234.

Hirano, M. (**1974**). “Morphological structure of the vocal cord as a vibrator and its variations,” Folia Phoniatr Logop **26**, 89–94.

Hirano, M. (**1981**). *Clinical Examination of Voice* (Springer, New York).

Hollien, H., Dew, D., and Philips, P. (**1971**). “Phonational frequency ranges of adults,” J. Speech Hear. Res. **14**, 755–760.

Hollien, H., and Shipp, T. (**1972**). “Speaking fundamental frequency and chronologic age in males,” J. Speech Hear. Res. **15**, 155–159.

Kasuya, H., Maekawa, K., and Kiritani, S. (**1999**). “Joint estimation of voice source and vocal tract parameters as applied to the study of voice source dynamics,” in Proceedings of the International Congress on Phonetic Sciences, San Francisco, CA, pp. 2505–2512.

Klatt, D., and Klatt, L. (**1990**). “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” J. Acoust. Soc. Am. **87**, 820–857.

Krishnamurthy, A., and Childers, D. (**1986**). “Two-channel speech analysis,” IEEE Trans. Acoust., Speech, Signal Process. **34**, 730–743.

Larar, J., Alsaka, Y., and Childers, D. (**1985**). “Variability in closed phase analysis of speech,” in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Tampa, FL, pp. 1089–1092.

Lecluse, F., Brocaar, M., and Verschuure, J. (**1975**). “The electroglottography and its relation to glottal activity,” Folia Phoniatr. **17**, 215–224.

Lehto, L., Laaksonen, L., Vilkman, E., and Alku, P. (**2008**). “Changes in objective acoustic measurements and subjective voice complaints in call-center customer-service advisors during one working day,” J. Voice **22**, 164–177.

Liljencrants, J. (**1985**). “Speech synthesis with a reflection-type line analog,” DS dissertation, Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, Sweden.

Makhoul, J. (**1975**). “Linear prediction: A tutorial review,” Proc. IEEE **63**, 561–580.

Markel, J., and Gray, A., Jr. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).

Milenkovic, P. (**1986**). “Glottal inverse filtering by joint estimation of an AR system with a linear input model,” IEEE Trans. Acoust., Speech, Signal Process. **34**, 28–42.

Miller, R. (**1959**). “Nature of the vocal cord wave,” J. Acoust. Soc. Am. **31**, 667–677.

Naylor, P., Kounoudes, A., Gudnason, J., and Brookes, M. (**2007**). “Estimation of glottal closure instants in voiced speech using the DYPSA algorithm,” IEEE Trans. Audio, Speech, Lang. Process. **15**, 34–43.

Oppenheim, A., and Schafer, R. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Plumpe, M., Quatieri, T., and Reynolds, D. (**1999**). “Modeling of the glottal flow derivative waveform with application to speaker identification,” IEEE Trans. Speech Audio Process. **7**, 569–586.

Price, P. (**1989**). “Male and female voice source characteristics: Inverse filtering results,” Speech Commun. **8**, 261–277.

Rabiner, L., and Schafer, R. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).

Riegelsberger, E., and Krishnamurthy, A. (**1993**). “Glottal source estimation: Methods of applying the LF-model to inverse filtering,” in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Minneapolis, MN, Vol. **2**, pp. 542–545.

Rothenberg, M. (**1973**). “A new inverse-filtering technique for deriving the glottal air flow waveform during voicing,” J. Acoust. Soc. Am. **53**, 1632–1645.

Shiga, Y., and King, S. (**2004**). “Accurate spectral envelope estimation for articulation-to-speech synthesis,” in the CD Proceedings of the Fifth ISCA Speech Synthesis Workshop, Pittsburgh, PA.

Stoicheff, M. L. (**1981**). “Speaking fundamental frequency characteristics of nonsmoking female adults,” J. Speech Hear. Res. **24**, 437–441.

Story, B. (**1995**). “Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract,” Ph.D. dissertation, University of Iowa.

Story, B. (**2005**). “Synergistic modes of vocal tract articulation for American English vowels,” J. Acoust. Soc. Am. **118**, 3834–3859.

Story, B., and Titze, I. (**1995**). “Voice simulation with a body-cover model of the vocal folds,” J. Acoust. Soc. Am. **97**, 1249–1260.

Story, B., Titze, I., and Hoffman, E. (**1996**). “Vocal tract area functions from magnetic resonance imaging,” J. Acoust. Soc. Am. **100**, 537–554.

Strik, H., and Boves, L. (**1992**). “On the relation between voice source parameters and prosodic features in connected speech,” Speech Commun. **11**, 167–174.

Strube, H. (**1974**). “Determination of the instant of glottal closure from the speech wave,” J. Acoust. Soc. Am. **56**, 1625–1629.

Strube, H. (**1982**). “Time-varying wave digital filters for modeling analog systems,” IEEE Trans. Acoust., Speech, Signal Process. **30**, 864–868.

Sundberg, J., Fahlstedt, E., and Morell, A. (**2005**). “Effects on the glottal voice source of vocal loudness variation in untrained female and male voices,” J. Acoust. Soc. Am. **117**, 879–885.

Švec, J., and Schutte, H. (**1996**). “Videokymography: High-speed line scanning of vocal fold vibration,” J. Voice **10**, 201–205.

Titze, I. (**2002**). “Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model,” J. Acoust. Soc. Am. **111**, 367–376.

Titze, I., and Story, B. (**2002**). “Rules for controlling low-dimensional vocal fold models with muscle activities,” J. Acoust. Soc. Am. **112**, 1064–1076.

Titze, I., Story, B., Burnett, G., Holzrichter, J., Ng, L., and Lea, W. (**2000**). "Comparison between electroglottography and electromagnetic glottography," J. Acoust. Soc. Am. **107**, 581–588.

Titze, I., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.

Veeneman, D., and BeMent, S. (**1985**). "Automatic glottal inverse filtering from speech and electroglottographic signals," IEEE Trans. Acoust., Speech, Signal Process. **33**, 369–377.

Vilkman, E. (**2004**). "Occupational safety and health aspects of voice and speech professions," Folia Phoniatr Logop **56**, 220–253.

Wong, D., Markel, J., and Gray, A., Jr. (**1979**). "Least squares glottal inverse filtering from the acoustic speech waveform," IEEE Trans. Acoust., Speech, Signal Process. **27**, 350–355.

Yegnanarayana, B., and Veldhuis, N. (**1998**). "Extraction of vocal-tract system characteristics from speech signals," IEEE Trans. Speech Audio Process. **6**, 313–327.

Perceptual learning of systematic variation in Spanish-accented speech

Sabrina K. Sidaras, Jessica E. D. Alexander, and Lynne C. Nygaard
Department of Psychology, Emory University, 532 Kilgo Circle, Atlanta, Georgia 30322

(Received 20 October 2008; revised 29 January 2009; accepted 20 February 2009)

Spoken language is characterized by an enormous amount of variability in how linguistic segments are realized. In order to investigate how speech perceptual processes accommodate to multiple sources of variation, adult native speakers of American English were trained with English words or sentences produced by six Spanish-accented talkers. At test, listeners transcribed utterances produced by six familiar or unfamiliar Spanish-accented talkers. With only brief exposure, listeners perceptually adapted to accent-general regularities in spoken language, generalizing to novel accented words and sentences produced by unfamiliar accented speakers. Acoustic properties of vowel production and their relation to identification performance were assessed to determine if the English listeners were sensitive to systematic variation in the realization of accented vowels. Vowels that showed the most improvement after Spanish-accented training were distinct from nearby vowels in terms of their acoustic characteristics. These findings suggest that the speech perceptual system dynamically adjusts to the acoustic consequences of changes in talker's voice and accent.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3101452]

PACS number(s): 43.71.Bp, 43.71.Es, 43.71.Hw [MSS] Pages: 3306–3316

I. INTRODUCTION

A signature problem in the study of speech perception is how listeners maintain stable linguistic percepts despite the large amount of variability inherent in the acoustic speech signal. Each talker's utterances are uniquely shaped by a host of talker-specific characteristics such as individual identity, emotional state, and region of origin (Frick, 1985; Labov, 1972; Van Lancker *et al.*, 1985). Although these properties are highly informative, differences in the way each talker produces an utterance introduce considerable variability into the speech signal. Listeners must somehow cope with this variability to arrive at the constant linguistic percepts necessary for subsequent stages of linguistic analysis.

Prior research suggests that variability among different talkers may not necessarily be a perceptual problem for listeners but rather a source of lawful variation that is learned, retained, and used during spoken language processing. A number of studies have shown that listeners both attend to variation in talker's voice (Green *et al.*, 1991; Magnuson and Nusbaum, 2007; Mullennix *et al.*, 1989; Mullennix and Pisoni, 1990; Nusbaum and Magnuson, 1997) and retain talker-specific characteristics of speech in memory (Bradlow *et al.*, 1999; McLennan and Luce, 2005; Nygaard *et al.*, 2000; Palmeri *et al.*, 1993). Further, when given experience with particular speakers, listeners appear to engage in perceptual learning of surface characteristics of speech (Allen and Miller, 2004; Ladefoged and Broadbent, 1957; Nygaard and Pisoni, 1998; Nygaard *et al.*, 1994; Yonan and Sommers, 2000), and this learning facilitates the processing of linguistic structure.

Other research has investigated the degree to which listeners can adapt to systematic variation in synthesized, noise-vocoded, and time-compressed speech (Davis *et al.*, 2005; Dupoux and Green, 1997; Greenspan *et al.*, 1988; Schwab *et al.*, 1985). Greenspan *et al.* (1988) exposed listeners to synthetic speech, either word- or sentence-length utterances, over a training period of several days. Listeners who received training showed better transcription accuracy than those listeners who did not receive training. Additional research suggests that listeners can even perceptually accommodate to drastic alterations in the acoustic speech signal, such as time-compressed (Dupoux and Green, 1997) and noise-vocoded speech (Davis *et al.*, 2005).

Although these results demonstrate that listeners perceptually adapt to the unique characteristics of synthetic and altered speech, the variation in these types of signals is highly systematic, altering the speech signal in regularized ways depending on the particular synthesis or resynthesis technique. As a consequence, this type of input is arguably less variable across utterances than are the types of embedded sources of variation found in natural speech. One such source of natural variation that listeners routinely encounter in everyday communication is speech produced by non-native speakers of a particular language or foreign-accented speech. Because utterances produced by non-native speakers are filtered through the articulatory habits and phonological structure of their native language, accentedness systematically affects the linguistic realization of multiple aspects of spoken language (Flege *et al.*, 1997; Flege and Fletcher, 1992; Flege *et al.*, 1999). Systematic variation due to accentedness has been found to influence the intelligibility of non-native speech such that non-native talkers are less intelligible than native talkers, and listening to accented speech requires increased processing effort and time (Goggin *et al.*, 1991; Munro, 1998; Munro and Derwing, 1995; Schmid and Yeni-Komshian, 1999; van Wijngaarden *et al.*, 2002).

One challenge for the listener is that variation due to accent is produced in conjunction with variation due to individual talkers' voices. In order to understand accented speech, listeners must identify the independent contributions of talker-specific variation and the accent-general variation introduced by speakers' non-native articulatory habits and native phonological structure. Only a handful of studies have begun to examine adaptation to this type of variation (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001). In a recent study, Bradlow and Bent (2008) exposed native English listeners to Chinese-accented English and then tested transcription of English utterances produced by a single novel Chinese- or Slovakian-accented talker. Listeners who received training showed better sentence transcription performance for a novel Chinese- than a novel Slovakian-accented talker at test. Additionally, listeners exposed to multiple accented talkers during training performed better than those trained with a single accented talker. Although this study as well as others suggest that listeners may be sensitive to the lawful variation inherent in accented speech, less clear is the extent to which listeners are learning general systematic attributes of the accent or instead, properties specifically relevant to the particular talker used at test. Studies to date have focused on assessing generalization to just a single novel accented speaker and as such, the extent to which systematic variation is learned during these tasks remains an open question. The current investigation examined the issue of whether listeners learn accent-general or talker-specific properties of variation by determining the extent to which listeners generalize to *multiple* talkers and utterances.

Another question that remains to be addressed concerns *what* properties of foreign-accented speech listeners might be learning with exposure to non-native speech. Previous research has focused almost exclusively on perceptual adaptation to sentence-length utterances (e.g., Bradlow and Bent, 2008) and the extent to which higher-level lexical, semantic, and syntactic constraints might be instrumental in tuning perceptual mechanisms to particular properties of altered or accented speech (Davis *et al.*, 2005; Norris *et al.*, 2003). However, because sentences contain multiple sources of information including prosodic and segmental structure, at issue are what accent-specific properties listeners are learning. When judging degree of accentedness, listeners appear sensitive to both prosodic and segmental aspects of non-native speech (Boula de Mareüil and Vieru-Dimulescu, 2006) and with sentence-length utterances, listeners may be adapting either to global properties such as prosodic and intonational contours or to regularities in the acoustic-phonetic structure of accented speech.

Certainly, previous research suggests that listeners are sensitive to systematic variation due to accent and alter their processing of linguistic structure accordingly (Evans and Iverson, 2004). One example of this perceptual precision comes from several recent studies (Eisner and McQueen, 2005; Kraljic and Samuel, 2006, 2007; Ladefoged and Broadbent, 1957; Norris *et al.*, 2003) demonstrating that listeners are able to use lexical support to shift their phonetic category structure to include unusual pronunciations of particular contrasts. Norris *et al.* (2003) found that when listeners were given experience with ambiguous phonetic segments in lexically constraining contexts, their phonetic category boundaries shifted in keeping with the lexically driven learning. Although these studies suggest that listeners track systematicities in variation at the segmental level and alter their linguistic category structure when relevant to linguistic processing, it is unclear to what extent perceptual adjustments occur when listeners are confronted with multiple talkers and items in a high-variability learning and test paradigm.

For the current investigation, a high-variability training paradigm was created in which native English-speaking listeners were exposed to Spanish-accented speech produced by multiple (3 males and 3 females) non-native talkers. At test, listeners were presented either with the same set of six talkers heard during training or with a different set of six Spanish-accented speakers. Assessing generalization to multiple familiar and unfamiliar accented talkers provided a crucial test of the degree to which listeners engage in perceptual learning of the overarching lawful variation found in accented speech. It was predicted that if listeners are simply learning properties that are specific to individual accented talkers encountered during training, then improved transcription performance should be found only for accented talkers that are familiar at test. However, if listeners are perceptually adapting to general, systematic properties of accent, then listeners should generalize both to novel utterances and to multiple unfamiliar speakers.

In addition to assessing generalization of learning, perceptual learning of accented speech was examined using both sentence- and word-length utterances. If listeners are primarily learning the global properties associated with accent, such as prosodic and intonational contours, then accent learning should occur only with sentence-length utterances. However, if listeners are also sensitive to segmental properties of speech that vary with accent, then perceptual learning should be observed with word-length utterances as well.

In addition to general measures of perceptual tuning, we conducted further analyses of the particular types of acoustic-phonetic cues listeners may be using to perceptually adapt to accented speech. Production and identification of accented vowels served as a starting point for the investigation of the fine structure of perceptual learning. Analysis of listeners' identification of a subset of accented vowels that were more or less confusable was performed on the word transcription data from the test phase of the perceptual learning task. If listeners are learning systematic segmental information during training, their identification of certain vowels should be better at test than listeners who were not exposed to the accented speech.

Finally, acoustic analyses were performed to investigate how the systematic variation at the phonetic level may have influenced learning. Both temporal and spectral analyses of the accented Spanish vowels as well as the same vowels produced by native English speakers were compared to determine whether the native Spanish speakers produced systematic cues to particular segments and to what extent those cues were similar to or different from those produced by native English speakers. It was predicted that those vowels that were distinct with respect to temporal or spectral cues would be identified more accurately and learned more readily than those vowels that tended to overlap in acoustic-phonetic space.

II. EXPERIMENT 1

Experiment 1 examined perceptual learning of accented speech using sentence-length utterances. Accented speech differs systematically from native speech not only in segmental characteristics but also in prosodic structure. Experiment 1 examined the extent to which listeners would be able to exploit these multiple sources to perceptually adapt to regularities in accented speech and generalize that learning to both novel utterances and multiple novel talkers.

A. Method

1. Listeners

Listeners were 80 undergraduates who received partial credit in an introductory psychology course. The participants in this and the following experiments were native speakers of English with no reported history of speech or hearing disorders and were not fluent speakers of Spanish.

2. Materials

Twelve native Spanish speakers (6 males and 6 females) from Mexico City were recruited from the Atlanta area. Their mean age in years at the time of recording was 32.75 (range 26–39), on arrival to the U.S. was 26.42 (range 21–34), and when speakers began to learn English was 16.67 (range 2–28). Native English speakers (3 males and 3 females) provided control stimuli.

A set of 100 Harvard sentences (IEEE Subcommittee, 1969) and 144 monosyllabic words was recorded onto digital audiotape and re-digitized at a 22.050 kHz sampling rate, edited into separate files, and amplitude normalized.[1] All sentences were monoclausal and contained five key words (e.g., The *birch canoe slid* on the *smooth planks*). Sentences used at test were mixed with white noise at a +10 signal to noise ratio.

Separate groups of ten listeners transcribed all 100 sentences and 144 words for each of the 12 accented talkers to determine baseline intelligibility. An additional ten native English-speaking listeners rated the accentedness of ten sentence-length utterances from each of the 12 talkers. Listeners rated the accentedness of each utterance on a seven-point Likert-type scale, from 1 = "not accented" to 7 = "very accented". Table I lists mean accent ratings as well as baseline word and sentence intelligibility scores for each talker. Accentedness ratings and baseline intelligibility were correlated, $r = -0.88$, $p < 0.05$, indicating that more intelligible speakers were judged as less accented.

Talkers were divided into two groups for counterbalancing purposes based on mean accentedness (based on sentences) and single word intelligibility score. Each group was made up of three males and three females with approximately equivalent intelligibility and accentedness. Groups did not differ significantly on accentedness, $t(5) = 0.61$, $p = 0.57$ ($M_{\text{group }1} = 4.12$, $M_{\text{group }2} = 4.33$) or on intelligibility, $t(10) = 0.23$, $p = 0.75$ ($M_{\text{group }1} = 46.6\%$, $M_{\text{group }2} = 49.9\%$).

TABLE I. Accentedness and intelligibility for Spanish-accented speakers.

Speaker group	Gender	Mean accentedness ratings	Mean intelligibility (sentences) (%)	Mean intelligibility (words) (%)
	Female	5.59	75.6	32.93
Spanish group 1	Female	4.43	83.0	39.73
	Female	3.10	89.8	68.80
	Male	4.77	65.9	54.50
	Male	2.83	90.5	58.27
	Male	4.01	82.9	49.20
	Female	4.31	85.5	48.93
Spanish group 2	Female	6.17	74.6	35.20
	Female	4.54	82.6	53.50
	Male	3.55	81.8	42.93
	Male	2.68	90.7	60.27
	Male	4.75	89.0	52.80

3. Procedure

Training varied across conditions but materials and speakers at test remained the same. This design allowed for the comparison of listeners' performance with the exact same items (words and talkers) at test. During training, listeners were exposed to spoken items produced by one of two groups of six Spanish-accented speakers, six native English speakers, or received no training at all. The English training and the no training groups served as controls. Listeners trained with the Spanish-accented speakers heard either the same voices during training and at test or different voices during training and at test. Talker group was counterbalanced such that half the listeners in each condition heard group 1 at test and half heard group 2 at test.

Training phase. Training consisted of four comparison blocks and three variability blocks that were presented in alternation. In each of the comparison blocks, listeners heard each of the six Spanish-accented talkers or native English-speaking controls (3 males and 3 females) produce four different sentences and rated the accentedness of each sentence on a scale of 1–7. In the variability blocks, listeners heard two repetitions of three sentences per speaker presented in random order, with novel sentences in each block. Across repetitions within a block, talker/sentence pairings changed so that listeners never heard the same sentence produced by the same talker more than once. Listeners were asked to type the sentences they heard and were given as much time as needed to transcribe each sentence. After each response, the intended target sentence was presented both on the screen and repeated over the headphones. The training period lasted approximately 40 min. All training sentences were presented in the clear.

Generalization test. Listeners in all conditions heard the same group of six Spanish-accented talkers producing 30 novel sentences at test. Five sentences produced by each talker were presented in random order and listeners performed the transcription task with no feedback. All of the sentences in the test phase were mixed in noise. The listeners trained with Spanish-accented speech all heard a familiar accent at test. What varied was whether the talkers were familiar (same condition) or unfamiliar (different condition). For the control groups, both accent and talkers' voices were unfamiliar.

Listeners were tested individually in a quiet room. Stimulus presentation and data collection were controlled using PSYSCOPE (Cohen *et al.*, 1993) on a PowerMac G3 computer. The auditory stimuli were presented binaurally over Beyerdynamic DT100 headphones at approximately 75 dB sound pressure level (SPL).

B. Results and discussion

Sentence transcription performance was scored for proportion total words correct in the sentences as well as for proportion key words correct. Proportion total words correct are reported as there were no differences in the effects using total or key words correct.

Training phase. Because performance in the English training group was uniformly high ($M_1=98.9$; $M_2=99.6$; $M_3=99.5$), transcription performance during training was only analyzed for the two Spanish-accented training groups. Participant (F_1) and item (F_2) analyses of variance (ANOVA) were conducted with variability blocks across training (blocks 1–3) and training group (same vs different) as factors. A significant main effect of training block was found for participants, $F_1(2,80)=28.72$, $p<0.001$, *partial* $\eta^2=0.42$ and $F_2(1,52)=1.00$, $p=0.374$, partial $\eta^2=0.037$. The main effect of training group was not significant for participants but was for items, $F_1(1,40)=1.62$, $p=0.21$, partial $\eta^2=0.039$ and $F_2(1,52)=6.39$, $p<0.02$, partial $\eta^2=0.11$. In general, transcription performance improved across blocks for both Spanish-accented training groups: same ($M_1=91.1$, $M_2=93.4$, $M_3=94.4$) and different ($M_1=92.2$, $M_2=94.6$, $M_3=95.3$), with better performance for the different than same group for items. No significant interaction between training group and training blocks was found either for participants or items. Planned comparisons (for participants) showed significant improvement in transcription performance between blocks 1 and 2, $F(1,40)=22.64$, $p<0.001$, partial $\eta^2=0.36$, and between blocks 2 and 3, $F(1,40)=6.00$, $p<0.02$, partial $\eta^2=0.13$.

Generalization test. Figure 1 shows percent correct transcription performance at test for each training group. One-way participant (F_1) and item (F_2) ANOVAs assessing listeners' performance at test revealed a significant main effect of training group, $F_1(3,76)=4.97$, $p<0.004$, partial $\eta^2=0.16$ and $F_2(3,90)=20.13$, $p<0.001$, partial $\eta^2=0.40$. Planned comparisons revealed no significant differences between the Spanish-accented training groups, same ($M=62.0$, SD=7.7) vs different ($M=61.1$, SD=5.5), $p_1=0.75$, $p_2=0.71$, or between the two control groups, English ($M=55.7$, SD=6.3) vs no training ($M=56.7$, SD=6.4), $p_1=0.66$, $p_2=0.34$, at test. However, listeners who received training with Spanish-accented speech ($M=60.1$, SD=6.4) performed better at test than listeners who received English or no training ($M=55.2$, SD=5.9), $F_1(1,76)=13.97$, $p<0.001$, partial $\eta^2=0.16$ and $F_2(1,90)=35.14$, $p<0.001$, partial $\eta^2=0.54$.

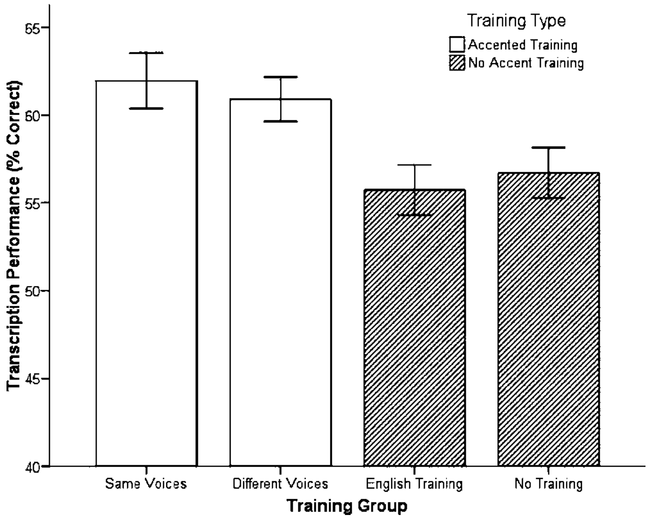


FIG. 1. Mean transcription performance at test for sentences as a function of training group.

These findings indicate that training with Spanish-accented speech resulted in perceptual adaptation to accent-general characteristics of non-native speech. Listeners generalized both to novel utterances and to novel voices within the same accent group suggesting that learning was not tied to particular tokens or talkers. In addition, improvement was observed after relatively brief exposure to accented speech suggesting that listeners adapted quickly to the lawful variation in accented speech.

Little evidence was found for talker-specific learning in addition to accent-general learning in this task. Perhaps because listeners received relatively more experience with the Spanish accent and relatively less experience with any particular talker, this type of training may have encouraged listeners to track commonalities *across* speakers rather than focus on the idiosyncrasies of any particular talker.

III. EXPERIMENT 2

Experiment 2 examined listeners' ability to perceptually adapt to properties of accented speech in single words. The results of experiment 1, along with previous research (e.g., Bradlow and Bent, 2008), suggest that listeners are sensitive to the regularities found in foreign-accented sentences. Using single words at training and test reduced the availability of global properties and allowed us to examine the extent to which listeners can learn systematic variation specific to the acoustic-phonetic structure of accented speech.

A. Method

1. Listeners

Listeners were 98 undergraduate students who received partial course credit in an introductory psychology course for their participation.

2. Materials

The same non-native Spanish and native English speakers that were used in the previous experiment also recorded a list of 144 monosyllabic English words (72 *easy* and 72 *hard*).[2] Easy words were high frequency words ($M=309.69$; Kučera and Francis, 1967) with few ($M=38.32$) low frequency neighbors (e.g., size, piece; Luce and Pisoni, 1998). Hard words were low frequency words ($M=12.21$) with many ($M=282.22$) high frequency neighbors (e.g., sane, lace). Both easy and hard words were rated as being highly familiar ($M=6.97$; on a scale of 1–7 with 1 being not familiar at all and 7 being highly familiar (Nusbaum *et al.*, 1984).

3. Procedure

Training phase. The same design was used as in experiment 1. Because more words than sentences were available, in each variability block, listeners heard two repetitions of each talker producing four different English words, with novel words in each block. All training words were presented in the clear, and listeners received feedback as in experiment 1 on their transcriptions.

Generalization test. At test, listeners transcribed a total of 48 novel accented words, eight words from each talker. All of the words in the test phase were presented in the clear, and no feedback was given. All other aspects of the procedure were the same as in experiment 1.

B. Results and discussion

Transcription accuracy was averaged across words for each participant. Words were scored as correct if listeners provided either the correct spelling or a homophone equivalent.

Training phase. As in experiment 1, since transcription performance was uniformly high in the English control condition ($M_1=93.6$, $M_2=92.6$, $M_3=93.5$), training performance was only evaluated for the two Spanish-accented conditions. Participant (F_1) and item (F_2) ANOVAs were conducted with training block (blocks 1–3) and training group (same vs different) as factors. A significant main effect of training block was found for participants, $F_1(2,96)=29.00$, $p<0.001$, partial $\eta^2=0.38$ and $F_2(2,70)=0.99$, $p=0.37$, partial $\eta^2=0.01$. The main effect of training group was not significant for participants, but was for items, $F_1(2,96)=1.10$, $p=0.30$, partial $\eta^2=0.02$, and $F_2(2,70)=4.5$, $p<0.05$, partial $\eta^2=0.06$. Transcription performance changed as a function of block for both Spanish-accented training groups: same ($M_1=59.3$, $M_2=58.7$, $M_3=66.3$) and different ($M_1=58.5$, $M_2=56.1$, $M_3=64.6$) with indication of better performance for the same than different group. No significant interaction between training group and training block was found for participants or items. Planned contrasts (for participants) showed significant improvement in transcription performance between blocks 2 and 3, $F(1,48)=50.65$, $p<0.001$, partial $\eta^2=0.51$, but not between blocks 1 and 2, $p=0.16$.

Generalization phase. Figure 2 shows transcription performance during the generalization test for each training group condition. One-way participant (F_1) and item (F_2) ANOVAs revealed a significant main effect of training group, $F_1(3,94)=4.08$, $p<0.01$, partial $\eta^2=0.11$ and $F_2(3,141)=3.06$, $p<0.05$, partial $\eta^2=06$. Planned comparisons showed no significant differences between the Spanish-accented training groups, same ($M=48.1$, SD=6.2) vs different ($M=46.9$, SD=4.6), $p_1=0.399$, $p_2=0.40$, or between the two control groups, English ($M=43.9$, SD=5.8) vs no training ($M=43.8$, SD=4.4), $p_1=0.96$, $p_2=0.96$. However, a significant difference was found between listeners that received Spanish-accented training ($M=47.5$, SD=5.4) and those that did not ($M=43.9$, SD=5.1), $F_1(1,96)=11.7$, $p<0.05$, partial $\eta^2=0.06$ and $F_2(1,47)=6.36$, $p<0.05$, partial $\eta^2=0.12$.

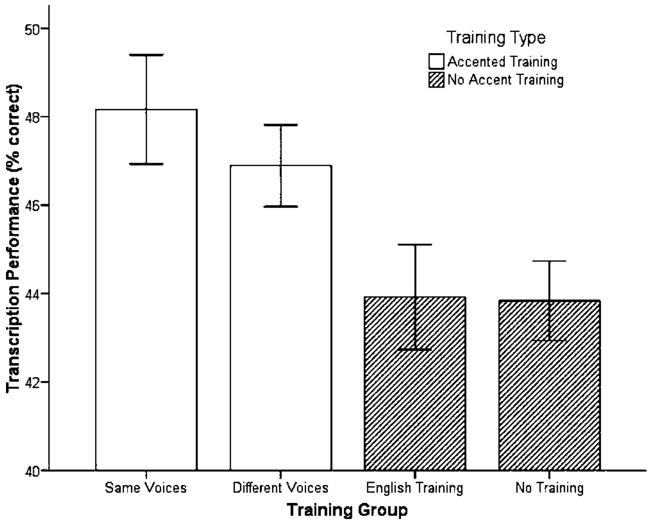


FIG. 2. Mean transcription performance at test for words as a function of training group.

These results indicate that a brief training session with isolated accented words produced perceptual adaptation. As in experiment 1, the intelligibility benefits of the training session generalized both to novel utterances and to novel talkers. The finding that perceptual learning occurred with single words suggests that listeners can attend to and learn not only the unique prosodic structure of accented speech but also the fine-grained details of the acoustic-phonetic structure of accented speech.

IV. PERCEPTION AND PRODUCTION OF SPANISH-ACCENTED VOWELS

In order to examine precisely what properties of Spanish-accented speech listeners were learning, we examined the perception of individual accented vowels from experiment 2 to determine which ones showed improvement as a function of training. In addition, we conducted acoustic analyses of the Spanish-accented vowels and the same vowels produced by native English speakers to determine if the native Spanish speakers produced reliable cues to particular segments.

Accented vowel production and perception was deemed a good starting place because the accuracy of both production and perception of vowels in a non-native language varies as a function of native language background (Bohn and Flege, 1992; Flege *et al.*, 1997; Flege *et al.*, 1999; Flege *et al.*, 2003; Munro, 1993). With respect to the present study, the Spanish vowel inventory /i, e, a, o, u/ differs from English both in number (Spanish has 5 vowels and English has approximately 11) and in their realization in spectral and temporal space (Bradlow, 1995). Based on this previous research, the native Spanish speakers in the present study

TABLE II. Confusion matrix for error analyses of word transcriptions.

Intended targets	Listeners' responses						
	/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other
			No accented training				
/i/	**45**	46	6				3
/ɪ/	31	**48**	10				11
/e/	1	2	**95**				2
/æ/			1	**71**	15	1	12
/ʌ/		1		10	**37**	25	27
/a/	1				15	**76**	8
			Spanish-accented training				
/i/	**51**	41	4	1			3
/ɪ/	28	**48**	13	1			10
/e/	1		**98**	1			
/æ/				**82**	13	1	4
/ʌ/		1		11	**36**	24	28
/a/		1			15	**80**	4

Values represent percent responses to target.

should have difficulty producing vowels that have no counterpart in their native language vowel inventory. In turn, the native English-speaking listeners would be expected to have difficulty identifying those same vowels. To that end, patterns of errors or confusions for vowel identification for trained and untrained listeners were calculated. The error analyses then served as a guide for the acoustic analyses to determine how the native Spanish speakers were producing the English vowels and which cues the English listeners were using to perceptually learn the systematic variation in the accented speech.

A. Error analyses

Analyses of vowel identification and confusions were calculated using the word transcription responses of listeners who participated in experiment 2. Evaluations of listeners' responses at test were thus necessarily limited by the orthographic constraints of written English. It should be noted, however, that these constraints were the same for both the trained and untrained groups. Listeners trained on accented voices, whether same or different, were grouped together ($n=49$) and listeners not trained with Spanish-accented speech were grouped together ($n=49$). The vowel identification analyses were carried out for target words with the vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/. Other vowels were excluded either because they were less frequent in our set or because the initial or final consonant heavily influenced the vowel (e.g., words with /r/ immediately following the vowel and words that began with /r/ or /w/). The data used for the analyses included listener responses to multiple words in each vowel category; 490 responses to /i/, 588 responses to /ɪ/, 588 responses to /e/, 294 responses to /æ/, 686 responses to /ʌ/, and 294 responses to /a/.

Table II shows confusion matrices of target vowels for trained and untrained listeners. Cells reflect percent identifications, which take into account the number of possible tokens. Regardless of training, the high front vowels /i/ and /ɪ/ were frequently confused with one another, while the /e/ vowel was relatively well identified. These confusions follow from a mapping between Spanish and English vowels with /ɪ/, a vowel not found in Spanish, being confused with other high front vowels /i/ and /e/. Likewise, for both trained and untrained listeners, the low vowels /æ/, /ʌ/, and /a/ were highly confusable. The accented /ʌ/, a vowel not in the Spanish inventory, was particularly difficult for the native English-speaking listeners. The pattern of confusions suggests that the native Spanish speakers had difficulty producing vowels that fell outside their vowel inventory (/ɪ/, /æ/, and /ʌ/) and that speakers were referencing their own vowel categories in order to approximate the non-native vowels (/i/ and /a/).

In addition to the overall pattern of confusions, the results show that listeners transcribed at least a subset of accented vowels more accurately after training with accented speech. Targeted comparisons of identification performance for listeners who did and did not receive accented training were completed for each of the accented vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/ that were analyzed. The vowels /i/, /æ/, and /a/ showed significantly higher accuracy for trained than untrained listeners, p's <0.05. The vowels /ɪ/, /e/, and /ʌ/ did not show a significant difference between trained and untrained listeners, all p's >0.05. The improvement in identification for particular accented vowels indicates that listeners might have been learning specific information during training that allowed them to better discriminate and identify particular vowels.

The vowel-specific nature of the learning guided the analysis of the acoustic-phonetic correlates to identification performance. If training with Spanish-accented speech reduced the confusability of vowels such as /i/, /æ/, and /a/, then acoustic-phonetic characteristics of these vowels should distinguish them from other vowels in the listeners' repertoire. In particular, temporal and spectral characteristics of the Spanish-accented vowels were examined to determine

TABLE III. Mean values and standard deviations for duration, F1, and F2 for native English and Spanish speaker groups.

		Duration		F1		F2	
Speaker group	Vowel	*M*	SD	*M*	SD	*M*	SD
English	/i/	194.16	32.31	347.79	48.36	2580.52	237.60
	/ɪ/	161.24	24.69	526.87	126.61	2079.26	162.67
	/e/	227.50	45.09	444.87	97.62	2462.76	164.09
	/æ/	228.52	30.92	837.04	208.61	1852.45	109.34
	/ʌ/	188.21	28.33	686.44	190.17	1413.56	132.22
	/a/	235.61	44.23	768.69	198.51	1221.27	72.73
Spanish	/i/	178.19	36.63	356.76	53.84	2433.36	229.45
	/ɪ/	169.59	40.71	392.07	48.71	2395.58	282.15
	/e/	235.87	40.51	439.96	52.00	2367.90	263.70
	/æ/	217.56	35.55	777.84	143.82	1637.71	144.31
	/ʌ/	189.02	41.32	634.20	85.69	1352.06	208.50
	/a/	195.63	34.55	664.78	89.72	1310.77	157.48

Mean values represent averages of tokens of each vowel for each speaker and standard deviations represent the variance of the means of these tokens.

which properties might be contributing both to the overall identification of these vowels and to the improvement that listeners achieve with training.

B. Acoustic analyses

Acoustic analyses of vowel duration and first (F1) and second (F2) formant center frequencies were carried out using PRAAT sound analysis software (Boersma and Weenink, 2006) for the English vowels embedded in words produced by the 12 native Spanish and 6 native English speakers from experiment 2. Only words with the target vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/ were analyzed. For each of these vowels, between 12 and 16 tokens were analyzed for each speaker (for a total of 144–192 tokens per vowel). Three trained coders completed all acoustic analyses, with a single coder completing all analyses for vowels produced by a single talker. Recall that the vowels were embedded in words that contained a variety of consonant contexts. Although the context varied, it was consistent across both the Spanish-accented and native English speakers. Criteria for determining vowel onset and offset were taken from Munson and Solomon (2004). Vowel duration was determined from onset and offset times. Measurements of the first and second formant frequencies were taken at the midpoint of the vowel. Inter-rater reliability for vowel onset and offset measures was assessed using a subset of six vowels for all talkers (12 Spanish-accented talkers, 6 native English talkers; 108 tokens). Reliability was good with 86% agreement among all three coders. Table III reports mean values and standard deviations of duration, F1, and F2 for each vowel.

Based on the differences in the identification scores and patterns of confusion for each vowel, we expected temporal or spectral overlap for vowels that were confusable (e.g., /i/ and /ɪ/) and less overlap for those that were not confusable (e.g., /i/ and /e/). Further, we expected that the specific vowels that were better identified after learning (/i/, /æ/, and /a/) would have temporal or spectral properties that distinguished them from other intended vowels. Separate focused analyses were conducted on the three high front vowels /i/, /ɪ/, and /e/ and on the three low vowels /æ/, /ʌ/, and /a/. All follow-up comparisons used a Bonferroni corrected alpha of 0.0125.

Temporal characteristics. Figures 3(a) and 3(b) show mean duration measures for each English vowel produced by the native English and Spanish speaker groups. Separate ANOVAs were performed on duration with speaker group (English or Spanish) as a between group factor and either vowel group /i/, /ɪ/, and /e/ or vowel group /a/, /æ/, and /ʌ/ as the within group factor. For the /i/, /ɪ/, and /e/ vowel grouping, a main effect of vowel, $F(2,32)=63.90$, $p<0.001$, partial $\eta^2=0.80$, but no main effect of speaker group or interaction was found. The pattern of duration differences across the three vowels, /i/, /ɪ/, and /e/, for native Spanish speakers was similar to those of native English speakers. Native Spanish speakers produced the English vowel /e/ with longer durations than either /i/ or /ɪ/, both p's<0.001. In addition, the vowel /i/ had longer durations than /ɪ/, $p<0.001$. The relative differences in duration among these vowels are consistent with previous findings (e.g., Flege *et al.*, 1997) and suggest that duration is a reliable cue that listeners may use to distinguish among Spanish-accented productions of these vowels.

For the vowels /a/, /æ/, and /ʌ/, the ANOVA revealed a significant interaction between speaker group and vowel, $F(2,32)=10.80$, $p<0.001$, partial $\eta^2=0.40$, indicating that the pattern of durations across vowels differed as a function of speaker group. Comparisons across vowels for the native Spanish speakers revealed significant differences in duration between productions of /æ/ and /ʌ/, $p<0.001$, and between /æ/ and /a/, $p<0.001$, but not between /ʌ/ and /a/, $p=0.113$. Spanish speakers did not distinguish the vowels /ʌ/ and /a/ in terms of duration and exhibited a pattern of durations across vowels that differed from the native English speakers. Recall that listeners who were trained with Spanish-accented speech showed better identification of the /æ/ vowel than those that were not trained. The pattern of differences in duration sug-

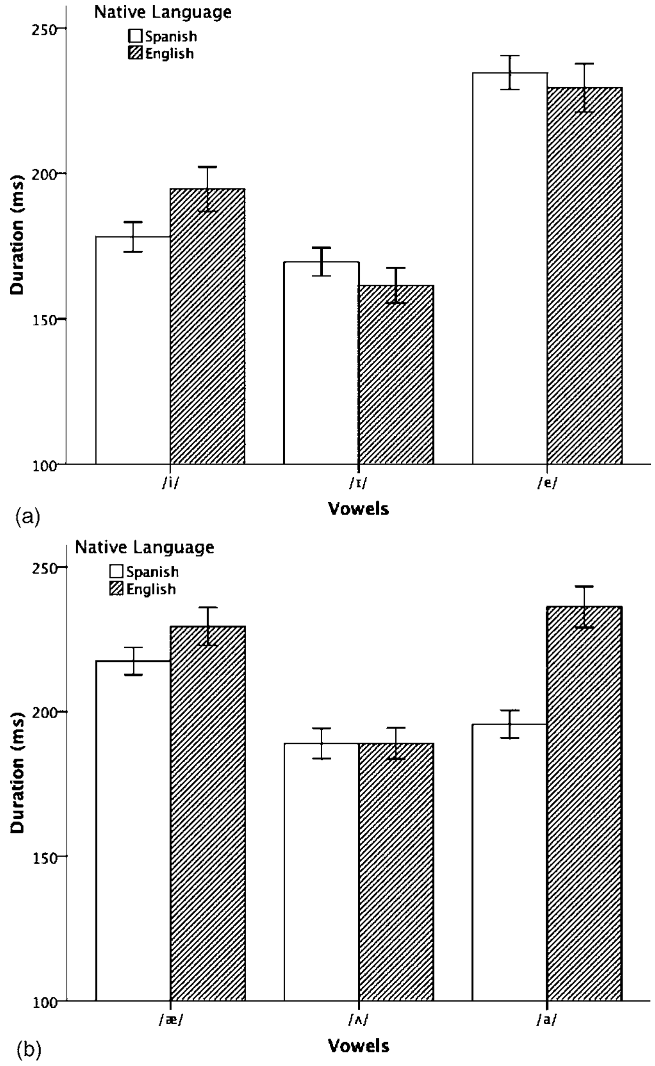


FIG. 3. Mean durations for native English and Spanish speaker groups for (a) vowels /i/, /ɪ/, and /e/ and (b) /æ/, /ʌ/, and /a/.

gests that this property could serve as one cue for the English listeners that distinguishes /æ/ from similar vowels for the native Spanish speakers.

Spectral characteristics. Figure 4 shows mean F1 and F2 values for the vowels /i/, /ɪ/, and /e/ for native English and Spanish speakers. Separate ANOVAs were performed on F1 and F2 with speaker group (English or Spanish) and vowel group (/i/, /ɪ/, and /e/) as factors.

For measures of F1, a significant interaction between speaker group and vowel was found, $F(2,32)=20.41$, $p<0.001$, partial $\eta^2=0.56$. Comparisons among vowels for the native Spanish speakers revealed that all pairwise comparisons were significant; /i/ and /e/, $p<0.001$, /ɪ/ and /e/, $p<0.001$, and /ɪ/ and /i/, $p<0.01$. Although the Spanish speakers were distinguishing among the three vowels, the pattern was very different from that produced by the native English speakers. For the Spanish speakers, F1 values for /ɪ/ fell between values for /i/ and /e/. For English speakers, F1 values for /e/ fell between /i/ and /ɪ/. The overlap in F1 frequencies among the three vowels coupled with the lower F1 frequency for /ɪ/ may have contributed to the confusability of /i/ and /ɪ/.

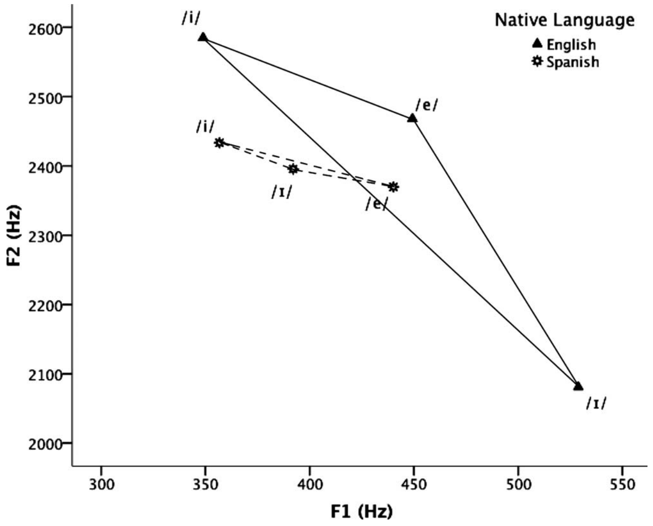


FIG. 4. Mean F1 and F2 values in hertz for /i/, /ɪ/, and /e/ for native English and Spanish speakers.

Turning to F2, a significant interaction between speaker group and vowel was found, $F(2,32)=38.57$, $p<0.001$, partial $\eta^2=0.71$. Comparisons across vowels for Spanish speakers revealed significant differences only between /e/ and /i/, $p=0.011$. It appears that accented speakers had difficulty producing /ɪ/, realizing the vowel with lower F1 and higher F2 values than native English speakers. These modified spectral characteristics overlapped with adjacent vowel categories and may have contributed to the confusability of /i/ and /ɪ/.

Figure 5 shows mean F1 and F2 values for the vowels /æ/, /ʌ/, and /a/ for native English and Spanish speakers. Again, separate ANOVAs were performed on F1 and F2 with speaker group (English or Spanish) and vowel (/æ/, /ʌ/, and /a/) as factors. For F1, there was a significant main effect of vowel $F(2,32)=23.28$, $p<0.001$, partial $\eta^2=0.59$, but no effect of speaker group and no interaction. For both groups of speakers, F1 values for /ʌ/ were significantly lower than for /a/, $p<0.003$, which in turn had lower F1 values than for the vowel /æ/, $p<0.003$. These results show that native

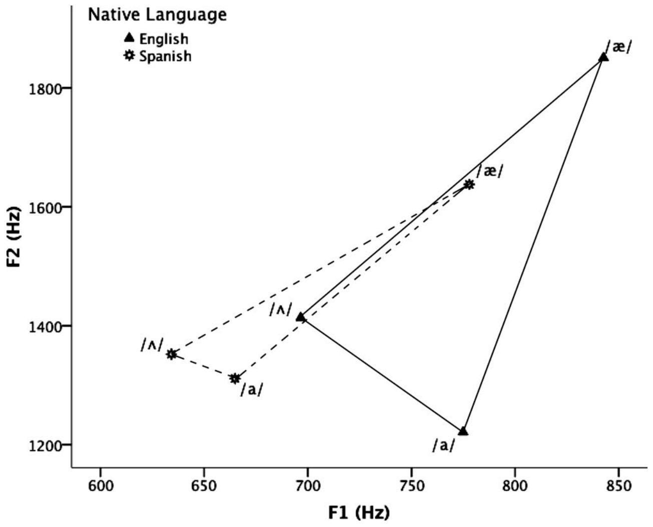


FIG. 5. Mean F1 and F2 values in hertz for /æ/, /ʌ/, and /a/ for native English and Spanish speakers.

Spanish speakers distinguished among these low vowels with respect to F1, approximating the pattern produced by native English speakers. In turn, these vowels were relatively less confusable than the high front vowels.

For measures of F2, there was a significant interaction between speaker group and vowel, $F(2,32)=8.36$, $p < 0.001$, partial $\eta^2=0.34$. Comparisons of relative vowel differences for the native Spanish speakers revealed differences between /æ/ and /ʌ/, $p<0.001$, and between /æ/ and /a/, $p < 0.001$. No significant difference was found between /a/ and /ʌ/, $p=0.221$.

Summary. These findings suggest that acoustic characteristics of the Spanish-accented English vowels may be related to the perceptual confusions observed for the native English listeners. Particular vowels or particular sets of vowels that were confusable to the English listeners also had temporal and/or spectral characteristics that overlapped in acoustic-phonetic space.

In particular, the vowels /i/ and /ɪ/ were found to be highly confusable in the error analyses (see Table II), which seems to correspond to the observed overlapping spectral characteristics of these vowels. Recall, however, that identification of the highly confusable /i/ was better for those listeners who received accentedness training than for those that did not. To speculate, listeners who received training with accented speech may have begun to distinguish among high front vowels within the Spanish speakers' relatively crowded vowel space by attending to the properties of accented productions of both /i/ and /e/ that proved to be similar to the native English productions, resulting in /i/ being susceptible to perceptual learning and /e/ being *a priori* less confusable.

For the trio of vowels /æ/, /ʌ/, and /a/, the vowel /ʌ/ was found to be highly confusable in the error analyses and the vowel /æ/ showed significant improvement as a function of training with accented speech. Likewise, the native Spanish speakers did not distinguish /ʌ/ from /a/ either with respect to duration or with respect to F2, perhaps making /ʌ/ less distinct perceptually. In contrast, the native Spanish speakers produced the vowel /æ/ with temporal and spectral properties that were significantly different from either /ʌ/ or /a/. Although the accented /æ/ was produced with a significantly lower F2, listeners with accentedness training may have been sensitive to the distinctive constellation of cues that set /æ/ apart, at least in this limited set of productions and analyses.

V. GENERAL DISCUSSION

The objective of this study was to investigate the nature and extent of perceptual learning of foreign accented speech. Perceptual learning of Spanish-accented sentence- and word-length utterances was examined in a high-variability training and test paradigm. We sought to determine whether listeners learn the systematic variation specific to accent by examining generalization of learning to multiple familiar and unfamiliar accented talkers. The results showed that after only a brief training period with sentences or words, listeners showed an increased ability to transcribe novel accented words and sentences produced by familiar talkers. Most remarkably, listeners generalized, showing increased transcription performance for a group of six unfamiliar talkers from the same accent group.

Previous research has demonstrated perceptual learning in speech and language processing in general (Davis *et al.*, 2005; Dupoux and Green, 1997; Greenspan *et al.*, 1988; Nygaard and Pisoni, 1998) and for accented speech in particular (e.g., Bradlow and Bent, 2008). However, studies of accommodation to accented speech have focused on generalization to a single novel accented talker (Bradlow and Bent, 2008; Weil, 2001; Clarke and Garrett, 2004). The current findings demonstrate generalization to multiple talkers, suggesting that perceptual learning occurs for accent-general properties of speech and is not tied to particular talker- or item-specific characteristics.

Our findings also begin to pinpoint the nature of the perceptual learning process. Previous studies have almost exclusively used sentence-length utterances to evaluate perceptual adaptation to accented speech (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001) leaving open the question of whether listeners are learning global prosodic features or regularities in phonological form. In the current investigation, listeners used information present in both sentence- and word-length utterances, suggesting a sensitivity to regularities in the acoustic-phonetic structure of accented speech.

In order to confirm that learning was taking place at a segmental level, perceptual confusions for a subset of vowels were examined for listeners who did and did not receive training. The results showed that those listeners who received training with accented speech showed better identification of certain accented vowels (/i/, /æ/, and /a/) than untrained listeners. It appears that listeners learned specific segmental information during training that allowed them to better discriminate and identify particular vowels.

In addition to perceptual confusions, acoustic analyses were performed to investigate which acoustic-phonetic cues of the accented vowels listeners may have learned. The pattern of listeners' vowel confusions suggests that, not surprisingly, the native Spanish speakers had difficulty producing vowels that fall outside their native vowel inventory (/ɪ/, /æ/, and /ʌ/). However, although vowels that were highly confusable to listeners had overlapping temporal and/or spectral characteristics, the native Spanish speakers did appear to produce systematic segmental acoustic-phonetic variation that may have contributed, at least in part, to the perceptual learning of the Spanish-accented speech. For instance, the low vowels /ʌ/ and /a/ were not distinct with respect to duration, but were distinguished by the native Spanish speakers with spectral properties. Thus, with training English listeners may have learned to rely to a greater extent on particular spectral cues for these vowels.

These findings are generally consistent with previous experiments that have shown perceptual adjustments of phoneme categories as a result of experience with unusual pronunciations (Eisner and McQueen, 2005; Kraljic and Samuel, 2006; Norris *et al.*, 2003). The present findings confirm that when listeners are exposed to variation in accented speech, they are able to extract specific systematic information on a segmental level that generalizes to novel talkers' voices. Although in some studies (Kraljic and Samuel, 2006, 2007; Norris *et al.*, 2003) perceptual learning of alternate pronunciations generalizes to different talkers' voices, other research with different contrasts has found that learning seems to be talker-specific (Eisner and McQueen, 2005). In the present experiment, listeners did generalize, indicating that listeners were able to learn which characteristics of the accented speech should be attributed to consistent properties of talker's voice and which characteristics are due to cross-speaker regularities in accent.

Exposure to extensive variability during training may be necessary for listeners to extract the systematicities present in accented speech. Previous studies have shown that high stimulus variability during training facilitates second language vocabulary learning (Barcroft and Sommers, 2005; Sommers and Barcroft, 2007) as well as the learning of non-native phonetic categories (Logan *et al.*, 1991; Lively *et al.*, 1993, 1994). In the present experiments, although training with accented speech was extremely brief, listeners were exposed to many novel voice-word pairings during both training and at test. The opportunity both to compare tokens across the training blocks as well as from multiple talkers may have allowed listeners to generalize learning to novel accented utterances and speakers.

It should be noted that all this variability, while potentially necessary for robust learning, made the listeners' task extremely difficult both during training and at test. Recall that listeners encountered spoken utterances produced by *multiple* familiar or unfamiliar accented talkers, and consequently, were forced to readjust to a new talker's voice on a trial-by-trial basis. Previous research has established that changes from trial to trial in characteristics of spoken language such as talker's voice incur a processing cost (Mullennix *et al.*, 1989). Nevertheless, listeners learned to parse multiple sources of variability, dynamically attributing variance in the speech signal to changes in the linguistic, talker-specific, and accent-general structure of speech.

These findings are consistent, in a broad sense, with accounts that assume that representation of spoken language includes both perceptual and linguistic properties of speech (Goldinger, 1998; Johnson, 1997; Jusczyk, 1997; Nygaard *et al.*, 1994; Pisoni, 1997). In this sense, perceptual learning of accented speech may be a form of perceptual expertise or automaticity that relies on the accumulation of representations which include the lawful variation in Spanish-accented speech (see Logan, 1988; Ettlinger, 2007). Alternatively, listeners may be tuning their procedural memory or normalization routines in an accent-general fashion (Kolers and Roediger, 1984; Nusbaum and Morin, 1992). Rather than explicitly representing perceptual details of spoken language, listeners may engage in a normalization procedure that becomes tuned to unravel the combined contributions of a particular accent, talker's voice, and other sources of variation.

Taken with previous findings, our data suggest that listeners appear to be exquisitely sensitive to systematic variation in speech and alter their processing or representation of linguistic structure accordingly (e.g., Eisner and McQueen, 2005; Norris *et al.*, 2003; Kraljic and Samuel, 2006, 2007). Perceptual processing and representation of spoken language appear to include and utilize surface characteristics of speech in linguistic processing. Listeners perceptually adapt as they build up a repertoire of experiences with accented speech that in turn facilitates later processing of the linguistic structure of speech. By engaging in perceptual learning of the lawful variation inherent in accented speech, listeners appear to be sensitive to the details of segmental variability resulting from the complex relationship between linguistic environment, idiosyncratic talker-specific variability, and variation due to properties of the accent itself.

ACKNOWLEDGMENTS

We would like to thank Jennifer Queen, Sue Ann Patton, Matt Ross, Kathy Jernigan, and Lisa Allen for their help with stimulus development, acoustic analyses, and data collection. Portions of these data were presented at the ISCA Workshop on Plasticity in Speech Perception in London, UK (June, 2005); the 46th Annual Meeting of the Psychonomic Society in Toronto, Canada (November, 2005); and the Fourth Joint Meeting of the Acoustical Society of America and The Acoustical Society of Japan in Honolulu, HI (November, 2006). This research was supported in part by an Emory University Research Committee grant and Research Grant R01 DC 008108 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

[1] Spanish speakers were given all stimulus materials before the date of recording to familiarize themselves with the materials in order to decrease the chance of making production errors during recording.

[2] Easy and hard words were used in order to evaluate the effects of lexical properties on perceptual learning. In this experiment, lexical properties influenced overall performance level but did not interact with any other variables in experiment 2. Participant (F_1) and item (F_2) ANOVAs with training groups (same, different, English, and no training) and word types (easy vs hard) factors revealed no interaction between training groups and easy/hard word performance, $F_1(3,94)=0.275$, $p=0.844$, partial $\eta^2=0.009$ and $F_2(1,70)=0.413$, $p=0.523$, partial $\eta^2=0.006$, but there were main effects of both word type and training groups. Transcription performance for hard words ($M=30.5$, $\mathrm{SD}=8.2$) was significantly worse than for easy words ($M=60.9$, $\mathrm{SD}=7.8$), $F_1(1,94)=660.9$, $p<0.001$, partial $\eta^2=0.875$ and $F_2(1,70)=19.98$, $p<0.001$, partial $\eta^2=0.12$. The main effect of word type indicates that neighborhood density and word frequency contributed overall to transcription performance, but did not seem to affect or be affected by the learning process in this task. As such, these properties are not discussed further in the current investigation.

Allen, J. S., and Miller, J. L. (**2004**). "Listener sensitivity to individual talker differences in voice-onset-time," J. Acoust. Soc. Am. **115**, 3171–3183.

Barcroft, J., and Sommers, M. S. (**2005**). "Effects of acoustic variability on second language vocabulary learning," Stud. Second Lang. Acquis. **27**, 387–414.

Boersma, P., and Weenink, D. (**2006**). "Praat: Doing phonetics by computer," from http://www.praat.org (Last viewed January, 2006), Version 5.0.23, computer program.

Bohn, O.-S., and Flege, J. (**1992**). "The production of new and similar vowels by adult German learners of English," Stud. Second Lang. Acquis. **14**, 131–158.

Boula de Mareüil, P., and Vieru-Dimulescu, B. (**2006**). "The contribution of prosody to the perception of foreign accent," Phonetica **63**, 247–267.

Bradlow, A. R. (**1995**). "A comparative acoustic study of English and Spanish vowels," J. Acoust. Soc. Am. **97**, 1916–1924.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (**1999**). "Effects of talker, rate, and amplitude variation on recognition memory," Percept. Psychophys. **61**, 206–219.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). "PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers," Behav. Res. Methods Instrum. Comput. **25**, 257–271.

Davis, M. H., Johnsrude, I. S., Hervaise-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. Hum. Percept. Perform. **23**, 914–927.

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**, 224–238.

Ettlinger, M. (**2007**). "Shifting categories: An exemplar-based computational model of chain shifts," Paper presented at the 29th Annual Meeting of the Cognitive Science Society, Nashville, TN.

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.

Flege, J., 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., MacKay, I., and Meador, D. (**1999**). "Native Italian speakers' production and perception of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Flege, J., Schirru, C., and MacKay, I. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.

Flege, J. E., and Fletcher, K. L. (**1992**). "Talker and listener effects on the degree of perceived foreign accent," J. Acoust. Soc. Am. **91**, 370–389.

Frick, R. W. (**1985**). "Communicating emotion: The role of prosodic features," Psychol. Bull. **97**, 412–429.

Goggin, J., Thompson, C., Strube, G., and Simental, L. (**1991**). "The role of language familiarity in voice identification," Mem. Cognit. **19**, 448–458.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., and Stevens, E. B. (**1991**). "Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect," Percept. Psychophys. **50**, 524–536.

Greenspan, S., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of synthetic speech produced by rule," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 421–433.

IEEE Subcommittee (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Johnson, K. (**1997**). in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 145–166.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).

Kolers, P. A., and Roediger, H. L. III (**1984**). "Procedures of mind," J. Verbal Learn. Verbal Behav. **23**, 425–449.

Kraljic, T., and Samuel, A. G. (**2006**). "Generalization in perceptual learning for speech," Psychon. Bull. Rev. **13**, 262–268.

Kraljic, T., and Samuel, A. G. (**2007**). "Perceptual adjustments to multiple speakers," J. Mem. Lang. **56**, 1–15.

Kučera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English* (Brown University Press, Providence, RI).

Labov, W. (**1972**). *Sociolinguistic Patterns* (University of Pennsylvania Press, Philadelphia, PA).

Ladefoged, P., and Broadbent, D. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /l/. III: Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, G. D. (**1988**). "Toward an instance theory of automatization," Psychol. Rev. **95**, 492–527.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–885.

Luce, P. A., and Pisoni, D. D. (**1998**). "Recognizing spoke words. The neighborhood activation model," Ear Hear. **19**, 1–36.

Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol. Hum. Percept. Perform. **33**, 391–409.

McLennan, C. T., and Luce, P. A. (**2005**). "Examining the time course of indexical specificity effects in spoken word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **31**, 306–321.

Mullennix, J. M., Pisoni, D. B. , and Martin, C. S. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**, 365–378.

Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.

Munro, M. J. (**1993**). "Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings," Lang. Speech **36**, 39–66.

Munro, M. J. (**1998**). "The effects of noise on the intelligibility of foreign-accented speech," Stud. Second Lang. Acquis. **20**, 139–154.

Munro, M. J., and Derwing, T. M. (**1995**). "Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech," Lang. Speech **38**, 289–306.

Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.

Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cognit. Psychol. **47**, 204–238.

Nusbaum, H. C., and Magnuson, J. S. (**1997**). in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 109–132.

Nusbaum, H. C., and Morin, T. M. (**1992**). in *Speech Perception, Speech Production, and Linguistic Structure*, edited by Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (OHM, Tokyo), pp. 113–134.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," Research on Speech Perception: Progress Report **10**, 357–372.

Nygaard, L. C., Burt, S. A., and Queen, J. S. (**2000**). "Surface form typicality and asymmetric transfer in episodic memory for spoken words," J. Exp. Psychol. Learn. Mem. Cogn. **26**, 1228–1244.

Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific perceptual learning in spoken word recognition," Percept. Psychophys. **60**, 355–376.

Nygaard, L. C., Sommers, M., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). "Episodic encoding of voice attributes and recognition memory for spoken words," J. Exp. Psychol. Learn. Mem. Cogn. **19**, 309–328.

Pisoni, D. B. (**1997**) in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenni (Academic, San Diego, CA), pp. 9–32.

Schmid, P. M., and Yeni-Komshian, G. H. (**1999**). "The effects of speaker accent and target predictability on perception of mispronunciations," J. Speech Lang. Hear. Res. **42**, 56–64.

Schwab, E. C., Nusbaum, H. C., and Pisoni, D. B. (**1985**). "Some effects of training on the perception of synthetic speech," Hum. Factors **27**, 395–408.

Sommers, M. S., and Barcroft, J. (**2007**). "An integrated account of the effects of acoustic variability in first language and second language: Evidence form amplitude, fundamental frequency, and speaking rate variability," Appl. Psycholinguist. **28**, 231–249.

Van Lancker, D., Kreiman, J., and Emmorey, K. (**1985**). "Familiar voice recognition: Patterns and parameters. Part I. Recognition of backward voices," J. Phonetics **13**, 19–38.

van Wijngaarden, S. J., Steeneken, H. J., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.

Weil, S. A. (**2001**). "Foreign accented speech: Encoding and generalization," J. Acoust. Soc. Am. **109**, 2473 (A).

Yonan, C. A., and Sommers, M. S. (**2000**). "The effects of talker familiarity on spoken word identification in younger and older listeners," Psychol. Aging **15**, 88–99.

Cues to perception of reduced flaps

Natasha Warner[a)]
Department of Linguistics, University of Arizona, Tucson, Arizona 85721-0028 and Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands

Amy Fountain
Department of Linguistics, University of Arizona, Tucson, Arizona 85721-0028

Benjamin V. Tucker
Department of Linguistics, University of Alberta, Edmonton, Alberta T6G 2E7, Canada

(Received 25 March 2008; revised 19 February 2009; accepted 20 February 2009)

Natural, spontaneous speech (and even quite careful speech) often shows extreme reduction in many speech segments, even resulting in apparent deletion of consonants. Where the flap ([ɾ]) allophone of /t/ and /d/ is expected in American English, one frequently sees an approximant-like or even vocalic pattern, rather than a clear flap. Still, the /t/ or /d/ is usually perceived, suggesting the acoustic characteristics of a reduced flap are sufficient for perception of a consonant. This paper identifies several acoustic characteristics of reduced flaps based on previous acoustic research (size of intensity dip, consonant duration, and F4 valley) and presents phonetic identification data for continua that manipulate these acoustic characteristics of reduction. The results indicate that the most obvious types of acoustic variability seen in natural flaps do affect listeners' percept of a consonant, but not sufficiently to completely account for the percept. Listeners are affected by the acoustic characteristics of consonant reduction, but they are also very skilled at evaluating variability along the acoustic dimensions that realize reduction.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097773]

PACS number(s): 43.71.Es, 43.71.Sy [AJ] Pages: 3317–3327

I. INTRODUCTION

A quick look at any corpus of spontaneous speech shows that speakers do not produce every segment of a word and do not produce sounds as one would expect (Greenberg, 1997, 1999; Johnson, 2004; Pluymaekers *et al.*, 2005a, 2005b). For example, one of our recordings includes an American English utterance [bɹʌʒləɪˀ], "but I was like," in which the speaker deleted some segments, shifted the qualities of others, and inserted r-coloration. Still, native listeners understand the utterance easily. Which ways that speech sounds vary during reduction are important for speech recognition? How acceptable is it to a listener if the speaker changes the manner of articulation of a consonant, weakening it to an approximant, vs if the speaker shortens a consonant, or fails to produce a drop in intensity for the consonant, nearly deleting it? In this article, we focus on how reduction affects American English intervocalic /t/ and /d/ in flapping ([ɾ]) position (e.g. "pretty, prejudice"). Previous work (e.g., Koopmans-van Beinum, 1980; Arai, 1999; Ernestus *et al.*, 2002) shows that reduced words and sounds are quite difficult to perceive when removed from their context, although they are perceived well in context. This finding leads to the question of whether specific acoustic characteristics of reduction hinder perception.

Acoustically, prototypical flaps are characterized by a very brief closure, resembling a voiced stop closure except for its brevity (Port, 1977; Zue and Laferriere, 1979). Zue and Laferriere (1979) found an average duration of just 26–27 ms for flaps, as compared to 75 and 129 ms for prestress /d/ and /t/. They also found that the duration of the consonant does not differ for flap derived from /t/ vs /d/, although the duration of the preceding vowel does differ slightly. Fisher and Hirsh (1976) found some differences in duration and intensity between flapped /t, d/. Flaps are not expected to have a burst, as they are expected to be so short that air pressure cannot build up behind the closure. However, Zue and Laferriere (1979) also found flap-like tokens with surprisingly long closures more similar to a [d] but no burst, as well as tokens with short, flap-like closures with a clear burst. Horna (1998) confirmed that several variants, ranging from more stop-like to more approximant-like, are possible, with the approximant-like variants more common in conversation than in read speech. De Jong (1998) and Fukaya and Byrd (2005) both gave articulatory data on flaps, and argued that gradient gestural differences lead to the percept of flap as an acoustically different sound. Son (2008) showed that even if a clear tongue tip gesture occurs, there may be little or no acoustic sign of an expected flap in Korean. As for perception of flaps, Port (1977) found that short consonantal duration is such a strong cue that the percept of "rabbit" shifts entirely to "ratted" ([b] to [ɾ]) if the [b] is made short enough, despite the conflicting place cues. McLennan *et al.* (2003, 2005) and Connine (2004) investigated the perception of flapped /t, d/ in comparison to nonflapped stop [t] or [d] in words where flap would be expected (e.g., "atom, Adam"). However, since flap rather than stop is

[a)]Author to whom correspondence should be addressed. Electronic mail: nwarner@u.arizona.edu

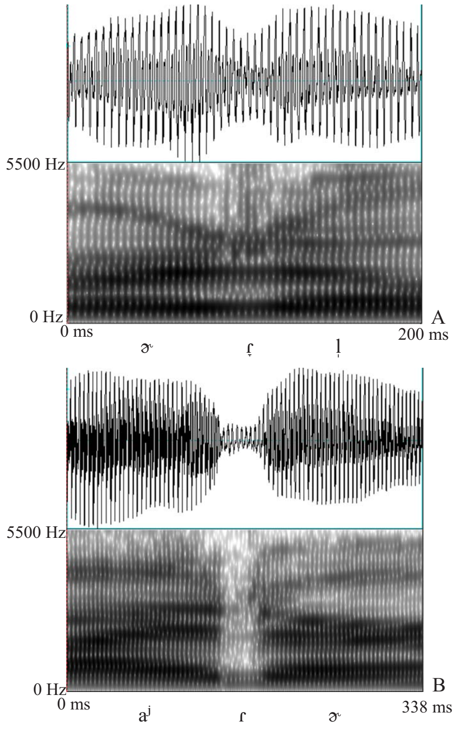


FIG. 1. (Color online) Spectrograms and waveforms of a reduced, approximant-like consonant in "fertilizer" (A) and a clear flap in "spider" (B) for intervocalic /t/or /d/.

the normal pronunciation in this environment, this investigates processing of allophonic variation rather than of speech reduction.

In our previous work (Warner, 2005; Warner and Tucker, 2007) on /t/ and /d/ in flapping position, we encountered many tokens with approximant-like /t, d/ (Fig. 1), in carefully read speech as well as conversation. Out of more than 1900 measurable tokens of /t/ or /d/ in flapping environment, produced by seven speakers, the second and/or third formants continued at least faintly throughout the consonant in 88% of tokens. In 56% of tokens, they continued strongly throughout the consonant, as one would expect for an approximant but not a true flap. During the /t/ or /d/, intensity dipped by an average of only 10.11 dB relative to the peaks of surrounding vowels, whereas /p/ and /k/ in comparable environments showed an average dip of 32.07 dB. Of the /t, d/ tokens in our study, 70% lacked a burst, 97.5% had voicing throughout the consonant, and many lacked a clear onset and offset. The average duration for the /t, d/ was 32 ms. An additional 78 tokens had the /t, d/ so thoroughly deleted that we could not locate any acoustic trace of it to measure. Furthermore, some tokens had a large valley in the fourth formant timed to the /t, d/ [Fig. 1(A)], even if the consonant was extremely reduced (Dungan *et al.*, 2007; Warner and Tucker, 2008). This valley in the F4 was particularly common following /r/ (e.g., "party, quarter"), visible in 46% of tokens (out of 120 read speech tokens with preceding /r/, from six speakers). It occurred in only 2% of tokens before the vowel /i/ and not after /r/ (e.g., "city").

Most reduced tokens in our previous work, despite resembling an approximant in the spectrogram, sounded clearly like a /t/ or /d/. Lexical and phonotactic expectations might contribute: "forty" is unlikely to be misperceived as a non-word /foɹi/, or "status" as phonotactically impossible /stæəs/. However, even when deletion of the flap would form a real word (e.g., "powder/power, needle/kneel"), very reduced tokens rarely sound ambiguous. Thus, even a very small acoustic cue may be sufficient for listeners to perceive the consonant.

This study investigates whether the acoustic dimensions that vary in natural productions influence listeners' percept of a /t, d/ consonant. Previous research on reduced speech has shown that listeners need surrounding context in order to recognize reduced segments or words well (Bard *et al.*, 1988; Arai, 1999; Ernestus *et al.*, 2002). Listeners also take account of how often segments reduce in various environments when compensating for reduction (Mitterer and Ernestus, 2006). Here, we turn to specific acoustic dimensions that vary during reduction in a particular segment, the flap. We investigate the effects of degree of intensity dip (size in decibels) (Experiment 1), duration of the consonant (Experiment 2), and size of F4 valley (Experiment 3) on perception of real-word pairs such as "powder/power." Intensity dip and duration of the consonant were the most reliable and variable measures in our previous production data. We choose to manipulate the F4 valley as well because such large, clear F4 valleys were a surprising finding in the production study. Furthermore, F4 valleys sometimes occurred even in otherwise very reduced /t, d/ tokens, so we wish to determine whether this acoustic characteristic could be a perceptual cue. We use re-synthesized continua in a phonetic identification task. Thus, when listeners hear a reduced /t, d/ as in Fig. 1(A), do they attend to the intensity dip (Experiment 1), the duration of the consonant (Experiment 2), and/or the valley in F4 (Experiment 3) when reconstructing the sound? The over-arching question, then, is what listeners attend to within the extreme variability of natural speech. We investigate flaps as one case of this variability, and in each experiment, we test one dimension observed in natural speech variability.

II. EXPERIMENT 1: MANIPULATING DEGREE OF INTENSITY DIP

The first experiment manipulates the degree of dip in intensity at the /t, d/. We have previously observed a wide range of intensity dips for flaps, from a large intensity dip indicative of tongue closure, even with cessation of voicing, through small intensity dips, to a few tokens with no intensity dip (Fig. 2, and cf. Fig. 1) (Warner, 2005; Warner and Tucker, 2007). Experiment 1 uses PRAAT resynthesis (Boersma and Weenink, 2008) to create two continua with a range of intensity dips, based on one word with a flap ("needle") and a matched word without ("kneel"). We predict that listeners will be less likely to hear a /t/ or /d/ in stimuli with only a small dip in intensity at the consonant. However, because we see many tokens with only minor intensity dips in natural productions, we also predict that listeners will still be able to perceive an intended /t/ or /d/ relatively often even if the intensity dip is small. That is, we predict an effect of degree of intensity dip on consonant perception, but not a large shift such as from 0% to 100%.

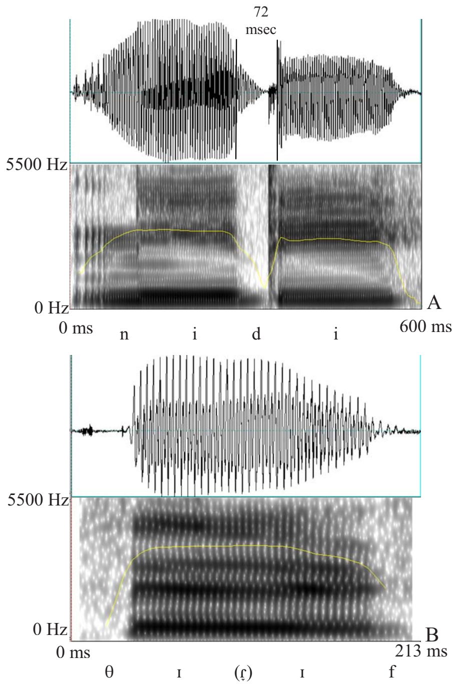


FIG. 2. (Color online) Waveforms, spectrograms, and overlaid intensity curves of /t, d/ realized with differing degrees of intensity dip. (A) Large intensity dip indicative of tongue closure, with voicing ending by the burst, in "needy." Consonant duration is marked by vertical lines (defined by F2 offset/onset). (B) No visible intensity dip for the consonant, in "…with it if…." Consonant duration cannot be measured.

A. Methods

1. Materials

A female speaker of American English produced multiple tokens of flap (VDV) and no-flap (VV) pairs, such as "powder/power, title/tile." This recording provided the tokens from which to create the stimuli for this experiment (intensity manipulation) as well as Experiments 2 and 3 (duration and F4 manipulations), which are reported below. Both underlying /t/ and /d/ words were included, but the speaker's dialect is that of Southern California, so she did not have a clear distinction between pairs such as "writer/rider" based on Canadian vowel raising or vowel length. She was recorded in a sound-attenuated booth at 44.1 kHz directly to hard drive, using a high-quality stand-mounted microphone. The word list consisted of 12 pairs of words, more than were planned for inclusion as stimuli, so that appropriate items for resynthesis could be selected from among them. The recording list used a pseudo-random order, with the members of a pair not contiguous. The speaker, a linguist but not a phonetician, did not know the topic was reduction. She was asked to produce the words several times, varying her speech from "careful but natural" to "sloppy." She produced varied VDV (flap) tokens, none unnaturally careful (e.g., not $[\text{t}^{\text{h}}\text{a}^{\text{j}}\text{t}^{\text{h}}\text{ɫ̩}]$ for "title"), but all within a range from flap to near-deletion, similar to the tokens produced by non-linguist subjects in our acoustic study (Warner, 2005; Warner and Tucker, 2007).

Experiment 1 used two continua, one based on a token of needle and the other based on a token of kneel. Tokens of other words from the recording were used for the other experiments, as described below. The base token of needle (for the VDV continuum) had a moderate intensity dip (4.3 dB relative to average of surrounding vowel peaks, approximately 49 ms) and clear second and third formants through the consonant. The speaker pronounced kneel as bisyllabic ([nijɫ̩]), but a token with minimal intensity dip (1.1 dB relative to average of neighboring vowel peaks, approximately 42 ms) was selected for the VV continuum. Both tokens were resynthesized using PRAAT's intensity editor (8 ms steps) to flatten the intensity contour throughout the natural dip, while the rest of the signal was multiplied by a constant to maintain the shape of the intensity contour outside that area. (This was done to maintain the natural onset and offset of the word.) This token, flattened throughout the consonant, was used as step 1 of each continuum (which should sound least like it contains a /t, d/), and it was used as the base from which to synthesize steps 2–8. For each subsequent step, we decreased the intensity values for eight time points (at 8 ms intervals) at the time of the original, natural dip. The third through the fifth time points of the dip were reduced in increments of 3 dB per step (e.g., 3 dB for step 2, 6 dB for step 3, and 21 dB for step 8). Thus, the original signal was not used as any continuum step. 21 dB was the most extreme, because a larger decrease sounded like a computer glitch rather than a consonant. (The continuum range was chosen not based on our previous production results, but as being the largest practical range. In our production data, the 5th and 95th percentiles for /t, d/ intensity dips fall at 2.82 and 18.66 dB, and the average at 10.11 dB, so the synthesized continuum is slightly larger than the typical natural range.) In order to ensure that extreme drops in intensity did not *de facto* generate longer durations of perceived intensity dip, the time points other than the most extreme part of the dip (points 1, 2, and 6 through 8) were set to 1–2 dB less than the surrounding consonant, and were held constant for all continuum steps. Thus, only the degree (in decibels) of the consonant-like dip varied, not its duration. In total, a 16 ms stretch was at the lowest amplitude, and the entire dip including the gradual ramp covered 72 ms. Figure 3 illustrates several resulting stimuli for the needle continuum. This procedure produced stimuli at the large-dip end of the continuum with the appearance of a brief closure and a slight ramping of intensity going into and out of it, as one often sees in natural productions that have clear flaps.

2. Participants and procedures

Thirty-four native speakers of American English, students in introductory linguistics courses at the University of Arizona, participated in the experiment. None reported any speech or hearing disorders. Data from additional participants who did not identify themselves as monolingual English speakers were not analyzed. Participants received a small amount of course credit.

Participants sat in a sound-attenuated booth and heard the stimuli over headphones. The task was two-alternative forced choice, with a button box for responses. Six repetitions of each stimulus (normalized for overall amplitude) were presented, in a single session randomized with the stimuli for Experiments 2 and 3 below. They were not

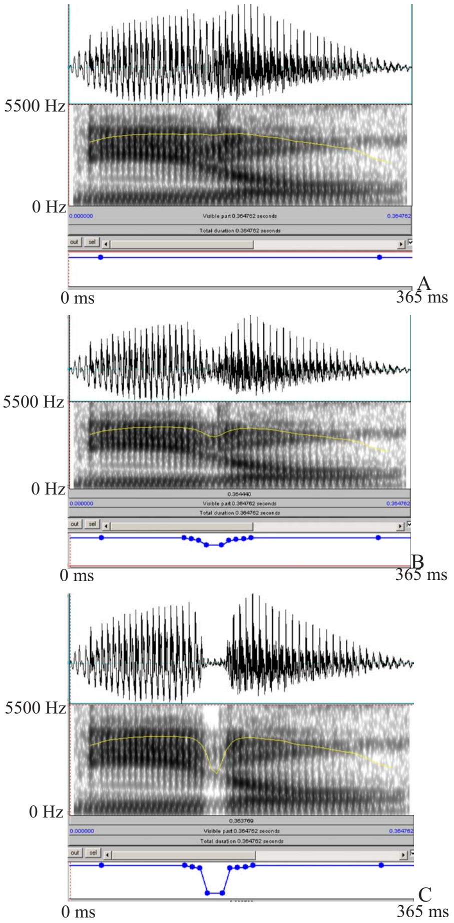


FIG. 3. (Color online) Example stimuli for the intensity VDV-base continuum (needle). Waveform, spectrogram, calculated intensity contour (overlaid on spectrogram), and stipulated intensity contour (multiplied by identical base token), for Steps 1 (flat), 3, and 8 (largest dip). (A), (B), and (C) are Steps 1, 3, and 8 respectively. Intensity is displayed over the same range for all three stimuli.

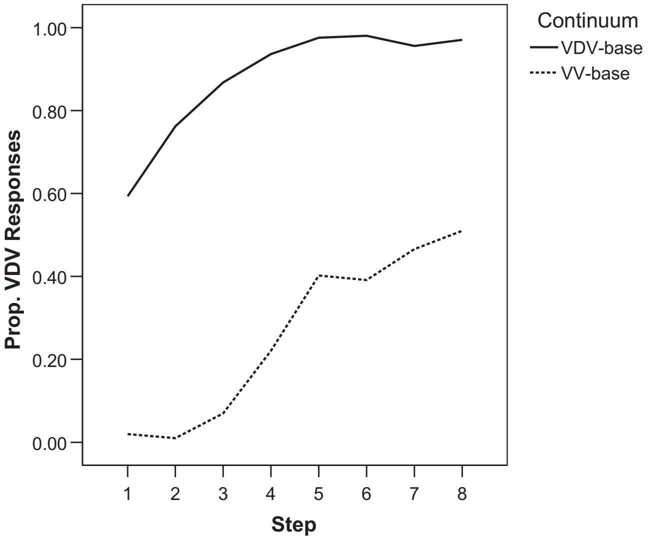


FIG. 4. Responses (proportion VDV) for Experiment 1, size of intensity dip continua. Low-step numbers have flat intensity through the consonant; high-step numbers have a large intensity dip.

blocked by continuum or experiment, as that might induce listeners to focus on the acoustic dimension manipulated in each continuum. For each stimulus, the two words of the pair were presented on a computer screen visible through the booth window. For half the subjects, the VDV (flap) item appeared consistently on the left and the VV item on the right, with the reverse for the other half. Subjects responded which word they had heard using the button box. For a few word pairs where deletion of the flap could lead to two alternative real words (e.g., "waiter vs weigher/wear"), both no-flap alternatives appeared.

Subjects first read a list of the target words, to be sure they would have all response options in mind. They then performed a practice test on similar materials, followed by the real test, with one break. For each item, the response options appeared on the screen, and 1 s later, the auditory stimulus began. Subjects had a 3 s window from onset of the auditory stimulus in which to respond. After a response, there was a 1 s pause before the following visual stimulus appeared. The EPRIME software (Psychology Software Tools, Inc.) controlled the experiment and recorded responses. The experiment took approximately 20 min. Subjects answered questions about their language and dialect background after the experiment.

B. Results

Proportion of VDV responses (Fig. 4) and reaction times (RTs), averaged over the six presentations of each stimulus, were analyzed using within-subjects analyses of variance (ANOVAs). RTs will not be presented, however, as they generally supported the patterns in the proportion VDV data, adding little information. The factors were step (1–8) and continuum (VDV needle vs VV kneel). A between-subjects control factor (VDV presented on left or right of screen) was included in the statistical analyses to remove variance. No subjects were outliers, so none were excluded.

The mean proportion of VDV responses showed significance for both main effects and their interaction [continuum: $F(1,32)=331.82$; step: $F(7,224)=71.75$; interaction: $F(7,224)=7.35$; all p's <0.001]. (The proportion of VV responses is always the inverse of VDV responses.) Both continua showed more VDV responses with larger intensity dips, with significant simple effects [VDV (needle) continuum: $F(7,224)=30.85$; VV (kneel) continuum: $F(7,224)=36.68$; both p's <0.001]. Still, listeners perceive both continua predominantly as the word from which the continuum was formed: the VV-base continuum shifts from approximately 0% to 50% VDV responses, while the VDV-base continuum covers the range from 60% to 100%.

The significant interaction shows that the shape of the identification curve differs for the two continua. The VV-base continuum shows lesser slope at the small dip (VV percept, low-step number) end of the continuum, and the VDV-base continuum shows a flattening of the curve at the opposite end of the continuum. Thus, the two continua seem to represent separate parts of a categorical perception curve, with neither continuum alone achieving a complete shift. To test this, we used interaction comparisons over restricted step ranges. An interaction comparison with the factors step

(Steps and 1 and 2 only) and continuum showed significance for both main effects [continuum: $F(1,32)=252.16$, $p<0.001$; step: $F(1,32)=12.90$, $p<0.005$] and their interaction [$F(1,32)=15.61$, $p<0.001$]. The simple effect of continuum step (Steps 1 and 2 only) was significant for the VDV-base continuum [$F(1,32)=14.96$, $p<0.005$], but not the VV-base continuum [$F(1,32)=1.00$, $p>0.05$]. A second interaction comparison of Step 5 vs Step 8 also showed significance for both main effects and their interaction [continuum: $F(1,32)=86.34$, $p<0.001$; step: $F(1,32)=6.24$, $p<0.02$; interaction: $F(1,32)=7.39$, $p<0.02$], but this time, the simple effect of step (5 vs 8) was significant for the VV-base [$F(1,32)=7.50$, $p<0.02$] but not the VDV-base ($F<1$) continuum. This shows that the VV-base continuum is flat at low steps (little or no intensity dip), while the proportion of VDV judgments is already increasing for the VDV continuum. For the large-dip steps (5 vs 8), the VDV continuum has already reached ceiling, but the VV continuum is still increasing. Both continua show parts of a categorical perception curve, as both have a plateau and a range showing increase. However, they show opposite parts of the complete curve, even though they cover the same range of intensity dips: 0–21 dB decrease.

C. Discussion

These data verify that the size of the intensity dip is a cue to the perception of a /t, d/ consonant. A large dip in intensity increases perception of a flap. Thus, when we see variability in natural speech in how deeply intensity drops during a word with flapped /t/ or /d/, this variability is indeed along an acoustic dimension that influences how consonantal the token sounds.

However, the effect of base continuum in this experiment is also quite large, and the two continua cover different parts of the categorical perception curve despite their equal acoustic range. Thus, there must be other acoustic cues. Even a large intensity dip (21 dB) does not make kneel sound as if it contained a /d/ more than about half the time. This is not a failure to include a sufficient continuum range: a larger dip in pilot stimuli sounded like a non-speech sound, and the needle continuum verifies that a larger dip is not necessary to reach 100% VDV judgments. In the other direction, even deletion of the intensity dip fails to make needle into kneel more than 50% of the time. This matches with our observation from the production study that even tokens with little intensity dip often have a clear consonantal percept. In Experiment 2, we turn to another acoustic dimension: duration, rather than degree, of the intensity dip.

III. EXPERIMENT 2: MANIPULATING THE CONSONANT DURATION

In natural speech data, the /t, d/ varies greatly in duration (Warner, 2005; Warner and Tucker, 2007). Even a clear flap closure is short, since this sound is defined by its quick "flap" of the tongue against the roof of the mouth, but there is still a considerable range of consonant duration. Our production data for /t, d/ had the 5th and 95th percentiles of consonant duration at 15 and 56 ms, respectively, with 32 ms as the average. Although reduced /t, d/ can be difficult to measure, consonant duration correlated well with degree of intensity dip, with clearer flaps being longer. We posited that duration might also be a significant cue. Experiment 2 manipulates duration of the consonant independently in order to determine whether duration is itself a salient cue to the /t, d/. Because of the correlation of duration with other measures of reduction (shorter durations for more reduced tokens), we predict that listeners will be less likely to perceive a /t, d/ with shorter duration. However, since the flap is inherently a very short sound, even an extremely short flap may be rather perceptible, at least if it has a clear intensity dip or gap in formant structure.

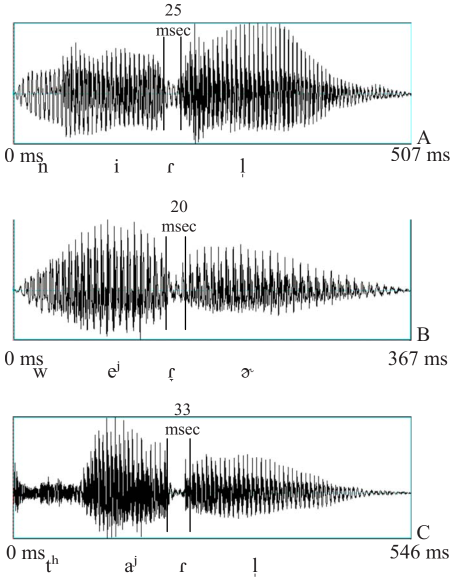


FIG. 5. (Color online) Waveforms of the three base tokens used to make the duration continua. Vertical lines delimit the consonant (defined by onset/offset of clear F2), and the duration of the consonant is shown. (A) Needle, a typical clear flap. (B) Waiter, an approximant realization. (C) Title, a very clear flap with extremely low amplitude and near-voicelessness during the closure.

A. Methods

Three tokens of VDV words (Fig. 5) were chosen from the same recording as in Experiment 1, from which to resynthesize the duration continua. One token of needle was a clear flap realization, with a substantial intensity dip (11.5 dB relative to surrounding vowel peaks), sudden onset and offset of the tongue closure as evidenced by a gap in formants, and voicing throughout. One token of "waiter" was an approximant-like realization, with a smaller dip in intensity (7.3 dB) and no sudden change in formants indicative of closure. Finally, one token of "title" represented an unusually clear flap realization: this token had at most extremely low amplitude voicing during the consonant (16.0 dB intensity dip), and could perhaps be considered voiceless.

For this experiment, no VV word was used as a base form, because it would not be clear what portion of the signal to lengthen or shorten to manipulate consonant duration. However, the use of three VDV-base words probes whether the effect of consonant duration differs for various realizations of the consonant. Because all of the base tokens were

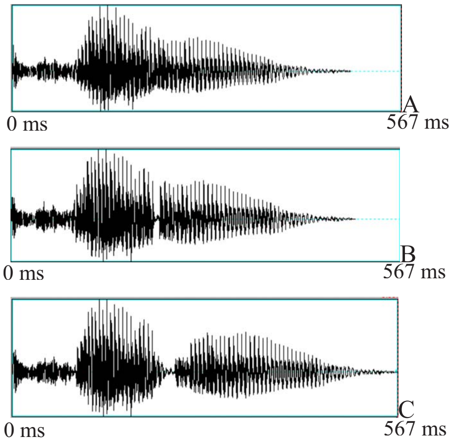


FIG. 6. (Color online) Waveforms of Steps 1 (A), 2 (B), and 8 (C) of the title duration continuum.

intended as VDV words, they were used as the sixth step of the eight-step continuum (near the VDV end).

We used PSOLA resynthesis within PRAAT to manipulate consonant durations (Fig. 6). A region covering the intensity dip was located for each base form. This region extended to near the intensity maxima for the surrounding vowels, and was thus larger than what would be measured as the consonant duration. PSOLA was then used to lengthen this region to 1.2 and 1.4 times its original duration for continuum Steps 7 and 8, and to shorten it to 0.8, 0.6, 0.4, 0.2, and 0 times its duration for Steps 5 through 1. Since shortening the region of the intensity dip also tends to lessen the degree of the dip (in decibels), we then located the lowest amplitude glottal period in the original signal and spliced it into the resynthesized forms for Steps 2–5 and 7 and 8, replacing the lowest amplitude period of each one. This guaranteed that each stimulus above Step 1 did drop to the same extent (in decibels) as it originally did, however, briefly. This was not done for Step 6 (the original item), or for Step 1, which had no intensity dip. This single low-amplitude period had a duration very similar to the period it replaced, usually within 1 ms, so that this manipulation did not affect the step-wise manipulation of duration. The duration range for the continuum was thus chosen not based on our production data, but on the base tokens used. For Step 2 (the shortest dip without deletion of it, 0.2 of original duration), the resulting duration of the manipulated portion of the signal (more than the consonant itself) was 10–11 ms for each continuum. Despite the necessity of rapid intensity changes, splicing at zero-crossings at consistent points of the glottal pulse, and the use of PSOLA, prevented the introduction of spurious burst-like noises. For Step 8 (the longest dip, 1.4 of original duration), the resulting duration of the manipulated portion was 66 ms for needle and waiter, and 74 ms for title.

The subjects and procedures were identical to those for Experiment 1. As described above, the stimuli for all three experiments were presented in a single session, in random order.

B. Results

The proportion VDV results (Fig. 7) was analyzed using an ANOVA with the within-subjects factors continuum (near-voiceless closure, voiced closure, approximant) and step (1–8), and the same between-subjects control factor (response side of screen) as above. Both main effects and the interaction were significant [continuum: $F(2,64)=15.94$; step: $F(7,224)=65.99$; interaction: $F(14,448)=17.69$; all p's <0.001]. The simple effect of step showed significantly more VDV responses at longer consonant durations for all three continua [near-voiceless: $F(7,224)=8.77$; voiced closure: $F(7,224)=70.21$; approximant: $F(7,224)=21.29$; all p's <0.001]. However, this includes Step 1, which lacks an intensity dip entirely.

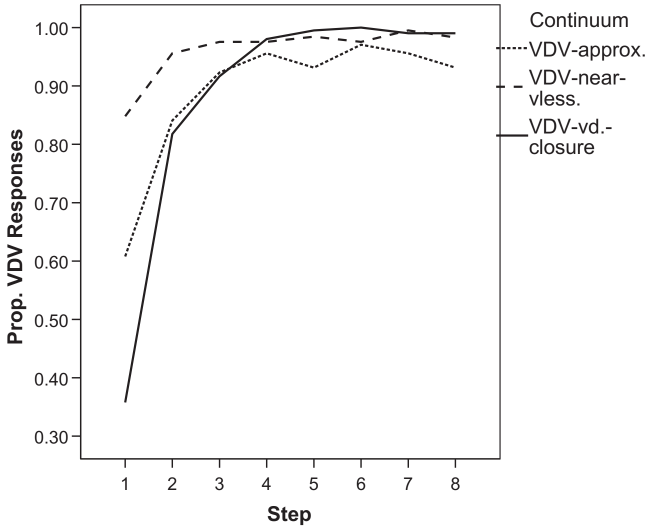


FIG. 7. Responses (proportion VDV) for duration continua. Low-step numbers have shortened or removed consonantal duration; high-step numbers have lengthened consonant duration. The original productions are at Step 6.

To be sure that duration of the consonant's dip, rather than just its presence, affects the percept, we used an interaction comparison of only Steps 2–5, the region containing ambiguity outside the no-dip first step. Both main effects as well as the interaction were significant [continuum: $F(2,64)=5.10$, $p<0.01$; step: $F(3,96)=12.54$, $p<0.001$; interaction: $F(6,192)=3.94$, $p<0.005$], and the simple effects showed an increase across this duration range for the voiced closure continuum [$F(3,96)=14.19$, $p<0.001$] and the approximant continuum [$F(3,96)=4.11$, $p<0.01$], but not the near-voiceless continuum [$F(3,96)=1.21$, $p>0.05$]. Thus, if the /t, d/ is realized as a less obstruent-like consonant, longer duration makes it sound more consonantal and shorter duration makes it sound more deleted. However, if its intensity dips very low, then even an extremely short dip of effectively one glottal period is sufficient to make the /t, d/ quite perceptible (VDV response at ceiling). Furthermore, even though two continua do show effects of consonant duration without Step 1, both already receive more than 80% VDV judgments by Step 2. Thus, even an approximant-like realization does not need a long duration to be perceived most often as containing a /t/ or /d/. It appears that almost any dip in intensity, no matter how small in degree or duration, can be perceived as a realization of /t, d/.

Even with extremely long consonant duration, the approximant continuum in Fig. 7 never reaches the high level of VDV responses the other two continua do. In an interaction comparison of Steps 4–8, only the main effect of continuum was significant [continuum: $F(2,64)=7.52$, $p<0.005$; step: $F<1$; Interaction: $F(8,256)=1.01$, $p>0.05$]. The voiced closure continuum had more VDV responses than the approximant continuum [main effect of continuum for just these two: $F(1,32)=10.29$, $p<0.005$], but it did not differ from the near-voiceless continuum [$F(1,32)=1.15$, $p>0.05$]. Thus, the voiced closure continuum patterns with the near-voiceless continuum in being at ceiling for longer durations, while the approximant continuum is not at ceiling. If a /t, d/ is realized as an approximant, rather than as a clear flap, even durations up to 1.4 times natural duration do not render it unambiguously consonantal.

These comparisons (Steps 2–5 and 4–8) together show that the voiced closure continuum patterns with the approximant continuum at short consonant durations, but with the near-voiceless continuum at long consonant durations. If the intensity dip is short, it has to be a very clear dip to definitively sound like a realization of /t, d/, but if it is long, a slight dip is sufficient.

Because Step 1 lacks the consonantal dip entirely, comparing Steps 1 and 2 shows how much the presence vs absence of even a very short dip affects the percept. (This is different from Experiment 1 above, which lacked short intensity dips.) In an interaction comparison of just Steps 1 and 2, the estimated effect size of the main effect of step was 0.71 (partial eta-squared), whereas in an interaction comparison of Steps 2–8, it was 0.34. The majority of the increase in VDV responses happens between Steps 1 and 2, not at longer durations (Fig. 7). Thus, the presence of any intensity dip at all has more impact on the percept of a consonant than even a large difference in the duration of that dip (from 0.2 to 1.4 times the original duration).

C. Discussion

These results clarify several points. First, consonant duration is clearly a perceptual cue to /t, d/ in flapping environment. However, in the continuum we tested with a very low-intensity consonant (near-voiceless closure), any duration of consonant at all is sufficient to cue its presence. In the continua with a fully voiced closure or a reduced, approximated consonant, the change from short to moderate duration increases perception of the consonant. These data also demonstrate that the presence vs absence of any intensity dip at all has a far greater effect on perception of a /t, d/ than even a large difference in consonantal duration. An extremely short dip, even one lacking consonantal closure, still contributes greatly to listeners' percept of a /t, d/.

Finally, the continuum with the middle degree of closure (voiced closure) patterns with the approximant continuum at short durations, but with the near-voiceless one for long consonants. If the consonant is short, only the particularly strong closure suffices for it to be definitively perceived. However, if the consonant is long, the weaker closures we tested still lead listeners to perceive the consonant. The approximant we tested, however, is not fully perceived as a realization of /t, d/, and lengthening it does not make it any more like a /t, d/. Thus, even though natural speech very often has /t, d/ realized as approximants, they are not fully accepted by the listener. However, this decrement for long approximated consonants is small: the /t, d/ is perceived in over 90% of such stimuli.

IV. EXPERIMENT 3: MANIPULATING THE F4 VALLEY

In our previous production work, we noticed a surprisingly large change in F4 in some tokens where a flap would be expected (Dungan *et al.*, 2007; Warner and Tucker, 2008). This valley in F4 [Fig. 1(A)] can traverse a range of 1000 Hz, and it is timed to the flap consonant, not to a neighboring segment. Some speech sounds do have systematic effects on F4, including retroflexes, American English /r/, and taps or flaps in some other languages (Espy-Wilson *et al.*, 2000; Avelino and Kim, 2002; Hamann, 2003; Zhou *et al.*, 2007). However, systematic effects on F4 are rare. In our previous work, we found that this F4 valley is most common after /r/ (e.g., 'hurdle, fertilizer,' 46% of tokens) and least common before /i/ (e.g., 'beauty,' 2% of tokens). It does not occur in the majority of tokens (visibly in 18% of tokens overall), and F4 is not always clear, but when the valley does occur, it is often a striking visual effect. Furthermore, even some tokens with extremely reduced /t, d/ have a clear F4 valley. Since such tokens appear to offer few other cues to the consonant, we wondered whether the F4 valley was a perceptual cue. Because the F4 valley can be so striking, and can be the only apparent acoustic realization of the consonant, we predict that listeners can make some use of an F4 valley to detect a /t, d/. However, listeners may have little reason to attend to F4 in the language overall, and F4 has low amplitude. For this reason, we predict that any effect of the F4 valley will be small. To test this, we manipulated F4, using LPC resynthesis in PRAAT.

A. Methods

1. Materials

Three tokens, two of "quarter" (VDV) and one of "core" (VV), were chosen from the same recording used for Experiment 1. The speaker pronounced quarter with onset /k/ matching core, not a /kw/ cluster. One token of quarter we selected had a clear F4 valley and a relatively large intensity dip. F4 was clearly visible in the spectrogram throughout, a requirement for this experiment. A second token of quarter had a clear F4 valley but minimal intensity dip. That is, the /t/ in this token was nearly deleted, but the F4 valley remained. Finally, a token of core with clearly visible F4 but no valley in it was chosen as a matched VV item. The use of two VDV items allows for comparison of a continuum where the F4 valley might be the primary cue to the /t/ to one where other cues are obvious. All three tokens were downsampled to 11 000 Hz, and formants were located through LPC analysis, using a prediction order of 10 for core and of 12 for both tokens of quarter. (The higher prediction order was necessary for accurate tracking of F4, in order to resynthesize without leaving traces of the original F4.) We then inverse filtered the tokens, using the LPC analysis, to obtain estimated glottal source functions. After manipulating the formant values

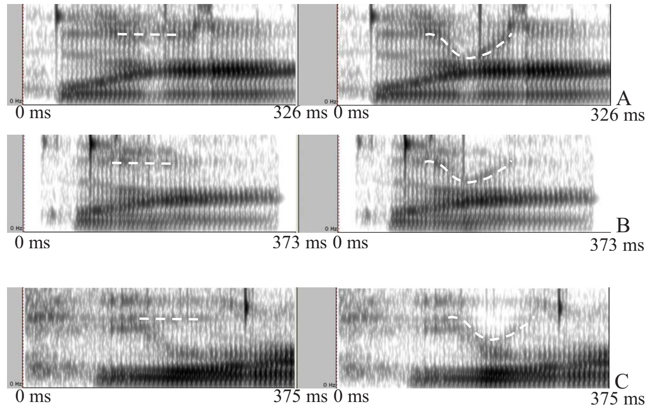


FIG. 8. (Color online) Spectrograms of Step 1 (no F4 valley, on left) and Step 8 (maximal F4 valley, on right) stimuli, for (A) quarter (VDV-intensity dip), (B) quarter (VDV-F4 valley only), and (C) core (VV). Overlaid dashed lines trace F4.

(5 ms time step) as described below, we resynthesized to create the stimuli. Figure 8 shows the resynthesized extreme steps for all three continua.

The formants were manipulated as follows. The time range of interest, lasting 85–100 ms, was located. It was from the beginning of rapid decrease to the end of rapid increase in F4 in the two tokens of quarter, and for a range timed similarly relative to the onset of the preceding vowel for core (which had no F4 valley). Any outlier points in the F4 LPC track during that time range were manually corrected. For the two VDV (quarter) continua, the F4 at Step 8 of the continuum (maximal F4 valley) used the original values, except for such outlier correction. For Step 1, the F4 was measured at the time points immediately outside this time period, and the F4 was interpolated (in barks) from the preceding to the following value over the time range. For Steps 2–7, the difference between the F4 for Step 8 (natural) and Step 1 (interpolated, straight F4) was divided into perceptually equal steps, as measured in barks. The range of the F4 valley size was thus determined by the base token, not by overall averages from production results. Because F4 is not clear in all naturally produced tokens and the valley does not always occur, we chose to base the continuum range on the naturally produced valley in a token with a clear F4 valley, rather than on production data averaged across tokens.

For the VV (core) continuum, there was no original F4 valley, so one had to be added to create Step 8. Furthermore, the speaker's F3 dropped for the /r/ considerably later in core than in quarter, and this meant that if F3 were not manipulated the added F4 valley would cross over the F3. Therefore, for core, the F3 was lowered to 504 Hz below the F4 value of Step 8 (lowest F4) for the descending portion of the F4 valley. After reaching the minimum, the F3 remained at that value until the natural F3 dropped below that value later in the word, at which point the natural F3 values were allowed to resume. (We did not wish to create a drop-rise pattern in F3 as well as F4.) This 504 Hz separation was based on the average for the vowel up until the manipulated time period. The same lowering of F3 was applied to all steps of the core continuum, so that only F4 would vary by continuum step. Once F3 was thus moved out of the way of F4, a somewhat parabola-shaped F4 valley of 1000 Hz was created for Step 8, modeled on the properties of the natural F4 valleys in the VDV continua. For the VV continuum, the natural F4 values were used as Step 1, and the difference between Step 1 (natural) and Step 8 (added F4 valley) was calculated and divided among the other steps as for the VDV continua.

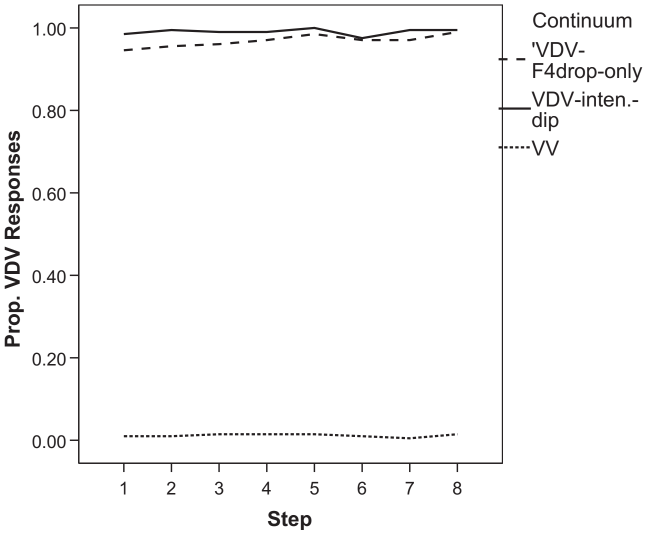


FIG. 9. Responses (proportion VDV) for the F4 continua. Low-step numbers have flat F4 (natural for VV); high step numbers have a large F4 valley (natural for VDV).

After resynthesis, the low-step stimuli lacked any F4 valley, showing steady F4. The high-step stimuli had a clear F4 valley resembling that of natural tokens. Those steps with original F4 values were also resynthesized, so that they would be equally degraded by LPC resynthesis. The stimuli for this experiment did contain LPC clicking noises that the other experiments' stimuli did not. However, they were still clear realizations of either quarter or core.

2. Subjects and procedures

Subjects and procedures were identical to Experiments 1 and 2. All three experiments were presented together, so the LPC-degraded stimuli were randomized among the higher quality PSOLA and intensity resynthesis items. The instructions mentioned that some items sounded like computer speech, and the practice items included some created through LPC resynthesis.

B. Results

ANOVAs were used to analyze the data (Fig. 9) with the within-subjects factors continuum (VDV-intensity dip, VDV-F4 valley only, VV) and step (1–8) and the usual between-subjects control factor (response side of screen). Only the main effect of continuum was significant, with both VDV-base continua perceived as VDV far more often than the VV continuum was [continuum: $F(2,64)=27.12$, $p<0.001$; step: $F(7,224)=1.72$, $p>0.05$; interaction: $F<1$].

One VDV continuum (F4 valley only) might show some effect of step, despite the lack of an interaction. In order to check for any possible effect of the F4 valley, we performed *post hoc* comparisons of Steps 1 and 2 to Step 8 for this continuum. Step 8 was identified as VDV significantly more often than either Step 1 $[F(1,32)=5.31,\ p<0.03]$ or Step 2 $[F(1,32)=6.13,\ p<0.02]$.

C. Discussion

The stimuli in this experiment were nearly always perceived as the base word from which they were formed, regardless of how F4 was manipulated. The presence of an F4 valley does affect perception, but the effect is extremely small, and is limited to the VDV-base continuum that lacked a strong dip in intensity. Based on what is known about perceptual cue-trading (e.g., Repp, 1983), it is not surprising that any perceptual effect of the F4 valley is limited to the continuum with the highest chance of ambiguity, where the /t/ was very approximant-like, with intensity nearly as great as that of the surrounding vocalic sounds. The formants continued strongly throughout the consonant, and there was certainly no consonant closure. The valley in F4 might be important, because it is the only visually clear trace of the /t/ remaining. If listeners ever attend to the F4 valley, it would be in an approximated token such as this one. However, the results show minimal use of the F4 valley. Furthermore, listeners perceived a /t/ in even this continuum at nearly ceiling, in more than 94% of tokens even for Step 1 (with flattened F4). Thus, there must be ample perceptual cues to the /t/ aside from F4 valley.

Still, the fact that there is any perceptual effect of the F4 at all is noteworthy. The fourth formant does not provide important cues to other segmental distinctions in English, as far as we know. Although American English /r/ may affect F4 (Zhou *et al.*, 2007), its low F3 is a far more likely cue (Best and Strange, 1992). The literature shows L2 listeners have difficulty learning to attend to acoustic dimensions not used for L1 distinctions (Best and Strange, 1992; Iverson *et al.*, 2003; Wagner *et al.*, 2006). Thus, English listeners should not be very good at using F4 variation as a cue. Furthermore, because we randomized the stimuli of all three experiments (total of 8 continua, varied on three dimensions), listeners probably could not learn to attend to the F4 valley over the course of the experiment. In addition to F4 not being a perceptual cue otherwise, it is also acoustically weak, with lesser amplitude than the lower formants. These factors seem to outweigh the fact that a drop of 1000 Hz in the F4 is a striking acoustic characteristic. The result is a minimal, but present, perceptual effect of the F4 valley.

V. GENERAL DISCUSSION

These experiments show a large effect of the degree of intensity dip on perception of a /t, d/ in flapping environment, a relatively large effect of the presence vs absence of any intensity dip at all, a smaller effect of consonant duration, and an extremely small effect of the F4 valley. Larger and longer intensity dips are perceived as more consonantal, but even a very short intensity dip can be enough to cue the presence of /t, d/. Furthermore, there must be other cues to the reduced consonant, as none of the continua spans the 0%–100% response range. It is possible that the dimensions we manipulated would show a greater shift in the absence of other cues (cue-trading, see Repp, 1983). However, we used a variety of base forms, making multiple continua that differed in the presence and strength of alternative cues, to allow for this possibility. Experiment 2, manipulating duration, exemplifies this with the use of three base tokens for resynthesis, differing widely in the other cues to the consonant. Because we did not systematically manipulate two or more cues at once through an entire range, we cannot be sure of potential cue trading. It could be that the F4 valley would show a larger effect with other cues more ambiguous, and it could be that duration might cause a larger effect if intensity were manipulated simultaneously. However, the use of several base tokens that vary widely on the most likely other cues does provide information about the probable range of effect sizes for each cue.

Experiment 3 manipulated F4, adding or removing a large drop-rise pattern. This was based on our observation of a striking valley in F4, often traversing 1000 Hz from surrounding vowels, in some tokens where flapped /t, d/ is expected. Experiment 3 showed that despite the large acoustic change, the F4 valley had minimal effect on listeners, and even that only when there were no other obvious cues to the consonant. This suggests that the valley observed in our production work is probably an articulatory artifact, having to do with a constriction the tongue moves through between targets for surrounding segments. This artifact may appear striking on a spectrogram, but not be so to listeners. However, listeners are slightly able to use this fine phonetic detail of the speech signal, despite the reasons for them to ignore it (i.e., low amplitude of F4 and lack of importance of F4 for other distinctions). In future research, it would be possible to test the F4 valley while manipulating duration of the pre- and post-valley portions and degree of intensity dip, to determine whether the F4 valley might play a larger role in a cue-trading relationship. However, we suspect that F4 is simply not used as a major perceptual cue.

Duration of the consonant clearly is a cue to its presence. However, Experiment 2 showed that the presence of any intensity dip in the signal at all, no matter how short it is, is more important than even a very large difference in consonant duration. Because the flap is normally a very brief consonant, usually with an intensity dip but otherwise with quite a bit of variability, listeners may categorize almost any brief intensity dip as a realization of /t, d/ most of the time.

Experiment 1 shows that the degree of the intensity dip (its size in decibels rather than its duration) is a relatively strong cue to the /t, d/. However, even these continua fail to traverse the 0%–100% range, despite covering the largest practical acoustic range. Furthermore, the continua based on a VDV vs a VV token represent different parts of the categorical perception curve. Thus, even in this case, there must be other perceptual cues.

In typical categorical perception experiments, the continuum should ideally cover the range from 0% to 100% responses. However, the fact that this did not happen here is not a failure of the experiments. Since we tested acoustic dimensions that vary in reduced vs careful realizations, if any of these dimensions led to a VDV-base stimulus receiving 0% VDV responses, it would mean that listeners were unable to understand reduced realizations of the word. We are testing for whether the dimensions that vary in natural productions of one category (VDV) affect perception of that category. They do, but not to the extent of fully turning tokens into the other category (VV). This makes sense: if a smaller intensity dip made "title" into "tile," then the amount of variability present in natural speech would mean that there was a merger in progress, rather than simply synchronic variability in what seem to be stable categories.

The three experiments together indicate that there must be additional cues to the presence vs absence of a "flapped" /t, d/, in addition to the ones tested here. Some possible cues are the duration of the preceding and following segments (vowels /r/ or syllabic /l/), and the timing of formant transitions between surrounding segments. We do not attempt to identify and test every potential cue to the VV/VDV distinction in this paper. Rather, we set out to determine whether the acoustic dimensions that vary when speakers reduce their flaps affect how /t, d/-like the resulting sound is. Since reduction sometimes resembles deletion, how do the dimensions of reduction affect the listener's percept? The results show that the dimensions that differ between clear vs reduced realizations do affect how likely listeners are to perceive a consonant.

However, approximated forms are accepted as realizations of /t, d/ almost, but not quite, at ceiling rates. A highly reduced flap does not entirely count as a flap, but it is very acceptable. Listeners are very tolerant of the kinds of variation that happen in natural reduced speech. They are sensitive to this variation, and our other work shows they use reduction in deciding what realizations to expect in upcoming speech, adjusting for reduction in the context (Tucker, 2007). However, these same dimensions that cue reduction are also cues to the presence or absence of the consonant. Listeners are very skilled at combining cues to segmental content with their knowledge of variability and reduction to perceive the intended word.

The vast majority of speech perception research investigates acoustic cues to the distinction between one sound and another—the presence of one sound vs the presence of another. There is a long history of research on place of articulation, voicing, or other distinctions such as fricatives vs affricates (Raphael, 2005). There has also been work on cues to the presence vs absence of some segments, e.g., in the "say/stay" and "slit/split" distinctions (Repp, 1983; Raphael, 2005). The current study is similar in investigating the presence vs absence of intervocalic /t, d/, as in "waiter/weigher, needle/kneel," etc. However, it is about more than cues to the presence of a consonant. It is also about perceptual use of the cues that differentiate a clear production of a segment from a reduced one. Reduction is rampant in natural speech (Greenberg, 1997, 1999; Pluymaekers *et al.*, 2005a, 2005b; Johnson, 2004), and surprisingly common even in careful speech (Warner and Tucker, 2007). Thus, to understand how listeners comprehend speech, we need to study acoustic characteristics that vary with reduction.

As a phonetician, when one records quarter or title and sees an approximant or a vocalic sequence rather than a flap, one is often surprised. We noticed many tokens in our production study (Warner, 2005; Warner and Tucker, 2007) that were articulatorily not flaps (no consonantal closure), and that was part of the motivation for this study: how do listeners react to variations in whether a consonant has a closure? Are approximated, non-canonical realizations acceptable where a flap is expected? Does reduction make the consonant less consonantal? Although we as phoneticians may be surprised by such non-flap-like realizations in a spectrogram, listeners are, in fact, less surprised. Reduction does affect how clearly listeners perceive a /t, d/, but listeners are very much able to cope with such variability. This is exactly what listeners should do: the acoustic dimensions of reduction tested here are dimensions that vary during natural speech. Listeners are tolerant of highly variable speech that contains both nearly deleted versions of flaps and clear versions of them, because such variability is typical in the spontaneous, natural language listeners hear most often. Therefore, listeners use the information available in the duration of the consonant and the degree of its intensity dip, but they also evaluate these cues relative to the wide range of realizations of flapped /t, d/ they normally hear.

ACKNOWLEDGMENTS

The authors would like to thank Holger Mitterer, James McQueen, Alexandra Jesse, and Anne Cutler, as well as three anonymous reviewers for their helpful comments. They also thank Jesse Tucker and Kayla McDaniel for assistance with the work. Any errors are, of course, our own.

Arai, T. (**1999**). "A case study of spontaneous speech in Japanese," in Proceedings of the International Congress of Phonetic Sciences (ICPhS), San Francisco, Vol. **1**, pp. 615–618.

Avelino, H., and Kim, S. (**2002**). "An articulatory and acoustic study of Pima coronals," J. Acoust. Soc. Am. **112**, 2419 (Abstract).

Bard, E. G., Shillcock, R. C., and Altmann, G. T. M. (**1988**). "The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context," Percept. Psychophys. **44**, 395–408.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305–330.

Boersma, P., and Weenink, D. (**2008**). "Praat: doing phonetics by computer (Version 5.0.08) [Computer program]," http://www.praat.org/ (Last viewed February, 2008).

Connine, C. M. (**2004**). "It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition," Psychon. Bull. Rev. **11**, 1084–1089.

de Jong, K. (**1998**). "Stress-related variation in the articulation of coda alveolar stops: Flapping revisited," J. Phonetics **26**, 283–310.

Dungan, M., Morian, K., Tucker, B. V., and Warner, N. (**2007**). "Fourth formant dip as a correlate of American English flaps," J. Acoust. Soc. Am. **121**, 3167 (Abstract).

Ernestus, M., Baayen, R. H., and Schreuder, R. (**2002**). "The recognition of reduced word forms," Brain Lang **81**, 162–173.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., and Alwan, A. (**2000**). "Acoustic modeling of American English /r/," J. Acoust. Soc. Am. **108**, 343–356.

Fisher, W. M., and Hirsh, I. J. (**1976**). "Intervocalic flapping in English," Papers from the Regional Meetings, Chicago Linguistic Society, pp. 183–198.

Fukaya, T., and Byrd, D. (**2005**). "An articulatory examination of word-final flapping at phrase edges and interiors," J. Int. Phonetic Assoc. **35**, 45–58.

Greenberg, S. (**1997**). "On the origins of speech intelligibility in the real world," in Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels, Pont-a-Mousson, France, pp. 23–32.

Greenberg, S. (**1999**). "Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation," Speech Commun. **29**, 159–176.

Hamann, S. (**2003**). *The Phonetics and Phonology of Retroflexes* (LOT, the Netherlands).

Horna, J. E. (**1998**). "An Investigation into the Acoustics of American English Flaps, with a secondary emphasis on Spanish Flaps, in fluent speech," Ph.D. thesis New York University.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Johnson, K. (**2004**). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the First Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa (The National International Institute for Japanese Language, Tokyo, Japan), pp. 29–54.

Koopmans-van Beinum, F. J. (**1980**). "Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions," Ph.D. thesis, University of Amsterdam, Amsterdam, The Netherlands.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (**2003**). "Representation of lexical form," J. Exp. Psychol. Learn. Mem. Cogn. **29**, 539–553.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (**2005**). "Representation of lexical form: Evidence from studies of sublexical ambiguity," Immunopharmacol Immunotoxicol **31**, 1308–1314.

Mitterer, H., and Ernestus, M. (**2006**). "Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch," J. Phonetics **34**, 73–103.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005a**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **118**, 2561–2569.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005b**). "Articulatory planning is continuous and sensitive to informational redundancy," Phonetica **62**, 146–159.

Port, R. F. (**1977**). "The influence of tempo on stop closure duration as a cue for voicing and place," Haskins Labs Status Report on Speech Res. **SR-51/52**, 59–73.

Raphael, L. J. (**2005**). "Acoustic cues to the perception of segmental phonemes," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford), pp. 182–206.

Repp, B. H. (**1983**). "Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization," Speech Commun. **2**, 341–361.

Son, M. (**2008**). "Pitfalls of spectrogram readings of flaps," J. Acoust. Soc. Am. **123**, 3079.

Tucker, B. V. (**2007**). "Spoken word recognition of the reduced American English flap," Ph.D. thesis, University of Arizona, Tucson, AZ.

Wagner, A., Ernestus, M., and Cutler, A. (**2006**). "Formant transitions in fricative identification: The role of native fricative inventory," J. Acoust. Soc. Am. **120**, 2267–2277.

Warner, N. (**2005**). "Reduction of flaps: speech style, phonological environment, and variability," J. Acoust. Soc. Am. **118**, 2035 (Abstract).

Warner, N., and Tucker, B. V. (**2007**). "Categorical and gradient variability in intervocalic stops," presented at the Linguistic Society of America Annual Meeting, Anaheim, CA.

Warner, N., and Tucker, B. V. (**2008**). "Fourth formant drop as a correlate of American English flaps," presented at the Linguistic Society of America Annual Meeting, Chicago, IL.

Zhou, X., Espy-Wilson, C., Tiede, M., and Boyce, S. (**2007**). "Acoustic cues of 'retroflex' and 'bunched' American English rhotic sound," J. Acoust. Soc. Am. **121**, 3168 (Abstract).

Zue, V. W., and Laferriere, M. (**1979**). "Acoustic study of medial /t, d/ in American English," J. Acoust. Soc. Am. **66**, 1039–1050.

Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing[a]

Olaf Strelcyk and Torsten Dau
Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Building 352, Ørsteds Plads, 2800 Kgs. Lyngby, Denmark

(Received 23 August 2008; revised 11 February 2009; accepted 12 February 2009)

Frequency selectivity, temporal fine-structure (TFS) processing, and speech reception were assessed for six normal-hearing (NH) listeners, ten sensorineurally hearing-impaired (HI) listeners with similar high-frequency losses, and two listeners with an obscure dysfunction (OD). TFS processing was investigated at low frequencies in regions of normal hearing, through measurements of binaural masked detection, tone lateralization, and monaural frequency modulation (FM) detection. Lateralization and FM detection thresholds were measured in quiet and in background noise. Speech reception thresholds were obtained for full-spectrum and lowpass-filtered sentences with different interferers. Both the HI listeners and the OD listeners showed poorer performance than the NH listeners in terms of frequency selectivity, TFS processing, and speech reception. While a correlation was observed between the monaural and binaural TFS-processing deficits in the HI listeners, no relation was found between TFS processing and frequency selectivity. The effect of noise on TFS processing was not larger for the HI listeners than for the NH listeners. Finally, TFS-processing performance was correlated with speech reception in a two-talker background and lateralized noise, but not in amplitude-modulated noise. The results provide constraints for future models of impaired auditory signal processing. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097469]

PACS number(s): 43.71.Ky, 43.66.Sr, 43.66.Nm, 43.66.Pn [BCJM] Pages: 3328–3345

I. INTRODUCTION

Hearing-impaired (HI) people often experience great difficulty with speech communication when background noise is present. While audibility has been shown to be the main determinant of speech reception in quiet, it does not account to the same degree for speech reception in noise (e.g., Plomp, 1978; Dreschler and Plomp, 1985; Glasberg and Moore, 1989). Consequently, for many HI listeners, the problem persists even if reduced audibility has been compensated for by hearing aids. Other impairment factors besides reduced audibility must be involved.

Relations between frequency selectivity and speech reception, particularly in noise, have been reported previously (e.g., Festen and Plomp, 1983; Dreschler and Plomp, 1985; Horst, 1987; van Schijndel *et al.*, 2001). Recently, also the processing of temporal fine-structure (TFS) information has received considerable attention with regard to speech reception (e.g., Tyler *et al.*, 1983; Buss *et al.*, 2004; Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008). While envelope cues are sufficient to achieve good speech reception in quiet (e.g., Shannon *et al.*, 1995), TFS cues may be required to ensure good speech reception in noise (e.g., Nie *et al.*, 2005; Lorenzi and Moore, 2008). In particular, it has been suggested that deficits in TFS coding might account for the limited ability of HI listeners to take advantage of amplitude fluctuations in a noise background, i.e., to listen in the dips of a fluctuating interferer (e.g., Qin and Oxenham, 2003; Lorenzi *et al.*, 2006; Gnansia *et al.*, 2008). However, the large variability of performance that is commonly observed across HI listeners makes it difficult to compare results across studies. Hence, only limited conclusions can be drawn about the relations between the different auditory functions, such as frequency selectivity and the processing of TFS. Also the relation between the deficits observed in monaural and binaural TFS processing remains unclear. Knowledge of these relations might shed light on the actual mechanisms and sites of the impairments.

Therefore, in the present study, individual performance on frequency selectivity, monaural and binaural TFS processing, and speech reception was measured using a common set of listeners. This is a similar concept to that used in the studies of Hall *et al.* (1984) and Gabriel *et al.* (1992), who examined binaural performance in individual HI listeners. Since the primary objective of the present study was to investigate impairment factors beyond audibility, ten HI listeners with similar high-frequency hearing losses were selected to provide a homogeneous group in terms of audibility. In this way, confounding effects of audibility were minimized and more direct conclusions could be drawn from a relatively small number of subjects about possible relations between the tested auditory functions. On the flip side, however, this group of HI listeners represents one homogeneous subset of the overall HI population and therefore one should act with caution in generalizing the results.

Besides the HI listeners, two further subjects were included in the present study. Despite normal audiograms, these subjects complained about difficulties with speech reception in noisy backgrounds. In literature, different terms have been used to refer to this phenomenon: auditory disabil-

[a] Portions of this work were presented at the 155th Meeting of the Acoustical Society of America, Paris, July 2008 (J. Acoust. Soc. Am. **123**, 3712).

© 2009 Acoustical Society of America

ity with normal hearing (King and Stephens, 1992), obscure auditory dysfunction (Saunders and Haggard, 1989), and King–Kopetzky syndrome (Hinchcliffe, 1992). For simplicity, in the present study, these subjects are referred to as having an obscure dysfunction (OD). In view of the heterogeneity of the clinical group of OD patients (e.g., Saunders and Haggard, 1989; Zhao and Stephens, 2000), these two listeners cannot constitute a representative sample and therefore should be regarded as cases. The comparison of performance between the two OD listeners and the HI listeners may provide valuable information on the nature of the underlying impairments in both groups.

Speech reception thresholds (SRTs) for full-spectrum and lowpass-filtered speech were measured in different diotic and dichotic interferers. The other psychoacoustic tests in this study were designed to examine basic auditory functions, mainly at a frequency of 750 Hz. Low-frequency information has been shown to play a dominant role both for monaural abilities, such as the perception of pitch of complex tones (e.g., Terhardt, 1974; Moore *et al.*, 1985), and for binaural abilities such as sound localization (e.g., Wightman and Kistler, 1992). Therefore, the frequency of 750 Hz was chosen to investigate the potential impact of a hearing impairment on auditory processing at low frequencies, even if a hearing loss in terms of elevated audiometric thresholds was present only at higher frequencies. As a basic auditory function, frequency selectivity was estimated via the notched-noise paradigm in simultaneous masking (e.g., Patterson and Nimmo-Smith, 1980).

Throughout the present study, the terms TFS information and TFS processing refer to the temporal fine structure at the output of the cochlear filters. This fine structure evokes phase-locked activity, i.e., synchronized timing of action potentials, in the subsequent stages of neural processing (see Ruggero, 1992, for a review). Apart from phase locking, TFS information may also be coded in terms of a conversion from frequency modulation to amplitude modulation (FM-to-AM) on the cochlear filter skirts, as has been suggested for the detection of high-rate FM (Zwicker, 1956; Moore, 2003). In the present study, however, the focus lies on TFS processing based on phase locking, rather than on the FM-to-AM conversion mechanism.

Evidence for TFS-processing deficits in HI listeners has been found in previous studies of monaural as well as binaural auditory functions. In terms of binaural processing, TFS deficits have been observed in the detection of interaural time or phase differences via lateralization (e.g., Hawkins and Wightman, 1980; Häusler *et al.*, 1983; Smoski and Trahiotis, 1986; Gabriel *et al.*, 1992; Koehnke *et al.*, 1995; Lacher-Fougère and Demany, 2005). Also studies on binaural masked detection or masking level differences (MLDs) have reported deficits in HI listeners (e.g., Hall *et al.*, 1984; Staffel *et al.*, 1990; Gabriel *et al.*, 1992). In both tasks, lateralization and binaural detection, the interaural phase or time differences in the stimuli can only be coded in terms of phase-locking-based TFS processing (see Stern and Trahiotis, 1995 and Colburn, 1996). Apart from these *binaural* measures of TFS processing, frequency discrimination of tones with frequencies of up to 4–5 kHz is thought to be determined by a temporal mechanism based on phase locking (see Moore, 2003). Hence, deficits observed in the frequency discrimination of steady pure tones (e.g., Turner and Nelson, 1982; Tyler *et al.*, 1983; Turner, 1987; Freyman and Nelson, 1991) and in the detection of low-rate FM (e.g., Zurek and Formby, 1981; Grant, 1987; Lacher-Fougère and Demany, 1998; Moore and Skrodzka, 2002; Buss *et al.*, 2004) have been interpreted to indicate deficits in *monaural* TFS processing in HI listeners. This conclusion has been further supported by studies of frequency discrimination with harmonic complex tones (e.g., Horst, 1987; Moore *et al.*, 2006; Hopkins and Moore, 2007). However, since none of the above mentioned studies has obtained both monaural and binaural measures of TFS processing, it remained unclear to what extent the deficits observed in the binaural tasks were due to monaural or independent binaural deficits.

Only a few studies have assessed the relation between TFS deficits and speech reception performance. Tyler *et al.* (1983), Glasberg and Moore (1989), Noordhoek *et al.* (2001), and Buss *et al.* (2004) found significant correlations between frequency discrimination performance and word recognition in speech-shaped noise (SSN) as well as quiet, while Horst (1987) did not find such correlations. Lorenzi *et al.* (2006) and Hopkins *et al.* (2008), using processed speech stimuli, presented evidence that HI listeners were less able to make use of the TFS information in speech than normal hearing (NH) listeners. However, in these studies, the potential contribution of reduced frequency selectivity to the observed TFS deficits remained unclear. Reduced frequency selectivity might have affected the processing of TFS information in several ways (see also Moore, 2008). For wideband signals, the outputs of broadened auditory filters would exhibit a more complex TFS than the outputs of "normal" filters (Rosen, 1987). In addition, the signal-to-noise ratio (SNR) in the presence of a wideband interferer would be smaller in the case of broadened filters, providing a less favorable input to the subsequent processing stages. Finally, parts of the preserved TFS information in the speech stimuli of Lorenzi *et al.* (2006) might have been coded in terms of FM-to-AM conversion through cochlear filtering (e.g., Zeng *et al.*, 2004; Gilbert and Lorenzi, 2006). In such a case, filter broadening would result in reduced AM depths at the filter output and a less distinct representation of frequency transitions (e.g., downward and upward glides) across adjacent filters. Hence, the observed deficits in the TFS processing of wideband stimuli could, in principle, have resulted from reduced frequency selectivity rather than from deficits in subsequent auditory processing stages.

Therefore, the present study investigated potential deficits in phase-locking-based TFS processing, where possible effects of frequency selectivity should play a minor role. Nevertheless, the relation between frequency selectivity and TFS processing was examined here since both might be affected by a common underlying impairment factor such as outer hair cell (OHC) damage. The TFS processing was addressed binaurally through measurements of binaural masked detection and lateralization of pure tones with ongoing interaural phase differences (IPDs). As a complementary monaural measure, detection thresholds for low-rate frequency

TABLE I. Audiometric information for the ten HI listeners and the two listeners with OD. The ears that were tested on monaural FM detection are marked by asterisks.

				Audiometric thresholds (dB HL)											
ID	Gender	Age	Ear	125	250	500	750	1000	1500	2000	3000	4000	6000	8000	Etiology
HI$_1$	F	24	L*	5	−5	0	5	15	25	35	60	60	55	60	
			R	0	−5	0	5	15	25	30	55	55	65	70	Hypoxia at birth
HI$_2$	M	53	L	5	0	5	10	5	15	30	45	40	50	55	
			R*	5	5	0	10	5	20	30	40	45	55	60	Unknown
HI$_3$	M	55	L*	0	5	10	15	15	10	55	70	55	60	55	
			R	0	5	15	15	20	10	30[a]	60	65	55	60	Noise induced
HI$_4$	M	56	L*	5	5	5	5	5	25	30	45	45	55	70	
			R	−5	0	−5	5	5	15	25	50	50	55	65	Hereditary
HI$_5$	M	60	L*	10	5	10	10	5	35	60	60	60	55	60	
			R	0	0	5	10	5	40	50	65	60	60	60	Noise induced
HI$_6$	M	67	L*	0	5	5	10	10	15	30	45	50	50	65	
			R	0	5	5	10	5	15	20	45	60	65[a]	65	Unknown
HI$_7$	M	70	L	5	5	0	10	15	10	20	50	65	60	60	
			R*	5	10	0	10	15	15	15	55	55	60	65	Noise induced
HI$_8$	F	70	L*	5	5	15	20	20	30	45	60	65	60	60	
			R	5	5	10	15	20	35	45	60	60	60	55	Unknown
HI$_9$	M	74	L	20	10	5	10	20	40	35	60	60	60	70	
			R*	20	15	5	10	10	50	60[a]	60	60	55	55[a]	Noise induced
HI$_{10}$	F	74	L*	0	10	10	15	20	45	55	65	65	60	70	
			R	5	5	10	15	15	35	55	55	60	60	60	Noise induced
OD$_1$	F	26	L	−5	−5	0	0	0	0	0	0	0	10	5	
			R*	−5	−5	0	0	0	0	0	5	0	5	5	None
OD$_2$	F	46	L	0	5	5	5	10	5	5	5	0	10	5	
			R*	0	0	0	5	10	5	10	0	10	10	5	None

[a] Thresholds differ by more than 10 dB between the ears.

modulation (FMDTs) were obtained. The IPD thresholds and FMDTs were measured in quiet as well as in continuous noise backgrounds in order to test the robustness of the TFS processing to interfering noise. Physiological animal studies (e.g., Rhode *et al.*, 1978; Abbas, 1981; Costalupes, 1985) have shown that phase locking to tones in the presence of background noise is generally preserved at SNRs near behavioral detection thresholds but ceases at sufficiently low SNRs. However, as no comparable studies exist in impaired hearing, it cannot be excluded that hearing impairment might potentiate the susceptibility of phase locking to noise disturbance.

II. METHODS

A. Listeners

The six NH listeners (three females and three males) were aged between 21 and 55 years (median: 28) and had audiometric thresholds better than 20 dB hearing level (HL; ISO 389-8, 2004) at all octave frequencies from 125 to 8000 Hz and 750 to 6000 Hz. The ten HI listeners (three females and seven males) were aged between 24 and 74 years (median: 63). Detailed audiometric information is given in Table I. Throughout the study, the HI subjects are sorted by age and the notation "HI$_n$" is used to refer to the individual subject with index n. The audiograms were "normal" up to 1 kHz (thresholds ≤20 dB HL) and sloping at higher frequencies to values of up to 70 dB HL. All listeners had bilaterally symmetric audiograms (within 10 dB, exceptions stated in Table I), to avoid the issue of level balancing in binaural testing, as discussed in Durlach *et al.*, 1981. The sensorineural origin of the hearing losses was established by means of bone-conduction measurements, tympanometry, and otoscopy. The etiologies stated in Table I were based on the subjects' reports. They ranged from hypoxia at birth (oxygen deficiency) and hereditary losses to noise-induced losses, either sudden or due to sustained exposure to intense sounds. The remaining two subjects had OD: Despite audiometric thresholds better than 15 dB HL at all test frequencies (see Table I), they approached the research center, complaining about difficulties with understanding speech in noisy backgrounds. Their middle-ear status was normal and they did not report any history of otitis media or excessive noise exposure. Auditory brainstem responses were measured for these two subjects and the HI subject HI_{10} (since HI_{10} showed diverging results in the lateralization task). As the responses were normal, there was no indication of eighth-nerve tumors, brainstem lesions, or auditory neuropathy. Additionally, all listeners were screened on a binaural pitch task, testing the ability to hear a Huggins' pitch *C*-scale (Santurette and Dau, 2007). Santurette and Dau (2007) suggested that the absence of a binaural pitch percept might indicate the presence of a severe central auditory deficiency. Since all listeners in the present study perceived the pitch there was no indication of such a deficiency.

Each subject completed all tests, with the exception of one NH listener, for whom SRTs were not measured. The average testing time was 24 h per listener. All experiments were approved by the Ethics Committee of Copenhagen County.

B. Apparatus

All stimuli were generated in MATLAB® and converted to analog signals using a 24-bit digital-to-analog converter (RME DIGI96/8). The sampling rate was 44.1 kHz for the speech reception measurement, 48 kHz for the masking and FM experiments, and 96 kHz for the lateralization task. The stimuli were presented in a double-walled sound-attenuating booth via Sennheiser HD580 headphones. Calibrations were done using a B&K 4153 artificial ear and, prior to playing, 128-tap linear-phase FIR equalization filters were applied to all broadband stimuli, rendering the headphone frequency response flat.

C. Statistical analyses

To accommodate the repeated-measures design, the statistical analyses were carried out using linear mixed-effects models (Laird and Ware, 1982; Pinheiro and Bates, 2000), as implemented in S-PLUS®. The between-subject variability that was not explained in terms of the fixed effect subject group (or interactions of other fixed effects such as stimulus condition with subject group) was accounted for in terms of subject-specific random effects. In addition to analyses of variance (ANOVAs) and multiple comparisons of the fixed effects (with simultaneous 95% confidence intervals, either based on the Dunnett method or Monte Carlo simulations), the estimated random effects were extracted. They served as ranks for the individual listeners' performance on a given test, for example, binaural masked detection or lateralization. In the following, the abbreviations SD and CI will be used for standard deviation and confidence interval, respectively.

III. SPEECH RECEPTION

A. Method

SRTs were measured for Danish closed-set Hagerman sentences (Dantale II, Wagener *et al.*, 2003) in the presence of different interferers: a stationary SSN with the long-term spectrum of the Dantale II sentences, a sinusoidally and a randomly amplitude-modulated noise (SAM and RAM), a multitalker and a reversed two-talker background (MULTITALK and TWOTALK), and a dichotic, lateralized speech-shaped noise (LATSSN). Specifically, the SAM noise was fully sinusoidally amplitude-modulated SSN, with a modulation rate of 8 Hz (cf. Füllgrabe *et al.*, 2006). The RAM noise was randomly amplitude-modulated SSN, with the Hilbert envelope of a 20-Hz-wide noise used as modulator. The MULTITALK noise was a reversed 20-talker babble (supplied as track 3 on compact disk CD101R3 "Auditory Tests Revised" by AUDiTEC of St. Louis). The TWOTALK noise consisted of running female and male speech, with silent gaps longer than 250 ms removed, mixed at equal level, and time-reversed (speech supplied as tracks 8 and 9 on compact disk CD B&O 101 "Music for Archimedes" by Bang & Olufsen). The LATSSN (noise) was SSN, which was lateralized to one side by means of a constant interaural time difference of 740 μs. For a given run, either the left or the right ear was leading, but the SRT was averaged across runs with lateralization to the left and right. In addition to these conditions with full-spectrum speech, two conditions with filtered speech were used, SSN_{filt} and SAM_{filt}, in which both target speech and interferer were lowpass filtered at 1 kHz (1024-tap FIR lowpass filter designed using the Parks–McClellan algorithm in MATLAB®). This was done to test the processing of speech information in the regions of normal hearing (as all listeners had normal audiometric thresholds up to 1 kHz). The SRTs in all the aforementioned conditions were measured binaurally with the target speech and interferer presented diotically, with the exception of the LATSSN condition, where the interferer was presented dichotically. In addition, SRTs in the SSN and SAM conditions were measured monaurally, for comparison with the other monaural tests of frequency selectivity and FM detection.

The SRT was defined as the SNR leading to 50% correct identification of the individual words in the Dantale II sentences. The interferer level was kept constant at 65 dB sound-pressure-level (SPL) while the sentence level was varied adaptively. In each condition, the listeners were trained on a single run of 20 sentences. Subsequently, the SRT was estimated as the average over two to three runs, depending on the condition. A monotonic improvement of threshold in a sequence of three runs was interpreted as a training effect. When such an effect occurred, further runs were taken until stable performance was reached, and the first runs were discarded. This procedure for dealing with training effects was applied to all the other tests in this study.

B. Results and discussion

Figure 1 shows the binaural SRTs for the NH (circles), the OD (bold numbers), and the HI listeners (plain numbers). The horizontal black bars denote the mean SRTs for the NH

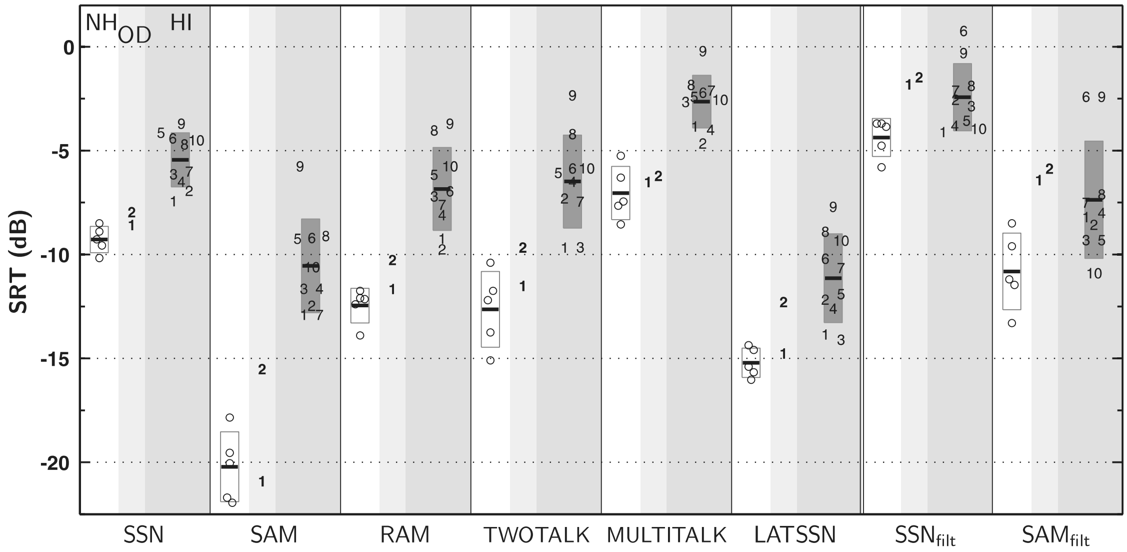


FIG. 1. SRTs for the NH listeners (circles), the two listeners with OD (bold numbers), and the HI listeners (plain numbers). The different conditions are indicated at the bottom of each panel. The horizontal black bars show the mean SRTs for the NH and HI listeners and the corresponding boxes represent ±1 SD.

and HI listeners and the corresponding boxes represent ±1 SD. Considering the first six conditions with full-spectrum speech, all listeners showed the lowest SRTs with the SAM and LATSSN interferers, while the highest SRTs were obtained with the MULTITALK interferer. The SRTs for the RAM and TWOTALK conditions lay slightly below those for the stationary SSN interferer. Performance in the conditions with lowpass-filtered speech (SSN$_{\mathrm{filt}}$ and SAM$_{\mathrm{filt}}$) was generally poorer than performance in the corresponding conditions with full-spectrum speech. An ANOVA was performed on the SRTs of the NH and HI listeners. The SRTs were found to be significantly higher for the HI listeners than for the NH listeners $[F(1,13)=36.1,\ p<0.0001]$. The SRTs differed significantly across conditions $[F(7,91)=238.7,\ p<0.0001]$ and the interaction between listener group and condition was significant $[F(7,91)=20.5,\ p<0.0001]$. Multiple comparisons revealed that the HI listeners performed more poorly than the NH listeners for all full-spectrum conditions $[p<0.001]$. For the two conditions with lowpass-filtered speech, the HI listeners' deficits were less pronounced. The deficit was significant for the SAM$_{\mathrm{filt}}$ condition $[p<0.01]$, but not for the SSN$_{\mathrm{filt}}$ condition $[p>0.05]$. Within the group of HI listeners, no significant correlation was observed between the SRTs for the filtered speech and the full-spectrum speech $[p>0.05]$. Hence, they did not seem to make equally good use of the low-frequency and high-frequency information in the speech stimuli. For example, listeners HI$_5$ and HI$_{10}$ performed relatively well in the filtered-speech task, but poorly in the full-spectrum speech task.

Previously, Horwitz *et al.* (2002) measured speech reception performance of HI listeners in regions of normal hearing, using lowpass-filtered speech in a SSN masker. In contrast to the present results (SSN$_{\mathrm{filt}}$ condition), they found significantly poorer performance for their HI than for their NH listeners. However, their speech stimuli were presented at a level of 77 dB SPL, where a substantial spread of excitation on the basilar membrane would be expected, particularly toward places corresponding to higher characteristic frequencies. Consequently, they interpreted their finding in terms of a reduced ability of their HI listeners to encode the information at places with high characteristic frequencies, where a hearing loss was present.

In addition to the SRTs, speech masking release was considered, i.e., the gain in terms of SRT for the SAM, RAM, TWOTALK, and LATSSN conditions when compared with the SRT for the stationary, diotic SSN condition. The group masking release values can be extracted from Fig. 1 as the differences in SRT between the corresponding conditions. The masking release values were significantly smaller for the HI listeners than for the NH listeners $[F(1,13)=21.8,\ p=0.0004]$. While the SAM masking release values for the full-spectrum speech differed strongly between the NH and HI listeners [by 5.8 dB, $p<0.001$], the difference for the filtered speech just reached significance [1.5 dB, $p=0.05$]. The finding of less pronounced deficits with lowpass-filtered speech may, at least partly, be attributed to the fact that the HI listeners had normal low-frequency hearing thresholds and that the full-spectrum speech stimuli were not amplified to fully restore audibility at high frequencies. It is interesting that the HI listeners did not benefit from high-frequency information in terms of the SAM masking release: While the NH listeners showed a significantly larger masking release with full-spectrum speech (SAM–SSN) than with filtered speech (SAM$_{\mathrm{filt}}$−SSN$_{\mathrm{filt}}$) [difference in dB: 4.5 (CI 3.3, 5.7)], the difference was not significant for the HI listeners [0.2 (CI –0.9, 1.2) dB].

As mentioned above, higher SRTs were observed in the MULTITALK masker than in the SSN masker, independent of listener group. Hence, in addition to the energetic masking present for the latter, another detrimental masking effect must have limited speech intelligibility in the case of the MULTITALK background. This could, for example, have been the complex harmonic structure of the background babble, which interfered with the use of spectro-temporal cues in the target speech, such as formant transitions.

As can be seen in Fig. 1, the OD listeners showed rather small deficits in the reception of full-spectrum speech. Consistent with previous reports in literature (e.g., Middelweerd *et al.*, 1990; Saunders and Haggard, 1992), they often performed at the lower limit of the NH group. Subject OD_1 showed elevated SRTs only in the two filtered-speech conditions. Subject OD_2 showed poorer performance than the NH listeners in all conditions except MULTITALK. Particularly in the SAM, LATSSN, and both filtered-speech conditions, her SRTs were increased relative to those for the NH group. Hence, for these two listeners, a deficit with speech reception was most apparent for the lowpass-filtered speech. This deficit might reflect a general difficulty understanding speech that is less redundant than full-spectrum speech. However, it could also reflect a specific problem with the processing of low-frequency information.

The monaural SRTs, which were measured only in the SSN and SAM conditions, closely followed the corresponding binaural results described above (monaural SRTs were, on average, 1.5 dB higher than binaural SRTs). The mean monaural SSN SRT of −8.7 (SD 0.9) dB for the NH listeners was consistent with the SRT of −8.4 (SD 1.0) dB reported by Wagener *et al.* (2003). Since the monaural results do not provide any further insights they are not presented in detail.

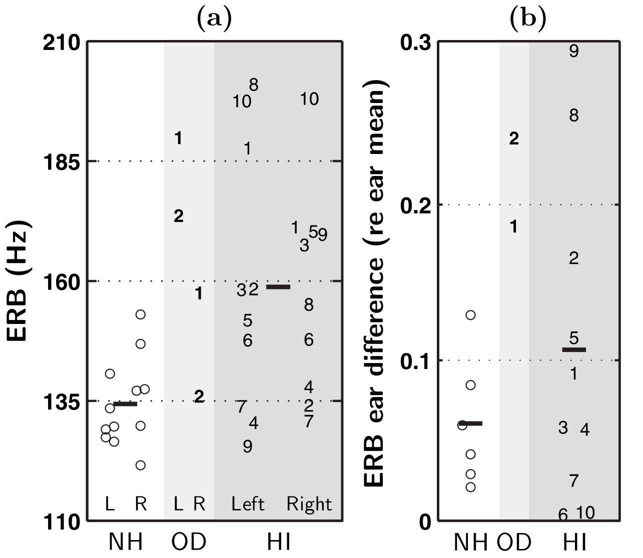


FIG. 2. (a) ERB of the roex(p,r) filter estimates at 750 Hz for the NH listeners (circles), the two listeners with OD (bold numbers), and the HI listeners (plain numbers). For each group, the left and right symbols or numbers correspond to the left and right ears, respectively. The horizontal black bars denote group means. (b) Absolute value of the ERB differences between the ears, divided by the mean ERB for the two ears.

IV. FREQUENCY SELECTIVITY

A. Method

Auditory-filter shapes at 750 Hz were determined separately for each ear using a notched-noise paradigm (cf. Patterson and Nimmo-Smith, 1980). Rosen *et al.* (1998) presented evidence that auditory-filter shapes are output driven. Under the assumption of the power-spectrum model (cf. Patterson and Moore, 1986) that a constant SNR at the output of the auditory filter is required for detection, this is equivalent to saying that the filter shape is determined by the level of the target signal rather than the noise masker. Therefore, here, in order to obtain a faithful filter estimate, the signal level was kept constant while the masker level was varied adaptively. The 750-Hz target tones of 440-ms duration were presented at a fixed level of 50 dB SPL and were temporally centered in the 550-ms noise maskers. Maskers and tones were gated with 50-ms raised-cosine ramps. The noise was generated in the spectral domain as fixed-amplitude random-phase noise (this holds also for the noises in all remaining tests). Five symmetric ($\delta f/f_0$: 0.0, 0.1, 0.2, 0.3, and 0.4) and two asymmetric notch conditions ($\delta f/f_0$: 0.2|0.4 and 0.4|0.2) were used, where δf denotes the spacing between the inner noise edges and the signal frequency f_0. The outside edges of the noise maskers were fixed at $\pm 0.8 f_0$.

A three-interval, three-alternative, forced-choice (3I-3AFC) weighted up-down method (Kaernbach, 1991) was applied to track the 75% correct point on the psychometric function. A run was terminated after 14 reversals. The threshold was defined as the arithmetic mean of all masker levels following the fourth reversal. Following a training run for each notch condition, the threshold was estimated as the average over three runs. If the SD of these three runs exceeded 1 dB, one or two additional runs were taken and the average of all was used.

A nonlinear minimization routine was implemented in MATLAB® to find the best-fitting rounded-exponential filter in the least-squares sense, assuming that the signal was detected using the filter with the maximum SNR at its output. Middle-ear filtering was taken into account, using the middle-ear transfer function supplied by Moore *et al.* (1997). However, the results presented in the following do not depend on this choice. Furthermore, besides the equivalent rectangular bandwidth (ERB) as a measure of filter tuning, also the 3-dB and 10-dB bandwidths were considered. However, because they yielded essentially identical results, for ease of comparison only the ERB results will be discussed further.

B. Results and discussion

The roex(p,r) filter model (Patterson *et al.*, 1982) provided a good description of the individual notched-noise threshold data, with a rms fitting error of 0.64 (SD 0.25) dB, averaged across all subjects.[1] Figure 2(a) shows the estimated ERBs for the NH and HI listeners as well as the two OD listeners. The HI listeners showed, on average, significantly higher bandwidths than the NH listeners [$F(1,14)=13.5$, $p=0.003$], by a factor of 1.2. However, the results varied considerably across the HI listeners, with four of them showing bandwidths in both ears within the range of the NH listeners. In addition to the ERB, significantly shallower lower and upper filter skirts were observed for the HI listeners than for the NH listeners [lower skirt: $F(1,14)=10.9$, $p=0.005$; upper skirt: $F(1,14)=5.6$, $p=0.03$].

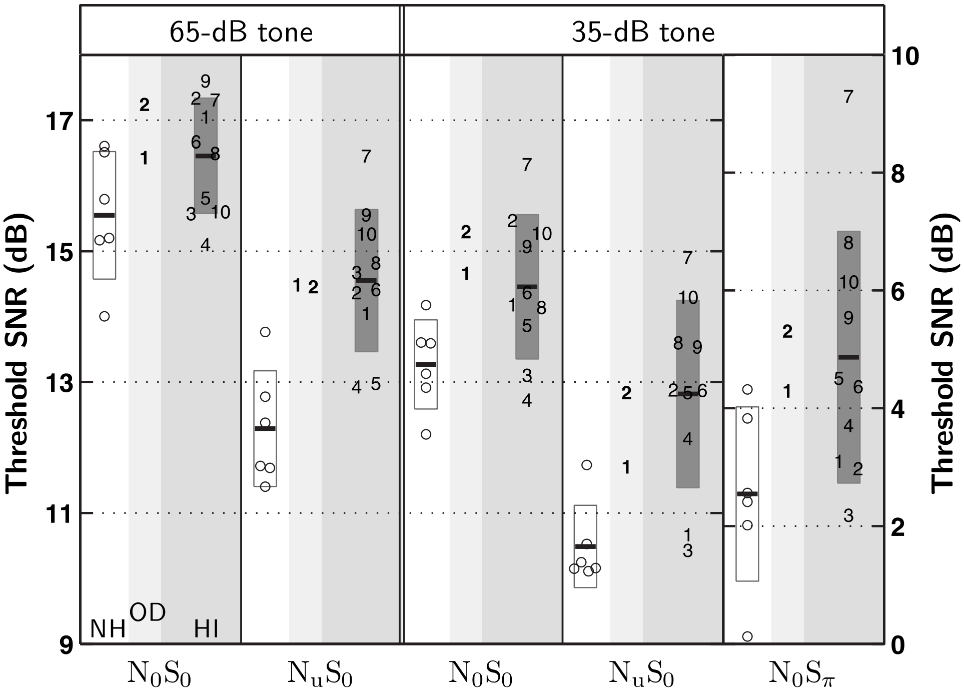


FIG. 3. Binaural masked thresholds, i.e., tone level re masker spectrum level at detection threshold, for the NH listeners (circles), the two listeners with OD (bold numbers), and the HI listeners (plain numbers), obtained in three different masking conditions (N_0S_0, N_uS_0, and N_0S_π) and at two different tone levels (65 and 35 dB SPL). Note the different offset of the ordinate for the N_0S_π condition. Otherwise as Fig. 1.

As can be seen in Fig. 2(a), abnormal filter bandwidths were also found for the two OD listeners. While OD_1 showed significantly elevated bandwidths (compared to the NH group) in both ears, OD_2 showed an increased bandwidth only in the left ear. The difference between the ERB of the left and right ears (divided by the mean ERB of the two ears) is depicted in Fig. 2(b). While this interaural bandwidth asymmetry did not differ significantly between the NH and HI listeners, both OD subjects showed larger differences between the ears than the NH listeners and most of the HI listeners. Decreased frequency selectivity, as found here, has been reported previously in the OD literature (e.g., Narula and Mason, 1988; Saunders and Haggard, 1992). It is also consistent with the finding of reduced distortion-product otoacoustic emission amplitudes (Zhao and Stephens, 2006), if these are taken as an indication of OHC integrity.

The mean ERB of 134 (SD 9) Hz for the NH listeners is larger than the value of 106 Hz predicted by the ERB function given in Glasberg and Moore, 1990. However, that function was designed to predict tuning in the presence of a masker with a constant spectrum level of about 35 dB SPL. It is known that the auditory-filter bandwidth increases with increasing (output) level (e.g., Rosen *et al.*, 1998). Hence, the larger bandwidths found here may be attributed to the higher masker levels applied (the average spectrum level here was 48 dB SPL). In fact, they are in good agreement with the bandwidths reported by Moore *et al.* (1990) who measured at comparable masker levels.

V. BINAURAL MASKED DETECTION

A. Method

The binaural masked thresholds for 750-Hz tones at fixed levels of 65 and 35 dB SPL were measured in band-limited noise (50–1500 Hz). Three different masking conditions were tested: a diotic tone presented in a diotic noise (N_0S_0), a diotic tone presented in an uncorrelated noise (N_uS_0), and a tone with an interaural phase shift of 180° presented in a diotic noise (N_0S_π). The first two conditions were measured using both tone levels whereas the last condition was measured only for the lower tone level. The tones of 500-ms duration were temporally centered in the 700-ms noise maskers. Maskers and tones were gated with raised-cosine ramps of 100-ms and 200-ms durations, respectively.

The same 3I-3AFC method as for the frequency selectivity measurement (including threshold estimation) was used. Also here, the signal level was kept constant while the masker level was varied adaptively. The final standard error of the masked threshold estimate, averaged across all listeners and conditions, was 0.4 dB.

B. Results and discussion

The masked thresholds for the NH, the OD, and the HI listeners are shown in Fig. 3, with SNRs given relative to the masker spectrum level. For all listeners, the thresholds were lower in the dichotic N_uS_0 and N_0S_π conditions than in the corresponding diotic N_0S_0 conditions. These MLDs reflected a release from masking in the dichotic configurations and will be discussed further below. An ANOVA revealed that the masked thresholds were significantly higher for the HI than for the NH listeners $[F(1,14)=14.7,\ p=0.002]$. Furthermore, the masked thresholds differed significantly between the different binaural conditions $[F(2,59)=536.9,\ p<0.0001]$ and also the interaction between listener group and masking condition was significant $[F(2,59)=4.2,\ p=0.02]$. While there was no significant difference between the NH and HI listeners for the diotic N_0S_0 condition [group difference: 1.1 (CI –0.3, 2.4) dB], thresholds for the dichotic conditions differed significantly [N_uS_0 group difference: 2.3 (CI 1.0, 3.6) dB; N_0S_π group difference: 2.3 (CI 0.8, 3.9) dB]. Furthermore, within the group of HI listeners, a significant correlation between the N_uS_0 and N_0S_π thresholds was observed $[r=0.87,\ p=0.001]$. Together, this suggests a deficit with TFS processing at threshold, which impaired N_uS_0 and N_0S_π detections in similar ways.

Significantly larger SNRs were required for the detection of the 65-dB tones than for the 35-dB tones [effect of level on SNR: 1.9 (CI 1.5, 2.4) dB]. This is consistent with the notion of decreasing sharpness of the auditory filters with increasing tone level, if detector efficiency is assumed to be invariant (as found by Rosen *et al.*, 1998). However, as can also be seen in Fig. 3 (mean results), the effect of tone level did not differ significantly across masking condition or listener group. The latter is in agreement with Baker and Rosen (2002), who found a differential effect of tone level on the ERBs of their NH and HI listeners only for levels above 70 dB SPL.

The following MLDs were observed for the NH listeners: $N_0S_0-N_uS_0$ 3.0 (SD 0.7) dB and $N_0S_0-N_0S_\pi$ 10.7 (SD 1.3) dB. Since tone level had no significant effect, here, the $N_0S_0-N_uS_0$ MLD was averaged across the two tone levels. The HI listeners showed significantly smaller $N_0S_0-N_uS_0$ MLDs than the NH listeners [reduced by 1.3 (CI 0.1, 2.4) dB]. However, no significant difference was found for the $N_0S_0-N_0S_\pi$ MLD [reduced by 1.1 (CI –0.2, 2.5) dB]. Hence, the deficits in terms of the MLDs were less significant than the deficits in terms of the masked thresholds. This was due to the fact that the HI listeners exhibited not only significantly increased dichotic thresholds, but also slightly increased diotic thresholds, as previously reported by Staffel *et al.* (1990) and Gabriel *et al.* (1992).

Figure 3 also shows the masked threshold results for the two OD listeners. While subject OD_2 performed clearly more poorly than the NH listeners, subject OD_1 showed performance at the "lower edge" of that for the NH group. However, this applied to both the diotic and the dichotic masking conditions, as reported previously by Saunders and Haggard (1992). Therefore, in terms of their MLDs, no deficits were found for the OD listeners.

VI. LATERALIZATION

A. Method

Lateralization thresholds were measured for 750-Hz tones of 500-ms duration, at fixed levels of 70 and 35 dB SPL. The tones were gated synchronously and were lateralized by introducing a carrier-phase delay to one of the ears, giving rise to an IPD. For NH listeners, interaural carrier delays have been shown to dominate interaural gating delays for frequencies below about 1.5 kHz (see Zurek, 1993). To further weaken potential gating cues to lateralization, long onset/offset ramps of 200 ms each were used. Pilot measurements confirmed that the lateralization was solely based on TFS cues, since no significant difference was found between the lateralization thresholds for tones with a waveform delay and tones with a carrier delay only. At each tone level, in addition to the lateralization threshold in quiet, three conditions with different bandlimited noise interferers (50–1500 Hz) were measured: diotic noise at a low (dioticLo) and a high sound level (dioticHi), and dichotic noise at an intermediate level (dichotic). The noise level in each condition was chosen relative to the individual's masked threshold (N_0S_0 or N_uS_0) to make sure that lateralization performance was not limited by tone detection and to reduce effects of frequency selectivity. The actual noise levels were as follows: dioticHi: 10 dB below masked threshold, for both tone levels; dioticLo: 40 dB below masked threshold for the 70-dB tones and 25 dB below masked threshold for the 35-dB tones; and dichotic: 20 dB below masked threshold for the 70-dB tones and 15 dB below masked threshold for the 35-dB tones.

A two-interval, two-alternative, forced-choice (2I-2AFC) weighted up-down method was used to track 75% correct lateralization. The first interval always contained the zero IPD reference tone while the second interval contained the tone, which was randomly lateralized to the left or right side. Listeners were instructed to indicate the direction of motion. The IPD was tracked logarithmically and the maximum IPD was restricted to 90°, since the extent of lateralization starts to decline for values above 90° (Kunov and Abel, 1981). The background interferer was presented continuously during the whole run. A run was terminated after 14 reversals and the threshold was defined as the geometric mean of all IPD values following the fourth reversal. Listeners were trained in at least two sessions and performed more than 1200 lateralization judgments (constant stimuli) prior to actual data collection. IPD threshold was estimated as the geometric mean over three runs. If the SD over these runs, relative to the mean IPD threshold, exceeded a factor of 0.2 (which corresponds to a constant criterion in logarithmic units), additional runs were taken and the average of all was used. The final relative standard error of the IPD threshold estimate, averaged across all listeners and conditions, was 0.13.

B. Results and discussion

The analysis of the lateralization results was performed on the log-transformed IPDs, as these satisfied the requirements of normal error distributions. This is in line with previous reports in literature on lateralization (e.g., Saberi, 1995; Lacher-Fougère and Demany, 2005). Figure 4 shows the IPD thresholds for the NH, OD, and HI listeners. The HI subjects HI_7 and HI_{10} (not shown) performed much more poorly on lateralization than the remaining HI listeners. Therefore, their IPD thresholds were not included in the group averages and will be discussed separately further below. However, the conclusions presented in the following would remain unchanged if they were taken into account. Two trends can be seen in Fig. 4. First, lateralization performance was better at the higher tone level than at the lower level. Second, the HI listeners showed generally higher IPD thresholds than the NH listeners.

An ANOVA confirmed both the significant difference between NH and HI listeners $[F(1,12)=8.7,\ p=0.01]$, and the effect of tone level $[F(1,94)=71.5,\ p<0.0001]$. The effect of interferer condition was also significant $[F(3,94)=27.8,\ p<0.0001]$, while interactions did not reach significance. The dichotic noise conditions led to the highest IPD

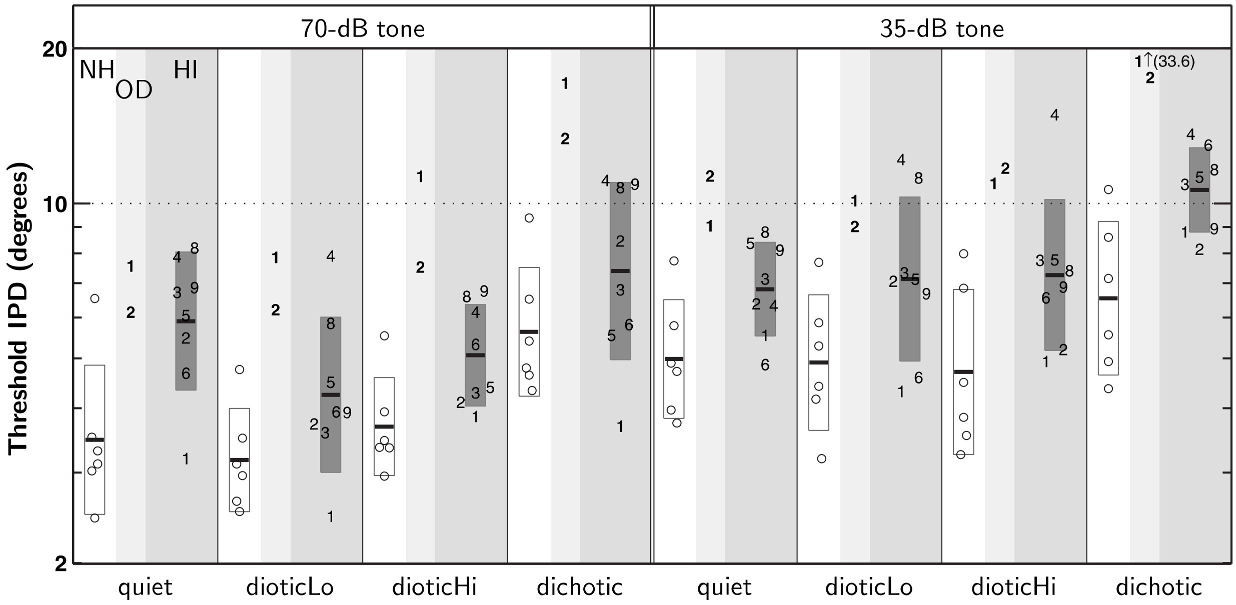


FIG. 4. Lateralization thresholds for the NH listeners (circles), the two listeners with OD (bold numbers), and the HI listeners (plain numbers), at two different tone levels (70 and 35 dB SPL) and for different interferer conditions (see text). Otherwise as Fig. 1.

thresholds, although the noise levels were actually lower than in the dioticHi conditions. This may, at least partly, be attributed to the fact that the dichotic noise gave rise to a diffuse, broad percept, while the diotic noise was lateralized in the midline. Hence, the latter provided an additional ongoing reference cue since the noise was switched on continuously during a run.

Comparing performance in the dioticLo and dioticHi conditions, the HI listeners seemed to cope as well as the NH listeners with the increase in noise level. Generally, noise did not have a greater effect on lateralization performance for the HI listeners than for the NH listeners, irrespective of tone and noise levels (as reflected in a lack of interaction between listener group and condition). Apart from higher thresholds for the HI listeners, the two groups of listeners showed a very similar pattern of results across conditions, with one exception, the quiet condition at the high tone level (leftmost panel in Fig. 4). Here, the lateralization thresholds for the HI listeners were a factor of 1.7 higher than for the NH listeners, while in the other conditions thresholds were, on average, a factor of 1.4 higher. For the dichotic condition at the same tone level (factor of 1.3), one might have expected a larger deficit than in the quiet condition: While in both conditions an ongoing reference cue was absent, a smaller fraction of nerve impulses would have been expected to be phase locked to the tone in the presence of the noise interferer, thus possibly producing more difficulties for the impaired auditory system. This was, however, not the case. Also, the HI listeners' deficit in quiet was actually smaller at the lower tone level (factor of 1.4) than at the higher level.

Hawkins and Wightman (1980) and Smoski and Trahiotis (1986) reported different effects of stimulus level on lateralization performance. For HI listeners with similar audiograms as in the present study, they measured lateralization thresholds in quiet at a low and a high stimulus level, in regions of normal hearing. For narrowband noise stimuli, the HI listeners MM and MD in Hawkins and Wightman, (1980) showed a smaller lateralization deficit at the higher stimulus level than at the lower level. In contrast to this, and consistent with the present results, Smoski and Trahiotis (1986) observed a larger deficit in lateralization at the higher level using pure tones. In the same study, this trend was less clear when using narrowband noise stimuli. Hence, the discrepancy between the studies may be at least partly attributable to the differences in the stimuli.

Smoski and Trahiotis (1986) suggested that the lateralization judgment at high levels could be based on the excitation of a large portion of the basilar membrane rather than only on local excitation, and that a hearing loss might affect the integration of the non-local information. This interpretation is consistent with the present results for lateralization in quiet and in noise. At the tone level of 70 dB SPL, one would expect a substantial spread of excitation, particularly toward places that correspond to higher characteristic frequencies. The NH listeners might have integrated the additional information present at these high-frequency places, whereas the HI listeners might not have been able to benefit from this information, as it fell in the sloping region of their hearing loss. Indeed, if actually included, information from defective neural units (as, e.g., desynchronized information across frequencies) might have had a detrimental effect on lateralization acuity. The role of spread of excitation is reduced at the lower tone level of 35 dB, but also at the higher level of 70 dB in the presence of background noise, as the latter partly masks non-local excitation. This would explain why the deficit observed for the HI listeners (relative to NH) was largest at the high tone level in quiet.

As mentioned above, the HI subjects HI_7 and HI_{10} performed more poorly on lateralization than the remaining HI listeners. Subject HI_7 showed markedly increased lateralization thresholds, independent of interferer condition and tone level. His IPD thresholds ranged from 21° to 27° at the high tone level, and from 32° to 40° at the low tone level, without showing a particular susceptibility to noise interference. For subject HI_{10}, lateralization thresholds could not be determined. Even after a considerable amount of training, her performance remained at chance level (even at the maximum IPD of 90°).[2]

The two OD listeners showed markedly higher lateralization thresholds than the NH listeners, for all interferer conditions and at both tone levels (see Fig. 4). On average, the IPD thresholds for subjects OD_1 and OD_2 were increased relative to those for the NH listeners, by factors of 2.6 and 2.2, respectively. Both showed the most pronounced problems with lateralization in the presence of the dioticHi and dichotic noise interferers. In fact, in these conditions, they performed even more poorly than most of the HI listeners.

VII. FREQUENCY MODULATION DETECTION

A. Method

FMDTs were measured monaurally for carrier frequencies of 125, 750, and 1500 Hz. Prior to gating, the stimulus was a frequency-modulated sinusoid defined by

$$s(t) = a \sin\left[2\pi f_c t + \frac{\Delta f}{f_m}\sin(2\pi f_m t + \varphi)\right], \qquad (1)$$

where f_c represents the carrier frequency, Δf represents the maximum frequency excursion, and f_m represents the FM rate. The FM phase φ was always 1.5π. The phase-locking-based temporal mechanism for FM detection has been found to be operative only at FM rates below 10 Hz, whereas at higher rates, FM detection is thought to be based primarily on a FM-to-AM conversion mechanism (e.g., Moore and Sek, 1996; Lacher-Fougère and Demany, 1998). Here, both mechanisms were tested, by using FM rates of 2 and 16 Hz. The tone levels were 30 dB sensation level (SL; individual hearing thresholds were determined by means of 3I-3AFC detection measurements) and 70 dB SPL. The impact of noise interference was tested by measuring the FMDT for 2-Hz FM tones at 750 Hz in a bandlimited noise (50–1500 Hz), at a level 10 dB below the individual masked threshold. At 1500 Hz, all measurements were undertaken in the presence of a low-level noise background (50–3000 Hz, with a spectrum level of 55 dB below the tone level), in order to mask low-frequency cues due to spread of excitation.

Finally, in order to assess the phase-locking-based mechanism further, similar to the paradigm used by Moore and Sek (1996), FMDTs for 2-Hz FM tones with a superimposed AM were measured at the carrier frequencies of 750 and 1500 Hz. In view of the findings of Grant (1987), who observed a significantly larger deficit in FM detection in HI listeners if the FM tones were randomly rather than sinusoidally amplitude modulated, here, a quasi-sinusoidal AM was used: While the modulation depth was fixed at a peak-to-valley ratio of 6 dB, the instantaneous modulation rate either increased or decreased as a linear function of time. According to Moore and Sek (1996), the peak-to-valley ratio of 6 dB should be large enough to disrupt FM-to-AM conversion cues, but still small enough not to induce substantial level-related pitch shifts. Hence, for the conditions with added AM, the amplitude a in Eq. (1) was time dependent,

$$a(t) \propto 1 + m \sin(2\pi F_a(t) + \vartheta). \qquad (2)$$

Here, m represents the AM depth and $F_a(t)$ is the integral of the instantaneous modulation rate

$$F_a(t) = \int_0^t d\tau\left(f_1 + \frac{f_2 - f_1}{T}\tau\right), \qquad (3)$$

with T representing the tone duration. The initial and final modulation rates f_1 and f_2 were each chosen randomly out of the interval between 1 and 3 Hz, under the constraint $|f_2 - f_1| > 1$ Hz. Also the AM phase ϑ was randomized. Independent of condition, the FM tones had a duration of 750 ms and were gated with 50-ms raised-cosine ramps.

A 3I-3AFC weighted up-down method was used to track 75% correct FM detection. In the conditions without AM, two of the intervals contained unmodulated tones, whereas the target interval contained the FM tone. In the conditions with added AM, all three intervals were independently amplitude modulated and the listeners were instructed to detect the interval containing the FM by listening for its characteristic high-low-high warble. The maximum frequency excursion Δf was tracked logarithmically. A run was terminated after 12 reversals and the threshold was defined as the geometric mean of all Δf values following the fourth reversal. Prior to data collection, a training session was given in which the listeners were trained on all conditions. Initially, both ears were tested on 2-Hz FM detection at 750 Hz in quiet and subsequently the worse ear was chosen for further testing. This was done in order to obtain the largest possible range of FMDTs among the HI listeners, particularly in view of the subsequent comparison with the results of the other tests such as frequency selectivity. Furthermore, it seemed reasonable to assume that the worse ear was limiting the binaural TFS-processing performance, particularly in the lateralization task. The FMDT was estimated as the geometric mean over three runs. If the SD over these runs, relative to the mean FMDT, exceeded a factor of 0.15, additional runs were taken and the average of all was used. The final relative standard error of the FMDT estimate, averaged across all listeners and conditions, was 0.08.

B. Results and discussion

The analysis of the FM detection results was performed on the log-transformed FMDTs, as these satisfied the requirements of normal error distributions. This is in agreement with previous reports in literature on FM detection (e.g., Zurek and Formby, 1981; Buss *et al.*, 2004). For all listeners, FM detection performance did not differ significantly between the tone levels of 30 dB SL and 70 dB SPL (two-tailed t-test: $p = 0.79$). Therefore, only the 30-dB results are considered in the following. Figure 5 shows the FMDTs for the NH, OD, and HI listeners. As can be seen, for all groups, the FMDTs increased with increasing carrier frequency, consistent with previous studies (e.g., Demany and Semal, 1989).

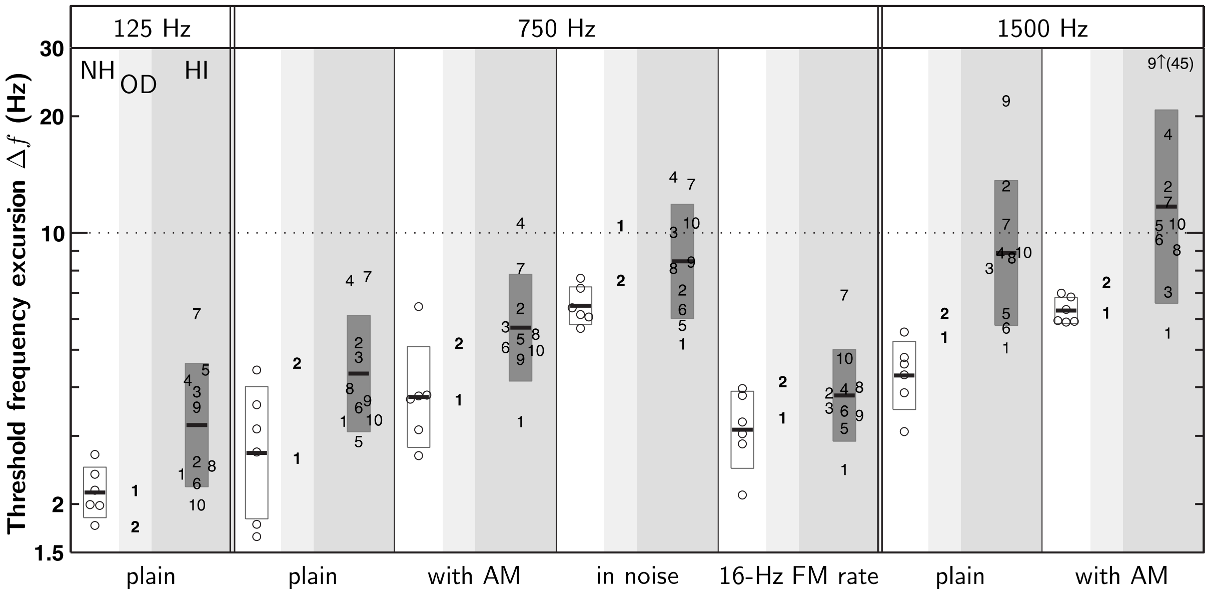


FIG. 5. FMDTs for the NH listeners (circles), the two listeners with OD (bold numbers), and the HI listeners (plain numbers), for three different carrier frequencies (125, 750, and 1500 Hz) and for different measurement conditions (see text; "plain" refers to 2-Hz FM in quiet). All results were obtained at 30 dB SL. Otherwise as Fig. 1.

The HI listeners performed generally more poorly than the NH listeners. On average, their FMDTs were a factor of 1.5 higher than for the NH listeners.

An ANOVA confirmed the statistical significance of the group difference [$F(1,14)=16.9$, $p=0.001$] as well as the effect of tone frequency [$F(1,89)=56.7$, $p<0.0001$]. No significant interaction between listener group and tone frequency was observed [$p=0.19$]. While the log-transformed FMDTs increased linearly as a function of frequency, the Weber fractions $\Delta f/f_c$ decreased from 125 to 750 Hz by a factor of 4 and then remained constant up to 1500 Hz. Zurek and Formby (1981) measured FMDTs in HI listeners and found larger deficits for low-frequency tones than for high-frequency tones, given the same degree of hearing loss (<30 dB HL) at the test frequency. However, the FM detection deficits at 125 Hz observed in the present study were substantially smaller than the ones reported in that study. This might be due to the fact that the HI listeners of Zurek and Formby (1981) showed slightly higher audiometric thresholds at 125 Hz and generally more severe losses below 1000 Hz than the HI listeners of the present study.

FMDTs differed significantly across measurement conditions [2-Hz FM in quiet ("plain"), added AM, noise interference, and higher FM rate; $F(3,89)=24.1$, $p<0.0001$]. The interaction between listener group and measurement condition reached only marginal significance [$F(3,89)=2.5$, $p=0.07$]. However, for the following multiple comparison analysis, the interaction term was kept in the mixed-effects model.

As revealed by the multiple comparisons, the group differences between NH and HI listeners were significant for the 2-Hz FM in quiet and the condition with added AM [group difference in terms of $\log_{10}$(FMDTs) for 2-Hz FM: 0.23 (CI 0.09, 0.37); group difference with added AM: 0.20 (CI 0.04, 0.35)]. For all listeners, the FMDTs with added AM were increased relative to those for the condition with FM only. However, as this increase was similar for the NH and HI listeners, it seems that both groups relied to a comparable extent on FM-to-AM conversion cues, when AM was absent. No significant group difference was found in the condition with the higher FM rate of 16 Hz [group difference: 0.09 (CI −0.08, 0.26)]. Thus, regarding the different FM rates (2 Hz vs 16 Hz), the HI listeners showed a significant deficit on FM detection at the low rate but not at the high rate, where the FM-to-AM conversion is supposed to be the dominant detection mechanism. This can be seen in Fig. 5 (second and fifth panels): While the HI listeners' performance was better for the higher FM rate, the NH listeners' performance was worse. Taken together, this suggests that the observed deficits in the detection of 2-Hz FM were indeed due to problems with phase-locking-based TFS processing.

In the presence of the noise interferer, all listeners performed worse than in quiet. However, the HI listeners did not perform significantly more poorly than the NH listeners in this condition [group difference: 0.11 (CI −0.06, 0.29)]. Hence, the HI listeners did not show an increased susceptibility to noise interference. This is in agreement with the results of Turner (1987), who measured pure-tone frequency difference limens in the presence of low-frequency masking noise for four NH and four HI listeners and found a similar effect of the noise upon performance for the two groups of listeners. Also, Horst (1987) measured frequency discrimination in noise. However, the question of a different impact of noise on the performance of the NH and HI listeners could not be addressed, since he did not measure the frequency difference limen for a given noise level but determined the noise level at which a given fixed frequency difference could just be perceived.

Figure 5 also shows the FMDTs for the two OD listeners. Their FMDTs did not differ substantially from those for the NH listeners. Subject OD_2 performed at the lower edge

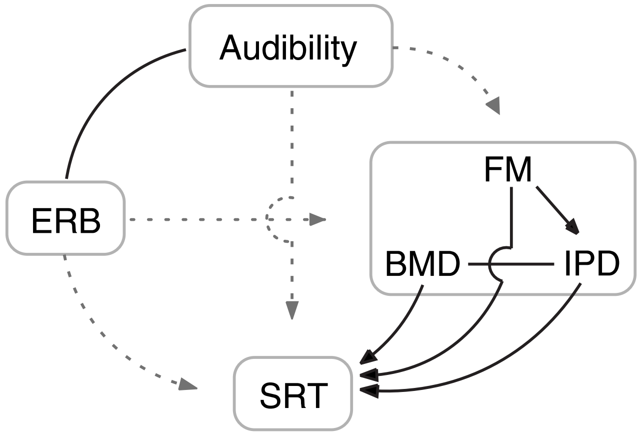


FIG. 6. Relations between the results for the different auditory tests within the group of HI listeners: pure-tone hearing thresholds (audibility), frequency selectivity (ERB), monaural frequency modulation detection (FM), binaural masked detection (BMD), tone lateralization (IPD), and speech reception (SRT). Solid lines indicate significant correlations whereas dotted lines indicate correlations that were not significant. The direction of the arrows is solely based on the assumed sequence of processing in the auditory pathway. Therefore, arrowheads were omitted where the order is uncertain or where the processing might take place in parallel rather than in sequence.

of the NH listeners except for the 125-Hz carrier, where her performance was good. For subject OD_1, a deficit was observed for the 750-Hz carrier with interfering noise. Otherwise her performance was essentially normal.

VIII. COMPARISON OF RESULTS ACROSS TESTS

A. Hearing-impaired listeners

Pearson correlations and two-tailed p values were examined to study the relations between the results of the different auditory tests within the group of HI listeners. The findings are schematized in Fig. 6.

1. Correlations with absolute hearing thresholds

Frequency selectivity in terms of the ERB at 750 Hz was significantly correlated with the individual hearing threshold at this frequency $[r=0.77,\ p=0.009]$. Here, the hearing threshold was estimated by means of a 3I-3AFC method with a 1-dB stepsize. When the standard audiometric threshold (with a 5-dB stepsize) was considered instead, the correlation was smaller $[r=0.53]$, but increased when thresholds were averaged in terms of the pure-tone average (PTA) threshold at 0.5, 1, 2, and 4 kHz $[r=0.8]$. The finding of a correlation between frequency selectivity and hearing threshold is consistent with previous reports in literature (e.g., Tyler *et al.*, 1983; Moore, 1996), although less distinct correlations have been observed for hearing losses below 30 dB HL (see Baker and Rosen, 2002).

No significant correlations were observed between individual hearing thresholds and performance in the three tests of TFS processing (binaural masked detection, lateralization, and FM detection). Tones with equal sound pressure levels were used for all listeners in the masked detection and lateralization tasks. Hence, the deficits in performance that were observed at the low tone level of 35 dB SPL could have been due to the slightly differing sensation levels (ranging from 32 to 38 dB SL for the NH group and from 23 to 34 dB SL for the HI group). However, the absence of correlations between hearing thresholds, and thereby sensation levels, and masked detection/lateralization performance makes this unlikely. With regard to FM detection, subject HI_9, who showed markedly worse performance at 1.5 kHz, also had the highest hearing thresholds at this frequency. Nevertheless, the correlation between the hearing thresholds and FMDTs at 1.5 kHz was not significant when considering all HI listeners $[r=0.39,\ p=0.27]$. Finally, the hearing thresholds were not significantly correlated with the results for speech reception, regardless of whether the hearing thresholds at single frequencies or averages across frequencies were considered.

The absence of correlations with the hearing thresholds can, to some extent, be attributed to the homogeneity of the HI group in terms of their audiograms. Also, given the limited number of listeners, only rather strong correlations would be expected to be significant. Hence, here and in the following, the absence of a significant correlation does not necessarily imply the absence of a relationship.

2. Correlations between the various tests of TFS processing and frequency selectivity

The deficits observed for the HI listeners with binaural masked detection, lateralization, and FM detection provide strong evidence for deficits with phase-locking-based TFS processing. However, no significant correlations were observed between frequency selectivity and these tests of TFS processing.[3] This can be illustrated by means of individual results among the HI listeners: Subject HI_1 showed poor frequency selectivity but good TFS-processing skills, whereas subject HI_7 performed well on the former but poorly on the latter. Subject HI_{10} showed poor performance in both domains. Hence, it seems that the deficits found in TFS processing cannot be attributed solely to a deficit in frequency selectivity, but must be, at least partly, due to another impairment factor. This is further supported by the finding of TFS-processing deficits in quiet, which cannot be explained in terms of frequency selectivity.

Significant correlations were found among the tests of TFS processing. When correlations between the tests were observed for multiple test conditions, such as for the different interferer conditions in the lateralization task, an overall correlation is given in the following, instead of reporting the correlations for each individual condition. The overall correlation is based on the listeners' average performance on that test. This average performance was measured in terms of the estimated random effect, which summarizes individual performance across multiple conditions. Here, it represents the performance deviation of an individual HI listener from the HI group mean. Since the random effect accounts for multiple measurement conditions simultaneously, the corresponding correlation results are more robust and more conservative in terms of significance. Using this statistic, a significant correlation was observed between lateralization performance and the binaural masked thresholds in the N_0S_π condition $[r=0.80,\ p=0.01]$, as has been observed previously (Hall *et al.*, 1984; Kinkel *et al.*, 1988; Koehnke *et al.*, 1995). While the correlation between lateralization performance and the N_uS_0 thresholds was rather marginal [$p \sim 0.08$], no such correlation was observed for the N_0S_0 thresholds [$p > 0.2$].[4] The above correlation between lateralization performance and N_0S_π detection thresholds remained significant when controlling for individual hearing thresholds by means of partial correlation [$r=0.83$, $p=0.01$].

Performance on monaural FM detection and binaural masked detection was not correlated significantly.[5] However, the monaural FMDTs at 750 Hz were significantly correlated with lateralization performance [$r=0.79$, $p=0.01$]. Considering the different FM conditions separately, the correlations were strongest for the conditions with noise interference and with added AM. The correlation remained significant when controlling for individual hearing thresholds by means of partial correlation [$r=0.79$, $p=0.02$]. The fact that binaural and monaural (suprathreshold) TFS processing were correlated for the HI listeners suggests that the binaural deficit might be mainly attributable to a monaural impairment factor.

3. Correlations with speech reception

As depicted in Fig. 7, two of the full-spectrum speech conditions, LATSSN and TWOTALK, showed significant correlations with the measures of TFS processing while no significant correlations were observed for the other speech conditions, including filtered speech.[6] Performance in the dichotic masked detection tasks (conditions N_0S_π and N_uS_0, in terms of the estimated random effects) was correlated with the SRTs in the LATSSN condition [$r=0.85$, $p=0.002$]; see Fig. 7(a). The correlation was also significant when the masking release instead of the SRT was considered [$r=0.80$, $p=0.005$]. For the sake of brevity in the following, a correlation with the masking release will only be given if it was stronger than the correlation with the corresponding SRT itself. The SRTs in the LATSSN condition were also significantly correlated with lateralization performance, but only for the dioticHi condition at the high tone level [$r=0.80$, $p=0.02$]; see Fig. 7(b).[7] The pattern of correlations between the LATSSN SRTs and the masked thresholds as well as the lateralization thresholds remained unchanged when partialing out the individual hearing thresholds [masked detection: $r=0.82$, $p=0.007$; lateralization: $r=0.76$, $p<0.05$]. For the TWOTALK condition, significant correlations were found with both the dichotic masked thresholds (N_0S_π and N_uS_0) and the lateralization thresholds in the dioticHi condition [masked detection: $r=0.68$, $p=0.03$; lateralization: $r=0.84$, $p=0.009$], as can be seen in Figs. 7(c) and 7(d), respectively. While the correlation with the masked thresholds was marginal when controlling for the individual hearing thresholds, the correlation with the lateralization thresholds remained significant [masked detection: $r=0.60$, $p=0.09$; lateralization: $r=0.81$, $p=0.03$].

No significant correlations between performance on speech reception and FMDTs at 125 and 750 Hz were found. However, at 1.5 kHz, the FMDTs with added AM were significantly correlated with the SRT in the TWOTALK condition [$r=0.75$, $p=0.013$]. Here, the correlation was stronger for the corresponding masking release [$r=-0.77$, $p=0.009$], as depicted in Fig. 7(e). When controlling for the individual hearing thresholds at 1.5 kHz, the correlation with the SRT was marginal, while the correlation with the masking release remained significant [SRT: $r=0.61$, $p=0.08$; masking release: $r=-0.67$, $p<0.05$]. Generally, the observed correlations were only slightly affected when the effect of absolute hearing thresholds was partialed out. To some degree, this can be attributed to the homogeneity of the HI group in terms of their hearing thresholds.

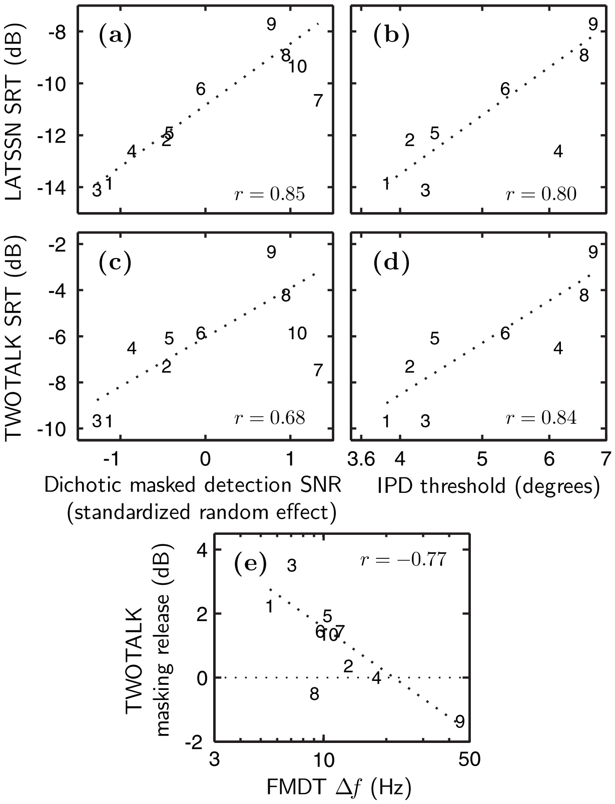


FIG. 7. Correlations between performance on speech reception and TFS processing within the group of HI listeners. The dotted regression lines were obtained by means of least trimmed squares robust regression. (a) Correlation between the LATSSN SRTs and performance for dichotic masked detection (N_uS_0 and N_0S_π conditions). The latter is given in terms of the standardized random effects, which measure the individual deviations from the HI group mean. Better/worse than average performance, i.e., a smaller/larger threshold SNR, results in a negative/positive random effect. The interval from -1 to 1 covers 68% of the HI "population." (b) Correlation between the LATSSN SRTs and the IPD thresholds in the dioticHi condition for the 70-dB tones. [(c) and (d)] Same as (a) and (b) but for TWOTALK SRTs. (e) Correlation between the TWOTALK masking release (re SSN) and the FMDTs for 1.5-kHz tones with added AM.

The finding of a correlation between the SRTs for the dichotic LATSSN masker and binaural low-frequency TFS processing seems reasonable in view of the results reported by Schubert and Schultz (1962) and Levitt and Rabiner (1967). They found that the release from masking for dichotic speech in noise ($N_0S_{0.5\text{ms}}$ or N_0S_π) was primarily determined by interaural time or phase disparity at low frequencies. Besides, in the present study, binaural masked detection and dichotic speech reception depended in the same way on binaural integration: While they could be accomplished monaurally, use of the binaural information would give rise to better performance. While the LATSSN condition assessed the ability to take advantage of an interaural timing mismatch between target speech and noise interferer, performance in the TWOTALK background depended particularly on the ability to separate the target talker and the two interfering talkers. Hence, the correlations found between the SRTs for TWOTALK background and the measures of TFS processing support the hypothesis of Zeng *et al.* (2005) that TFS cues might be utilized in talker separation in order to improve performance in listening situations with competing talkers. In this respect, the correlation between speech reception in the TWOTALK background (in terms of SRT and masking release) and FM detection performance at 1.5 kHz, observed here, may indicate a potential contribution of the second formant region (cf. Peterson and Barney, 1952) to talker identification and separation.

TFS processing was not correlated with SRTs (or masking releases) in the fluctuating backgrounds, SAM and RAM, neither for full-spectrum nor filtered speech. Hence, in contrast to Lorenzi *et al.*, (2006), no evidence was found for a relation between TFS processing and dip listening. This discrepancy might have been due to the fact that the HI listeners in Lorenzi *et al.*, (2006) had "flat" moderate hearing losses ($\sim$50 dB HL), whereas the HI listeners in the present study had "normal" hearing thresholds up to 1 kHz. Furthermore, Lorenzi *et al.* (2006) tested TFS processing with processed speech stimuli, which exhibited more complex TFS patterns than the tone stimuli used in the present study (with the exception of the uncorrelated noise maskers in the N_uS_0 masked detection task).

A correlation between frequency selectivity and speech reception, as previously reported in literature (e.g., Dreschler and Plomp, 1985; Horst, 1987), was not observed here. However, these studies often included estimates of frequency selectivity at frequencies above 1000 Hz, while, in the present study, frequency selectivity was estimated only at 750 Hz. This may explain the absence of a correlation in the case of the full-spectrum speech, but not for the lowpass-filtered speech. Another possible explanation, which might also account for the results with filtered speech, is that several impairment factors contributed to the observed speech reception deficits in complementary ways. Indeed, when the low-frequency slopes of the estimated filters and the monaural FMDTs at 1.5 kHz (with added AM) were considered as joint predictors in a multiple regression analysis, their combined effect on the monaural SRTs in the SSN and SAM conditions was significant [combined effect of filter slope and FMDT for SSN: $F(2,7)=9.6$, $p=0.01$; for SAM: $F(2,7)=8.5$, $p=0.01$]. The combined effect was less significant when the ERB instead of the filter slope was considered (for SSN: $p=0.04$; for SAM: $p=0.05$). However, regression results that rely on such conjunctions of variables, rather than on strong primary correlations, should be viewed with caution, particularly in view of the small number of subjects.

4. Possible relations to aging

One concern is that the NH listeners in the present study were, on average, younger than the HI listeners (median age 28 and 63 years, respectively). This raises the question of possible age effects, as previous studies have suggested a relation between aging and deficits in TFS processing as well as speech reception (e.g., Pichora-Fuller and Schneider, 1992; Strouse *et al.*, 1998; Schneider *et al.*, 2002; Ross *et al.*, 2007). Indeed, subject HI$_1$, who was the youngest of the HI listeners, performed better than the other HI listeners on the three tests of TFS processing, particularly lateralization and FM detection. However, apart from her age, HI$_1$ also differed in terms of etiology, as her hearing loss was due to hypoxia at birth. For the remaining HI listeners (53–74 years), dichotic masked detection was significantly correlated with age, while results for the other TFS tests were not (dichotic masked detection: $r=0.81$, $p=0.01$; lateralization: $r=0.36$, $p=0.37$; FM detection: $r=0.13$, $p=0.75$). Hence, it cannot be excluded that part of the TFS deficits observed for the HI listeners could be related to aging. Ross *et al.* (2007) recorded cortical auditory-evoked responses to tones with dynamic changes in IPD. They found that the highest carrier frequency, at which responses to changes in IPD could be detected, declined with age. This indicates that aging might induce or potentiate a degradation in the processing of TFS at a peripheral or central auditory level, which is not reflected in the pure-tone hearing thresholds.

B. Listeners with obscure dysfunction

The two OD listeners showed deficits in frequency selectivity and binaural masked detection, which were comparable to those of the HI listeners. In the lateralization task they performed even more poorly than most of the HI listeners, showing substantial deficits, particularly with lateralization in background noise. However, in contrast to the HI listeners, who showed similar deficits on binaural lateralization and monaural FM detection, the OD listeners did not show as clear deficits in the FM detection task as in the lateralization task. Since FM detection was assessed monaurally, one might conjecture that it was the non-tested ear that was actually limiting the lateralization performance. However, this can be excluded, as both ears were screened initially on FM detection and the worse ear was chosen for further testing. A possible reason for the poor binaural TFS performance of the OD listeners could be the large bandwidth differences between their ears. Colburn and Häusler (1980) suggested that the output of differing filters, given a diotic wideband input signal, would be partly uncorrelated at the two ears, resulting in lateralization blur. However, this explanation does not account for the observed poor performance in quiet and in dichotic (uncorrelated) noise. Hence, it seems that the TFS processing was affected at the stage of binaural integration rather than at a preceding monaural stage. Alternatively, even if the binaural TFS information was accurately integrated, it might not have been accessible at following stages of auditory processing.

The OD listeners showed rather small deficits in the reception of full-spectrum speech, but clear deficits in the reception of lowpass-filtered speech. These deficits might, at least partly, be attributable to the deficits in frequency selectivity and binaural TFS processing, which were observed to a similar extent for both OD listeners. However, additional personality-related factors, such as an individual's underestimation of their own hearing ability (lack of "auditory confidence"), may be involved in the phenomenon of obscure (auditory) dysfunction. Considering the heterogeneity of the clinical group of OD patients (e.g., Saunders and Haggard, 1989; Zhao and Stephens, 2000) and the fact that the diagnosis of OD is solely based on a self-rated disability, the necessity for such factors is almost self-evident.

IX. POSSIBLE UNDERLYING IMPAIRMENT MECHANISMS

Figure 6 illustrates that TFS processing was related neither to audibility nor to frequency selectivity, although deficits were found in all of the tests. One may speculate about possible impairment sites and mechanisms underlying these deficits. Damage to or loss of OHCs has been shown to result in a loss of sensitivity and frequency selectivity (e.g., Evans and Harrison, 1976; Liberman and Dodds, 1984), while damage to or loss of inner hair cells (IHCs) does not seem to have any substantial effect on sensitivity or tuning of the remaining intact IHCs (e.g., Wang *et al.*, 1997). Hence, OHC damage might have been responsible for the deficits in frequency selectivity and their relation to absolute threshold observed here (cf. Moore *et al.*, 1999).

Several factors might have contributed to the deficits in TFS processing. A loss of OHCs could have resulted in a reduced precision of phase locking (Woolf *et al.*, 1981). However, this is controversial, as other studies did not find evidence for such phase-locking anomalies (e.g., Miller *et al.*, 1997). Apart from this, Woolf *et al.* (1981) found the reduced phase locking to be related to elevated absolute thresholds, which was not observed for the TFS deficits in the present study. Also, a loss of OHCs might have altered the spatiotemporal response pattern of the basilar membrane. This could have affected TFS processing if TFS information was extracted by cross-correlation of the outputs of different places along the basilar membrane (e.g., Deng and Geisler, 1987; Shamma, 2001; Carney *et al.*, 2002). Since the present study assessed OHC integrity in terms of frequency selectivity only at a single frequency, this option cannot be ruled out here.

Alternatively, through partial section of the auditory nerve, it has been shown that a loss of auditory-nerve fibers of up to 90% does not necessarily result in elevated pure-tone thresholds (e.g., Schuknecht and Woellner, 1953). Hence, the observed TFS deficits in regions of normal hearing might be attributable to damage to or loss of auditory-nerve fibers or the innervated IHCs. A related possibility concerns the (monaural) enhancement of phase-locking synchrony to low-frequency tones that has been observed in the cochlear nucleus (e.g., Joris *et al.*, 1994) and might be reduced in impaired hearing. The alternative possibility, however, that a specific binaural processing stage, such as interaural coincidence detection, was affected in the HI listeners seems implausible given the clear correlation between the monaural and binaural TFS deficits found in these listeners.

X. SUMMARY AND CONCLUSIONS

In addition to deficits in speech reception, deficits in frequency selectivity and in phase-locking-based TFS processing were observed for HI listeners, despite testing in regions of normal hearing. The observed TFS deficits were not related to reduced frequency selectivity. Monaural and binaural TFS deficits, however, were found to be related, suggesting that the binaural deficits might have been attributable to a monaural impairment factor. Background noise did not have a larger effect on TFS processing for the HI listeners than for the NH listeners: Although the acuity of TFS processing was decreased for the HI listeners, it seemed to be as robust to noise interference as for the NH listeners. SRTs in a two-talker background and in lateralized noise, but not in amplitude-modulated noise, were correlated with TFS-processing performance, suggesting that TFS information might be utilized in talker separation and spatial segregation.

The OD listeners showed deficits in frequency selectivity and in binaural, but not monaural, TFS processing. Compared with the NH listeners, their SRTs were particularly elevated for lowpass-filtered speech.

These findings on auditory deficits, as well as preserved auditory abilities, may serve as constraints for future models of the impaired auditory system. Furthermore, they may help in defining an auditory profile for listeners with impaired hearing.

ACKNOWLEDGMENTS

We wish to thank Graham Naylor, Christian Lorenzi, Steven Greenberg, Eric R. Thompson, and our colleagues at the Centre for Applied Hearing Research for valuable comments and stimulating discussions. We also thank Brian C. J. Moore and three anonymous reviewers for their constructive criticism on an earlier draft of this paper. We are grateful to all the listeners for their participation in many hours of testing. This work was supported by the Danish Research Foundation, the Danish Graduate School SNAK "Sense organs, neural networks, behavior, and communication," and the Oticon Foundation.

[1] In addition to the roex(p,r) filter model, the more complex variants roex(p,w,t) and roex(p,w,t,p), as given in Oxenham and Shera (2003), were fitted to the threshold data. However, the gain in terms of goodness of fit was negligible, with an average change in the rms error by a factor of 0.96 and a maximum change by a factor of 0.7 in a single case. This is in contrast to Oxenham and Shera (2003), who found a considerably larger reduction in the rms error. The discrepancy may be due to the use of a short signal duration (10 ms) and low sensation levels (10 to 35 dB SL) in their study. Since the results obtained here with the more complex filter models were very similar to those obtained with the simple roex(p,r) model, the discussion was limited to the latter.

[2] While subject HI$_7$ showed consistently poor performance on all TFS-processing tests (poorest performance of all listeners on binaural masked detection and FM detection), subject HI$_{10}$, who was not able to lateralize at all, showed relatively poor performance on masked detection, but average performance in the FM detection task. Although it was ensured that HI$_{10}$ had understood the lateralization task, it cannot be excluded that her problem was, at least partly, due to the nature of the 2I-2AFC task, rather than a problem with lateralization per se.

[3] At first, it may seem surprising that the diotic masked thresholds (N_0S_0) and frequency selectivity were not correlated. However, in addition to the

filter bandwidth, the masked thresholds are determined by the detector efficiency, i.e., the SNR at the output of the auditory filter required for detection.

[4] The fact that no significant correlation was observed between the IPD thresholds and the N_0S_0 or N_uS_0 masked thresholds is not surprising, as the levels of the diotic and dichotic noise interferers in the lateralization task had been chosen according to the individual N_0S_0 and N_uS_0 detection thresholds, in order to make sure that lateralization performance was not limited by tone detection.

[5] A reason for this could be that FM detection constituted a suprathreshold measure of TFS processing, while masked detection assessed the latter at threshold. Apart from this, since the tone detection could have been accomplished monaurally, it seems reasonable to assume that the binaural detection performance was not solely determined by the "worse" ear, which was tested on FM detection.

[6] In the filtered-speech conditions, listeners HI_6 and HI_9 performed markedly more poorly than all other listeners (Fig. 1). Subject HI_9 showed the largest deficits in speech reception among the HI listeners. However, his poor performance on FM detection at 1.5 kHz, which might have been a sign of substantial deficits in the processing of high-frequency information, cannot account for the deficits in the reception of lowpass-filtered speech. Similarly, subject HI_6's problems with lowpass-filtered speech were not reflected in his performance on the auditory tests of frequency selectivity or TFS processing. The reason for this remains unclear.

[7] Given that a dichotic noise interferer was used in the LATSSN speech condition, it might seem counterintuitive that a correlation was found in the case of the dioticHi lateralization condition but not the dichotic condition. However, the dichotic noise interferer (as compared to a diotic one) exerted rather opposite effects on speech reception and lateralization: While it gave rise to a release from masking in the speech task, it represented an additional challenge in the lateralization task. Furthermore, the level of the dioticHi noise in the lateralization task was comparable to the level of the noise interferer in the speech task (if the level is considered relative to the corresponding masked threshold for tone and speech, respectively).

Abbas, P. J. (**1981**). "Auditory-nerve fiber responses to tones in a noise masker," Hear. Res. **5**, 69–80.

Baker, R. J., and Rosen, S. (**2002**). "Auditory filter nonlinearity in mild/moderate hearing impairment," J. Acoust. Soc. Am. **111**, 1330–1339.

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (**2002**). "Auditory phase opponency: A temporal model for masked detection at low frequencies," Acta. Acust. Acust. **88**, 334–347.

Colburn, H., and Häusler, R. (**1980**). "Note on the modeling of binaural interaction in impaired auditory systems," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft).

Colburn, H. S. (**1996**). "Computational models of binaural processing," in *Auditory Computation*, edited by H. Hawkins, T. McMullen, A. Popper, and R. Fay (Springer-Verlag, New York).

Costalupes, J. A. (**1985**). "Representation of tones in noise in the responses of auditory nerve fibers in cats. I. Comparison with detection thresholds," J. Neurosci. **5**, 3261–3269.

Demany, L., and Semal, C. (**1989**). "Detection thresholds for sinusoidal frequency modulation," J. Acoust. Soc. Am. **85**, 1295–1301.

Deng, L., and Geisler, C. D. (**1987**). "A composite auditory model for processing speech sounds," J. Acoust. Soc. Am. **82**, 2001–2012.

Dreschler, W. A., and Plomp, R. (**1985**). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," J. Acoust. Soc. Am. **78**, 1261–1270.

Durlach, N. I., Thompson, C. L., and Colburn, H. S. (**1981**). "Binaural interaction in impaired listeners. A review of past research," Audiology **20**, 181–211.

Evans, E. F., and Harrison, R. V. (**1976**). "Correlation between cochlear outer hair cell damage and deterioration of cochlear nerve tuning properties in the guinea-pig," J. Physiol. **256**, 43P–44P.

Festen, J. M., and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. **73**, 652–662.

Freyman, R. L., and Nelson, D. A. (**1991**). "Frequency discrimination as a function of signal frequency and level in normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **34**, 1371–1386.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.

Gabriel, K. J., Koehnke, J., and Colburn, H. S. (**1992**). "Frequency dependence of binaural performance in listeners with impaired binaural hearing," J. Acoust. Soc. Am. **91**, 336–347.

Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438–2444.

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). "Effect of masker modulation depth on speech masking release," Hear. Res. **239**, 60–68.

Grant, K. W. (**1987**). "Frequency modulation detection by normally hearing and profoundly hearing-impaired listeners," J. Speech Hear. Res. **30**, 558–563.

Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1984**). "Factors influencing the masking level difference in cochlear hearing-impaired and normal-hearing listeners," J. Speech Hear. Res. **27**, 145–154.

Häusler, R., Colburn, S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests," Acta Otolaryngol. Suppl. **400**, 1–62.

Hawkins, D. B., and Wightman, F. L. (**1980**). "Interaural time discrimination ability of listeners with sensorineural hearing loss," Audiology **19**, 495–507.

Hinchcliffe, R. (**1992**). "King-Kopetzky syndrome: An auditory stress disorder?," J. Audiol. Med. **1**, 89–98.

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Horst, J. W. (**1987**). "Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects," J. Acoust. Soc. Am. **82**, 874–885.

Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (**2002**). "Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing," J. Acoust. Soc. Am. **111**, 409–416.

ISO 389-8 (**2004**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones," International Organization for Standardization (Geneva).

Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. (**1994**). "Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency," J. Neurophysiol. **71**, 1022–1036.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys. **49**, 227–229.

King, K., and Stephens, D. (**1992**). "Auditory and psychological factors in auditory disability with normal hearing," Scandinavian Audiology, **21**(2), 109–114.

Kinkel, M., Holube, I., and Kollmeier, B. (**1988**). "Zusammenhang verschiedener Parameter binauralen Hörens bei Schwerhörigen (Relation between parameters of binaural hearing in hearing impaired subjects)," *Fortschritte der Akustik-DAGA 1988* (DPG Kongreß-GmbH, Bad Honnef), pp. 629–632.

Koehnke, J., Culotta, C. P., Hawley, M. L., and Colburn, H. S. (**1995**). "Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing," Ear Hear. **16**, 331–353.

Kunov, H., and Abel, S. M. (**1981**). "Effects of rise/decay time on the lateralization of interaurally delayed 1-kHz tones," J. Acoust. Soc. Am. **69**, 769–773.

Lacher-Fougère, S., and Demany, L. (**1998**). "Modulation detection by normal and hearing-impaired listeners," Audiology **37**, 109–121.

Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.

Laird, N. M., and Ware, J. H. (**1982**). "Random-effects models for longitudinal data," Biometrics **38**, 963–974.

Levitt, H., and Rabiner, L. R. (**1967**). "Binaural release from masking for speech and gain in intelligibility," J. Acoust. Soc. Am. **42**, 601–608.

Liberman, M. C., and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Lorenzi, C., and Moore, B. C. J. (**2008**). "Role of temporal envelope and fine structure cues in speech perception: A review," in *Auditory Signal Processing in Hearing-Impaired Listeners*, *First International Symposium on Auditory and Audiological Research (ISAAR 2007)*, edited by T. Dau, J. M. Buchholz, J. M. Harte, and T. U. Christiansen (Centertryk, Denmark).

Middelweerd, M. J., Festen, J. M., and Plomp, R. (**1990**). "Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram," Audiology **29**, 1–7.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel "eh" in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602–3616.

Moore, B. C. J. (**1996**). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–161.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, San Diego, CA), pp. 197–204.

Moore, B. C. J. (**2008**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399–406.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**, 224–240.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1990**). "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. **88**, 132–140.

Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.

Narula, A. A., and Mason, S. M. (**1988**). "Selective dysacusis—A preliminary report," J. R. Soc. Med. **81**, 338–340.

Nie, K., Stickney, G., and Zeng, F.-G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2001**). "Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region," J. Acoust. Soc. Am. **109**, 1197–1212.

Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).

Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory-filter asymmetry," J. Acoust. Soc. Am. **67**, 229–245.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.

Peterson, G., and Barney, H. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Pichora-Fuller, M. K., and Schneider, B. A. (**1992**). "The effect of interaural delay of the masker on masking-level differences in young and old adults," J. Acoust. Soc. Am. **91**, 2129–2135.

Pinheiro, J., and Bates, D. (**2000**). *Mixed-Effects Models in S and S-PLUS* (Springer-Verlag, New York).

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rhode, W. S., Geisler, C. D., and Kennedy, D. T. (**1978**). "Auditory nerve fiber response to wide-band noise and tone combinations," J. Neurophysiol. **41**, 692–704.

Rosen, S. (**1987**). "Phase and the hearing-impaired," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Springer-Verlag, New York).

Rosen, S., Baker, R. J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am. **103**, 2539–2550.

Ross, B., Fujioka, T., Tremblay, K. L., and Picton, T. W. (**2007**). "Aging in binaural hearing begins in mid-life: Evidence from cortical auditory-evoked responses to changes in interaural phase," J. Neurosci. **27**, 11172–11178.

Ruggero, M. A. (**1992**). "Physiology and coding of sound in the auditory nerve," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 34–93.

Saberi, K. (**1995**). "Some considerations on the use of adaptive methods for estimating interaural-delay thresholds," J. Acoust. Soc. Am. **98**, 1803–1806.

Santurette, S., and Dau, T. (**2007**). "Binaural pitch perception in normal-hearing and hearing-impaired listeners," Hear. Res. **223**, 29–47.

Saunders, G. H., and Haggard, M. P. (**1989**). "The clinical assessment of obscure auditory dysfunction—1. Auditory and psychological factors," Ear Hear. **10**, 200–208.

Saunders, G. H., and Haggard, M. P. (**1992**). "The clinical assessment of "obscure auditory dysfunction" (OAD) 2. Case control analysis of determining factors," Ear Hear. **13**, 241–254.

Schneider, B. A., Daneman, M., and Pichora-Fuller, M. K. (**2002**). "Listening in aging adults: From discourse comprehension to psychoacoustics," Can. J. Exp. Psychol. **56**, 139–152.

Schubert, E. D., and Schultz, M. C. (**1962**). "Some aspects of binaural signal selection," J. Acoust. Soc. Am. **34**, 844–849.

Schuknecht, H. F., and Woellner, R. C. (**1953**). "Hearing losses following partial section of the cochlear nerve," Laryngoscope **63**, 441–465.

Shamma, S. (**2001**). "On the role of space and time in auditory processing," Trends Cogn. Sci. **5**, 340–348.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Smoski, W. J., and Trahiotis, C. (**1986**). "Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **79**, 1541–1547.

Staffel, J. G., Hall, J. W., Grose, J. H., and Pillsbury, H. C. (**1990**). "NoSo and NoSπ detection as a function of masker bandwidth in normal-hearing and cochlear-impaired listeners," J. Acoust. Soc. Am. **87**, 1720–1727.

Stern, M., and Trahiotis, C. (**1995**). "Models of binaural interaction," in *Hearing*, edited by B. C. J Moore (Academic, San Diego, CA).

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.

Turner, C. W. (**1987**). "Effects of noise and hearing loss upon frequency discrimination," Audiology **26**, 133–140.

Turner, C. W., and Nelson, D. A. (**1982**). "Frequency discrimination in regions of normal and impaired sensitivity," J. Speech Hear. Res. **25**, 34–41.

Tyler, R. S., Wood, E. J., and Fernandes, M. (**1983**). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **74**, 1190–1199.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**2001**). "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 529–542.

Wagener, K., Josvassen, J. L., and Ardenkjaer, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.

Wang, J., Powers, N. L., Hofstetter, P., Trautwein, P., Ding, D., and Salvi, R. (**1997**). "Effects of selective inner hair cell loss on auditory nerve fiber threshold, tuning and spontaneous and driven discharge rate," Hear. Res. **107**, 67–82.

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.

Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). "Neural phase-locking properties in the absence of cochlear outer hair-cells," Hear. Res. **4**, 335–346.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Rio, E. D., Kong, Y.-Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.

Zhao, F., and Stephens, D. (**2000**). "Subcategories of patients with King-Kopetzky syndrome," Br. J. Audiol. **34**, 241–256.

Zhao, F., and Stephens, D. (**2006**). "Distortion product otoacoustic emissions in patients with King-Kopetzky syndrome," Int. J. Audiol. **45**, 34–39.

Zurek, P. M. (**1993**). "A note on onset effects in binaural hearing," J. Acoust. Soc. Am. **93**, 1200–1201.

Zurek, P. M., and Formby, C. (**1981**). "Frequency-discrimination ability of hearing-impaired listeners," J. Speech Hear. Res. **24**, 108–112.

Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs (The elementary foundations for determining the information capacity of the auditory system)," Acustica **6**, 365–381.

Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech

Jean C. Krause[a] and Louis D. Braida
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 4 March 2008; revised 10 February 2009; accepted 14 February 2009)

In adverse listening conditions, talkers can increase their intelligibility by speaking clearly [Picheny, M.A., *et al.* (1985). J. Speech Hear. Res. **28**, 96–103; Payton, K. L., *et al.* (1994). J. Acoust. Soc. Am. **95**, 1581–1592]. This modified speaking style, known as *clear speech*, is typically spoken more slowly than conversational speech [Picheny, M. A., *et al.* (1986). J. Speech Hear. Res. **29**, 434–446; Uchanski, R. M., *et al.* (1996). J. Speech Hear. Res. **39**, 494–509]. However, talkers can produce clear speech at normal rates (clear/normal speech) with training [Krause, J. C., and Braida, L. D. (2002). J. Acoust. Soc. Am. **112**, 2165–2172] suggesting that clear speech has some inherent acoustic properties, independent of rate, that contribute to its improved intelligibility. Identifying these acoustic properties could lead to improved signal processing schemes for hearing aids. Two global-level properties of clear/normal speech that appear likely to be associated with improved intelligibility are increased energy in the 1000−3000-Hz range of long-term spectra and increased modulation depth of low-frequency modulations of the intensity envelope [Krause, J. C., and Braida, L. D. (2004). J. Acoust. Soc. Am. **115**, 362–378]. In an attempt to isolate the contributions of these two properties to intelligibility, signal processing transformations were developed to manipulate each of these aspects of conversational speech independently. Results of intelligibility testing with hearing-impaired listeners and normal-hearing listeners in noise suggest that (1) increasing energy between 1000 and 3000 Hz does not fully account for the intelligibility benefit of clear/normal speech, and (2) simple filtering of the intensity envelope is generally detrimental to intelligibility. While other manipulations of the intensity envelope are required to determine conclusively the role of this factor in intelligibility, it is also likely that additional properties important for highly intelligible speech at normal speech at normal rates remain to be identified. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097491]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.70.Gr, 43.70.Fq [MSS] Pages: 3346–3357

I. INTRODUCTION

It has been well established that *clear speech*, a speaking style that many talkers adopt in difficult communication situations, is significantly more intelligible (roughly 17 percentage points) than conversational speech for both normal-hearing and hearing-impaired listeners in a variety of difficult listening situations (e.g., Picheny *et al.*, 1985). Specifically, the intelligibility benefit of clear speech (relative to conversational speech) has been demonstrated for backgrounds of wideband noise (e.g., Uchanski *et al.*, 1996; Krause and Braida, 2002) and multi-talker babble (e.g., Ferguson and Kewley-Port, 2002), after high-pass and low-pass filtering (Krause and Braida, 2003) and in reverberant environments (Payton *et al.*, 1994). Similar intelligibility benefits have also been shown for a number of other listener populations, including cochlear-implant users (Liu *et al.*, 2004), elderly adults (Helfer, 1998), children with and without learning disabilities (Bradlow *et al.*, 2003), and non-native listeners (Bradlow and Bent, 2002; Krause and Braida, 2003). These large and robust intelligibility differences are, not surprisingly, associated with a number of acoustical differences between the two speaking styles. For example, clear speech is typically spoken more slowly than conversational speech (Picheny *et al.*, 1986; Bradlow *et al.*, 2003), with greater temporal envelope modulations (Payton *et al.*, 1994; Krause and Braida, 2004; Liu *et al.*, 2004) and with relatively more energy at higher frequencies (Picheny *et al.*, 1986; Krause and Braida, 2004). In addition, increased fundamental frequency variation (Krause and Braida, 2004; Bradlow *et al.*, 2003), increased vowel space (Picheny *et al.*, 1986; Ferguson and Kewley-Port, 2002), and other phonetic and phonological modifications (Krause and Braida, 2004; Bradlow *et al.*, 2003) are apparent in the clear speech of some talkers. Any of these differences between clear and conversational speech could potentially play a role in increasing intelligibility. However, the particular properties of clear speech that are responsible for its intelligibility advantage have not been isolated.

One approach that can be used to evaluate the relative contribution of various acoustic properties to the intelligibility advantage of clear speech is to develop signal processing schemes capable of manipulating each acoustic parameter independently. By comparing the intelligibility of speech before and after an individual acoustical parameter is altered, the contribution of that acoustic parameter to intelligibility

[a] Author to whom correspondence should be addressed. Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL. Electronic mail: jkrause@cas.usf.edu

can be quantified. Of course, it may be necessary to modify a combination of multiple parameters to reproduce the full benefit of clear speech; however, manipulation of single acoustic parameters is nonetheless a convenient tool for characterizing their relative importance (at least to the extent that intelligibility is a linear combination of these parameters). The effect of each parameter on intelligibility can be isolated, and its role in improving intelligibility can be evaluated systematically.

Several studies have used this systematic approach to analyze the role of speaking rate in the intelligibility of clear speech, each employing a different time-scaling procedure to alter the speaking rate of clear and conversational sentences; clear sentences were time-compressed to typical conversational speaking rates (200 wpm), and conversational sentences were expanded to typical clear speaking rates (100 wpm) (Picheny *et al.*, 1989; Uchanski *et al.*, 1996; Liu and Zeng, 2006). Even after accounting for processing artifacts, none of the time-scaling procedures produced speech that was more intelligible than unprocessed conversational speech (although nonuniform time-scaling, which altered the duration of phonetic segments based on segmental-level durational differences between conversational and clear speech, was less harmful to intelligibility than uniform time-scaling). Using the same systematic approach, other studies have evaluated the role of pauses in the intelligibility advantage provided by clear speech. Results suggest that introducing pauses to conversational speech does not improve intelligibility substantially, unless the short-term signal-to-noise ratio (SNR) is increased as a result of the manipulation (Uchanski *et al.*, 1996; Liu and Zeng, 2006).

From these artificial manipulations of clear and conversational speech, it can thus be concluded that neither speaking rate nor pause structure alone is responsible for a large portion of the intelligibility benefit provided by clear speech. In fact, results obtained for naturally produced clear speech are consistent with this conclusion. With training, talkers can learn to produce a form of clear speech at normal speaking rates (Krause and Braida, 2002). This speaking style, known as *clear/normal* speech, is comparable to conversational speech in both speaking rate and pause structure (Krause and Braida, 2004). Nonetheless, it provides normal-hearing listeners (in noise) with 78% of the intelligibility benefit afforded by typical clear speech (i.e., 14 percentage points for clear/normal speech vs 18 percentage points for clear speech produced at slower speaking rates; Krause and Braida, 2002).

Taken together, these results suggest that clear speech has some inherent acoustic properties, independent of rate, that account for a large portion of its intelligibility advantage. In addition, the advent of clear/normal speech has simplified the task of isolating these properties, since conversational and clear speech produced at the same speaking rate can be compared directly. Such comparisons have revealed a number of acoustical differences between clear/normal speech and conversational speech that may be associated with the differences in intelligibility between the two speaking styles (Krause and Braida, 2004). In particular, the two properties of clear/normal speech that appear most likely responsible for its intelligibility advantage are increased energy in the 1000–3000-Hz range of long-term spectra and increased modulation depth of low-frequency modulations of the intensity envelope (Krause and Braida, 2004). In an attempt to determine the extent to which these properties influence intelligibility, the present study examines the effects of two signal processing transformations designed to manipulate each of these characteristics of conversational speech independently. One transformation alters spectral characteristics of conversational speech by raising formant amplitudes typically found between 1000 and 3000 Hz, while the other alters envelope characteristics by increasing the modulation depth of low frequencies ($<3-4$ Hz) in several of the octave-band intensity envelopes.

In this paper, each of these signal processing transformations is described, verified, and independently applied to conversational speech. Intelligibility results are then reported for both types of processed speech as well as naturally produced conversational and clear/normal speech. The goal of the intelligibility tests was to determine the relative contribution of each of these properties to the intelligibility advantage of clear speech. If altering either of these properties improved intelligibility substantially without altering speaking rate, the corresponding transformations would provide insight into signal processing approaches for hearing aids that would have the potential to improve speech clarity as well as audibility.

II. SIGNAL TRANSFORMATIONS

Based on previously identified properties of clear/normal speech (Krause and Braida, 2004), two signal transformations were developed to alter single acoustic properties of conversational speech. Specifically, the transformations were as follows.

(1) **Transformation SPEC**— This transformation increased energy near second and third formant frequencies. These formants typically fall in the 1000–3000-Hz range (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995) where differences in long-term spectra are typically observed between conversational and clear/normal speech. The spectral differences between the two speaking styles are thought to arise from the emphasis of second and third formants in clear/normal speech; higher spectral prominences at these formant frequencies are generally evident in the short-term vowel spectra of clear/normal speech relative to conversational speech, while little spectral change in short-term consonant spectra is evident across speaking styles (Krause and Braida, 2004).

Although Transformation SPEC is similar to a high-frequency emphasis of the speech spectrum, such as what would be accomplished by frequency-gain characteristics commonly used in hearing aids, this transformation manipulates only the frequency content of vowels and other voiced segments. As a result, the increase in level of F2 and F3 relative to F1 is somewhat greater than what would result from applying a high-frequency emphasis to the entire sentence, assuming that the long-term rms level of each sentence is held constant. Because a spectral boost of this magnitude is so common in the 1000–3000-Hz frequency range for vowels in clear/normal speech, it is important to quantify its contribution, if any, to the intelligibility advantage of clear speech.

(2) **Transformation ENV**— This transformation increased the modulation depth of frequencies less than 3–4 Hz in the intensity envelopes of the 250-, 500-, 1000-, and 2000-Hz octave bands. This type of change is often exhibited by talkers who produce clear speech at normal rates (Krause and Braida, 2004) and is also generally evident when talkers produce clear speech at slow rates (Payton *et al.*, 1994; Krause and Braida, 2004; Liu *et al.*, 2004). For these reasons, and also because modulations as low as 2 Hz are known to be important for phoneme identification (Drullman *et al.*, 1994a, 1994b), the increased modulation of lower frequencies in these octave bands is considered likely to contribute to improved intelligibility (Krause and Braida, 2004). Further evidence for this idea stems from the (speech-based) Speech Transmission Index (STI) (Houtgast and Steeneken, 1985): The speech-based STI is not only directly related to envelope spectra but also highly correlated with measured intelligibility scores for conversational speech and clear speech at a variety of speaking rates, suggesting that at least some of the differences in envelope spectra between the two speaking styles are associated with differences in intelligibility (Krause and Braida, 2004).

Before the intelligibility effects of these transformations were measured, acoustic evaluations were conducted to verify that each transformation had produced the desired change in acoustic properties. For the purposes of these acoustic evaluations, each transformation was applied to the conversational speech of the two talkers (T4 and T5) analyzed in Krause and Braida (2004) so that the acoustic properties of the processed speech could be directly compared to previously reported acoustic data for the clear/normal speech of these same two talkers (Krause and Braida, 2004). In that study, speech was drawn from a corpus of nonsense (grammatically correct but semantically anomalous) sentences previously described by Picheny *et al.* (1985), and one set of 50 nonsense sentences per talker was analyzed in both clear/normal and conversational speaking styles, such that 200 utterances (100 unique sentences) were analyzed between the two talkers. For the acoustic evaluations in this study, each transformation was applied to the 100 conversational utterances analyzed in that study. Thus, for each of the two talkers, it was possible to compare the acoustic properties of the 50 processed sentences to the acoustic properties previously reported for the exact same 50 sentences spoken in both conversational and clear/normal speaking modes.

A. Transformation SPEC: Formant frequencies

The first processing scheme, Transformation SPEC, increased energy near the second and third formants by first modifying the magnitude of the short-time Fourier transform (STFT) and then using the Griffin–Lim (Griffin and Lim, 1984) algorithm to estimate a signal from its modified STFT magnitude. The STFT magnitude, or spectrogram, was computed using an 8-ms Hanning window with 6 ms of overlap. The formant frequencies were then measured at 10-ms intervals for voiced portions of the speech signal using the formant tracking program provided in the ESPS/WAVES+ software package. For each 10-ms interval where voicing was present, the spectrogram magnitude was multiplied by a modified Hanning window, $w[F]$, whose endpoints in frequency, F_{start} and F_{end}, were calculated as follows:

$$F_{\text{start}} = F_2 - \min\left(2\text{BW}_2, \frac{F_1 + F_2}{2}\right), \quad (1)$$

$$F_{\text{end}} = F_3 + \min\left(2\text{BW}_3, \frac{F_2 + F_3}{2}\right), \quad (2)$$

where F_2, F_3, BW_2, and BW_3 are the second and third formants and their bandwidths, respectively. A Hanning window spanning this frequency range, $h[F]$, was modified according to the following formula:

$$w[F] = Ah[F] + 1, \quad (3)$$

where A, the scale factor used to control the amount of amplification, was set to 2 in order to achieve an energy increase comparable in magnitude to that previously reported between conversational and clear/normal speech (Krause and Braida, 2004). Finally, the Griffin–Lim (Griffin and Lim, 1984) iterative algorithm was used to derive the processed speech signal from the modified spectrogram, and the resulting sentences were then normalized for long-term rms value.

In order to evaluate whether Transformation SPEC achieved the desired effect on energy in the 1000–3000-Hz range, the long-term (sentence-level) and short-term (phonetic-level) spectra of processed speech were then computed for the two talkers from Krause and Braida (2004) and compared to the spectra previously obtained for these talkers' conversational and clear/normal speech (Krause and Braida, 2004). As in Krause and Braida (2004), FFTs were computed for each windowed segment (25.6 ms nonoverlapping Hanning windows) within a sentence, and then the rms average magnitude was determined over 50 sentences. A 1/3-octave representation of the spectra was obtained by summing components over 1/3-octave intervals with center frequencies ranging from 62.5 to 8000 Hz.

Figure 1 shows the long-term spectral differences of clear/normal and processed modes relative to conversational speech for T4 (results for T5 were similar), demonstrating that the processing had the desired effect on the long-term spectrum. Some spectral differences are apparent between clear/normal speech and conversational speech below 1 kHz (i.e., near F1). However, it is the spectral differences above 1 kHz that are thought to be related to the intelligibility benefit of clear/normal speech, as these differences are seen most consistently across talkers and vowel contexts (Krause and Braida, 2004). In this frequency range, the processed speech exhibits roughly the same increase in energy, relative to conversational/normal speech, as clear/normal speech does. Inspection of short-term spectra confirmed that this ef-

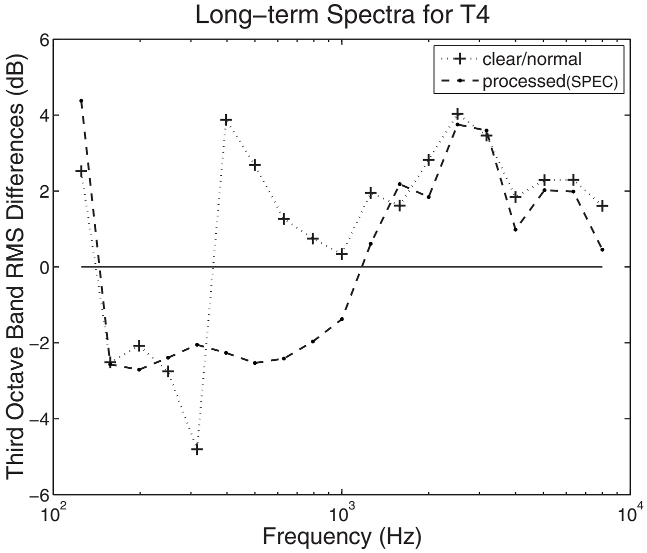


FIG. 1. Third-octave band long-term rms spectral differences for T4, a talker whose clear/normal speech has been previously analyzed in Krause and Braida (2004). Spectral differences were obtained by subtracting the conversational (i.e., unprocessed) spectrum from the clear/normal and processed spectra.

fect was due to the expected spectral changes near the second and third formants of vowels, with little change in consonant spectra.

B. Transformation ENV: Temporal envelope

Transformation ENV was designed to increase the modulation depth for frequencies less than 3–4 Hz in the intensity envelope of the 250-, 500-, 1000-, and 2000-Hz octave bands using an analysis-synthesis approach (Drullman *et al.*, 1994b). In the analysis stage, as in Krause and Braida (2004), 50 sentences from a talker were first concatenated and filtered into seven component signals, using a bank of fourth-order octave-bandwidth Butterworth filters, with center frequencies of 125–8000 Hz. The filters used in this stage of processing were designed so that the overall response of the combined filters was roughly ±2 dB over the entire range of speech frequencies. The filter bank outputs for each of the seven octave bands were then squared and low-pass filtered by an eighth-order Butterworth filter with a 60-Hz cutoff frequency in order to obtain relatively smooth intensity envelopes. After downsampling the intensity envelopes in each octave band by a factor of 100, a 1/3-octave representation of the power spectra for each band was computed, with center frequencies ranging from 0.4 to 20 Hz. Finally, the power spectra were normalized by the mean of the envelope function (Houtgast and Steeneken, 1985), such that a single 100% modulated sine-wave would result in a modulation index of 1.0 for the 1/3-octave band corresponding to the modulation frequency and 0.0 for the other 1/3-octave bands.

In the second stage of processing, the envelope of each sentence was processed separately. For the octave bands in which modification was desired, the original envelope of each sentence was modified by a 200-point FIR filter designed to amplify frequencies between 0.5 and 4 Hz. The amount of amplification was set so that the resulting modulation depths for these frequencies would be at least as large as those found in clear/normal speech and would generally fall within the range of values previously reported for clear speech, regardless of speaking rate (Krause and Braida, 2004). This range of modulation depths was targeted because the envelope spectra of both clear/normal speech and clear speech produced at slow rates (i.e., clear/slow speech) are associated with improved intelligibility (Krause and Braida, 2002). After the filter was applied, the modified envelope was adjusted to have the same average intensity as the original sentence envelope, and then any negative values of the adjusted envelope were set to zero. If resetting the negative portions of the envelope to zero affected the average intensity substantially, the intensity adjustment procedure was repeated until the average intensity of the modified envelope was within 0.5% of the average intensity of the original envelope. The modified envelope and original envelope were then upsampled to the original sampling rate of the signal in order to prepare for the final synthesis stage of processing.

Although it was convenient to work with intensity envelopes in the first two stages of processing so that the desired intensity envelope spectra could be achieved in the modified signal, the *amplitude* envelope was necessary for synthesis. Therefore, during the final processing stage, the time-varying ratio of the amplitude envelopes was calculated by comparing the square-root of the modified intensity envelope with the square-root of the original intensity envelope. The original octave-band signals were then transformed by multiplying the original signal in each octave band (with fine structure) by the corresponding time-varying amplitude ratio for that band. In order to ensure that no energy outside the octave band was inadvertently amplified, the result was also low-pass filtered by a fourth-order Butterworth filter with cutoff frequency corresponding to the upper cutoff frequency for the octave band. The processed version of the signal was then obtained by summing the signals in each octave band. Lastly, the processed sentences were normalized for long-term rms level.

After synthesis was completed, it was determined through informal listening tests that these modifications caused the speech of the female talker (T4) to sound more unnatural than the male talker (T5). The transformation was applied to two additional female talkers with the same results. A likely explanation for this problem was thought to be that the fundamental frequency of the female talkers tends to fall in the second octave band, and amplifying slowly varying modulations of voicing is not likely to occur in natural speech unless the talker slows down. This explanation was supported by the acoustic data, since an increase in modulation index was not exhibited in the 250-Hz band for clear/normal speech for the female talker, T4, although it was present in the 500-Hz band (Krause and Braida, 2004). Informal listening tests confirmed that eliminating the envelope modification in the 250-Hz band improved the quality of the female talkers' processed speech. Therefore, the signal transformation procedure was specified to modify only the 500-, 1000-, and 2000-Hz bands for female talkers.

In order to evaluate the effect of the processing on the

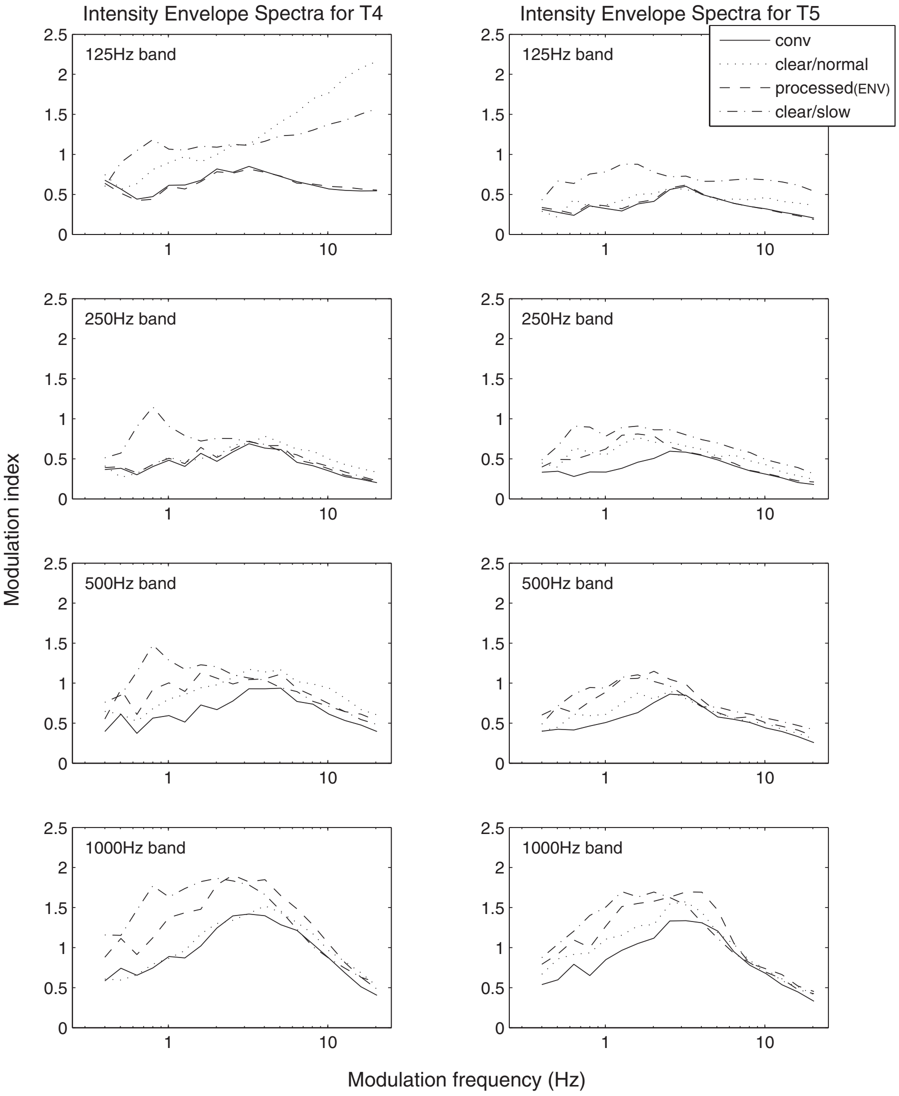


FIG. 2. Spectra of intensity envelopes in the lower four octave bands for T4 and T5, talkers whose clear/normal and clear/slow speech have been previously analyzed in Krause and Braida (2004).

intensity envelopes, the envelope spectra of the processed speech were computed for the two talkers from Krause and Braida (2004) and compared to the envelope spectra previously obtained for these talkers' unprocessed (conversational) speech, as well as their clear speech at both normal and slow rates (clear/normal and clear/slow). The spectra of the octave-band intensity envelopes for both talkers in each speaking style are shown in Figs. 2 and 3. From these figures, it can be seen that the processing had the intended effect on the spectra of the octave-band intensity envelopes, with envelope spectra of processed speech falling between that of previously measured clear/normal and clear/slow spectra (Krause and Braida, 2004) for frequencies less than 3–4 Hz in the specified octave bands (500-, 1000-, and 2000 Hz for both talkers as well as 250 Hz for T5, the male talker) and no substantial changes in the remaining octave bands.

III. INTELLIGIBILITY TESTS

To assess the effectiveness of the signal transformations, the intelligibility of the processed speech was measured and compared to the intelligibility of naturally produced conversational speech and clear speech at normal rates. Intelligibility in each speaking style was measured by presenting processed and unprocessed speech stimuli to normal-hearing listeners in the presence of wideband noise as well as to hearing-impaired listeners in a quiet background.

A. Speech stimuli

Speech stimuli were generated from speech materials recorded for an earlier study of clear speech elicited naturally at normal speaking rates (Krause and Braida, 2002). In that study, nonsense sentences (e.g., *His right cane could guard an edge.*) from the Picheny corpus (Picheny *et al.*, 1985) were recorded by five talkers in a variety of speaking styles. Materials were selected from one male (T5) and three female (T1, T3, and T4) talkers, because these four talkers obtained relatively large intelligibility benefits from clear/normal speech (11–32 percentage points relative to conversational speech) and were therefore most likely to benefit from signal transformations based on its acoustic properties. The materials selected for T4 and T5 were generally different than those used for these talkers in the acoustic evaluations of the transformations, except as noted below.

For each of the four talkers, 90 sentences recorded in a conversational speaking style and 30 sentences recorded in a

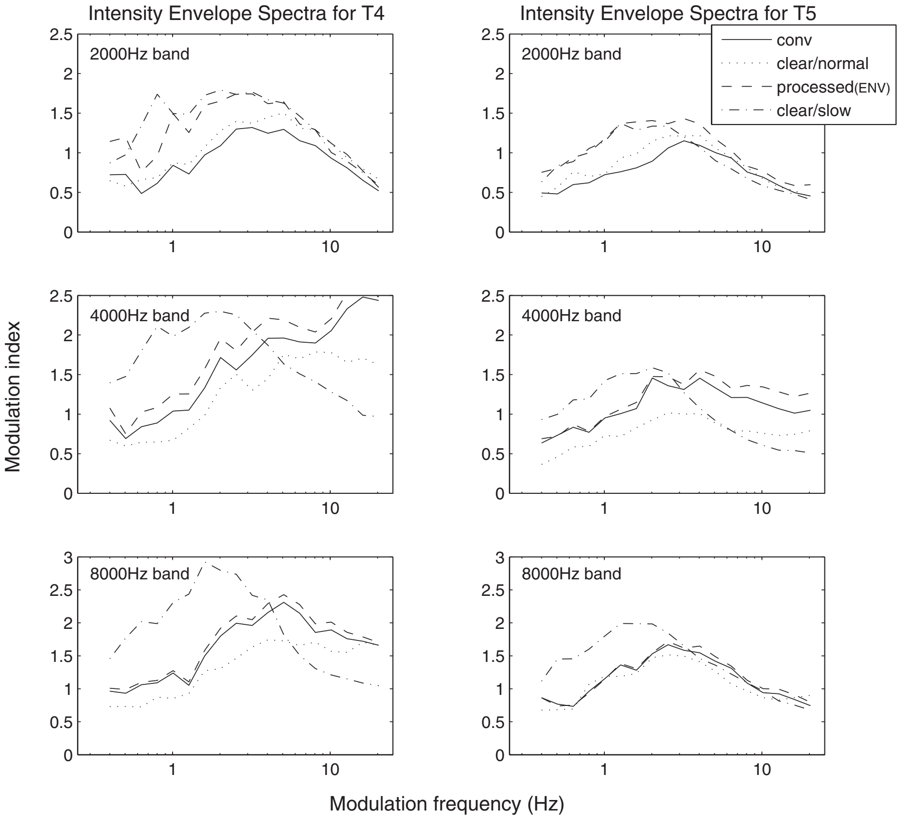


FIG. 3. Spectra of intensity envelopes in the upper three octave bands for T4 and T5, talkers whose clear/normal and clear/slow speech have been previously analyzed in Krause and Braida (2004).

clear/normal speaking style were used for this experiment. In one case (T1), 30 clear sentences that had been elicited at a quick speaking rate were used rather than those elicited in the clear/normal style, since these sentences were produced at about the same speaking rate as the clear/normal sentences but received higher intelligibility scores (Krause and Braida, 2002). Of the 90 conversational sentences used for each talker, 30 remained unprocessed, 30 were processed by Transformation SPEC, and 30 were processed by Transformation ENV (for T4 and T5, 10 of these sentences had been used previously in the acoustical evaluations), resulting in 30 unique sentences per condition per talker. Because different sentences were used for each talker, a total of 480 unique sentences (30 sentences × 4 conditions × 4 talkers) were thus divided evenly between the four different speaking conditions: conversational, processed(SPEC), processed(ENV), and clear/normal. The two conditions that were naturally produced (i.e., unprocessed) were included as reference points for the processed conditions, with conversational speech representing typical intelligibility and clear/normal speech representing the maximum intelligibility that talkers can obtain naturally by speaking clearly without altering speaking rate.

B. Listeners

Eight listeners were recruited from the MIT community to evaluate the intelligibility of the speech stimuli. All of the listeners were native speakers of English who possessed at least a high school education. Five of the listeners (one male and four females; age range: 19–43 years) had normal hearing, with thresholds no higher than 20 dB HL for frequencies between 250 and 4000 Hz, while three of the listeners (three males; age range: 40–65 years) had stable sensorineural hearing losses that were bilateral and symmetric. The audiometric characteristics for the test ears of the hearing-impaired listeners are summarized in Table I. For these listeners, the intelligibility tests were administered either to the ear with

TABLE I. Audiometric characteristics for the test ears of the hearing-impaired listeners. NR indicates no response to the specified frequency.

				Word[a] recognition	Thresholds (dB HL)					
Listener	Sex	Age	Test ear	(%)	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
L6HI	M	65	Left	100	55	60	45	45	55	85
L7HI	M	64	Left	92	10	20	40	60	65	NR
L8HI	M	40	Right	100	50	55	55	60	90	85

[a] Either NU-6 (Tillman and Carhart, 1996) or W-22 (Hirsh *et al.*, 1952).

better word recognition performance during audiometric testing or to the preferred ear if no such difference in word recognition performance was observed.

It is worth noting here that the purpose of the intelligibility tests was not to make comparisons between hearing-impaired listeners and normal-hearing listeners but simply to have all eight listeners evaluate the intelligibility benefit of clear/normal speech and the two signal-processing transformations. For nonsense sentences such as those used in this study, it has been shown that the intelligibility *benefit* of clear speech (relative to conversational speech) is roughly the same for normal-hearing listeners (in noise) as for hearing-impaired listeners (in quiet), despite differences in absolute performance levels (Uchanski *et al.*, 1996). For example, Uchanski *et al.* (1996) reported that clear speech improved intelligibility by 15–16 points on average for both normal-hearing listeners (clear: 60% vs conversational: 44%) and hearing-impaired listeners (clear: 87% vs conversational: 72%), including listeners with audiometric configurations similar to the listeners in this study. Similarly, Payton *et al.* (1994) reported that the clear speech intelligibility benefit obtained by each of their two hearing-impaired listeners fell within the range of benefits obtained by the ten normal-hearing listeners in that study. In both cases, the percentage change in intelligibility was different for the two groups [larger for normal-hearing listeners in the study by Uchanski *et al.* (1996); larger for hearing-impaired listeners in the study by Payton *et al.* (1994)], but the intelligibility benefit (absolute difference in percentage points between clear and conversational speech) was roughly the same. Given these data, listeners were not divided on the basis of hearing status in this study, because the only intelligibility measures planned were measures of intelligibility benefit.

C. Procedures

All listeners were tested monaurally over TDH-39 headphones. Normal-hearing listeners were tested in the presence of wideband noise, and hearing-impaired listeners were tested in quiet. As described above, clear speech typically provides about the same amount of benefit to both types of listeners under these test conditions (e.g., Payton *et al.*, 1994; Uchanski *et al.*, 1996).

Normal-hearing listeners were seated together in a sound-treated room and tested simultaneously. Each normal-hearing listener selected the ear that would receive the stimuli and was encouraged to switch the stimulus to the other ear when fatigued. For stimulus presentation, stereo signals were created for each sentence, with speech on one channel and speech-shaped noise (Nilsson *et al.*, 1994) of the same rms level on the other channel. The speech was attenuated by 1.8 dB and added to the speech-shaped noise, and the resulting signal ($\mathrm{SNR} = -1.8$ dB) was presented to the listeners from a PC through a Digital Audio Labs (DAL) soundcard.

Hearing-impaired listeners were tested individually in a sound-treated room. A linear frequency-gain characteristic was obtained for each hearing-impaired listener using the NAL-R procedure (Byrne and Dillon, 1986) and then implemented using a third-octave filter bank (General Radio, model 1925). This procedure provided frequency-shaping and amplification based on the characteristics of the individual's hearing loss. At the beginning of each condition, the listener was also given the opportunity to adjust the overall system gain. The speech was then presented through the system to the listener from a DAL card on a PC. These procedures ensured that the presentation level of each condition was both comfortable and as audible as possible for each hearing-impaired listener.

Because it has been shown that learning effects for these materials are minimal (Picheny *et al.*, 1985), all listeners heard the same conditions in the same order. However, the presentation order for test conditions was varied across talkers such that no condition was consistently presented first (or last). Listeners were presented a total of sixteen 30-sentence lists (4 talkers × 4 conditions/talker) and responded by writing their answers on paper. They were given as much time as needed to respond but were presented each sentence only once. Intelligibility scores were based on the percentage of key words (nouns, verbs, and adjectives) identified correctly, using the scoring rules described by Picheny *et al.* (1985).

IV. RESULTS

The clear/normal speaking condition was most intelligible overall at 58% key words correct, providing a 13 percentage point improvement in intelligibility over conversational speech (45%). The size of this intelligibility advantage was consistent with the 14 percentage point advantage of clear/normal speech measured previously (Krause and Braida, 2002) for normal-hearing listeners in noise. Neither of the signal transformations, however, provided nearly as large of an intelligibility benefit as clear/normal speech. At 49%, processed(SPEC) speech was just 14 points more intelligible than conversational speech on average, while processed(ENV) speech (24%) was considerably *less* intelligible than conversational speech.

An analysis of variance performed on key word scores (after an arcsine transformation to equalize variances) showed that the main effect of condition was significant [$F(3,256)=273$, $p< 0.01$)] and accounted for the largest portion of the variance ($\eta^2=0.371$) in intelligibility. Post-hoc tests with Bonferroni corrections confirmed that overall differences between all conditions were significant at the 0.05 level. As expected, the main effects of listener [$F(7,256) =90$, $p< 0.01$)] and talker [$F(3,256)=62$, $p< 0.01$)] were also significant (since, in general, some talkers will be more intelligible than others and some listeners will perform better on intelligibility tasks than others), but the listener × talker interaction was not significant. For the purposes of this study, it is more important to note that all interactions of talker and listener with condition were significant but accounted for relatively small portions of the variance (among these terms, listener × talker × condition accounted for the largest portion of the variance, with $\eta^2=0.057$). Nonetheless, examination of these interactions provides additional insight regarding the relative effectiveness of each signal transformation.

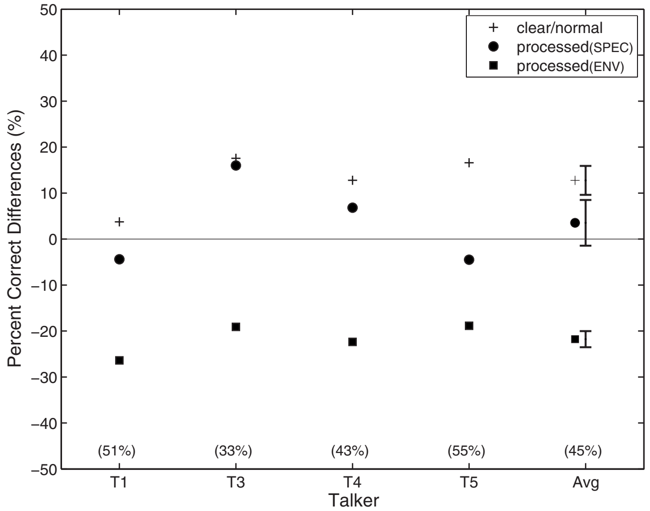


FIG. 4. Percent correct difference scores for each speaking condition (relative to conversational speech) obtained by individual talkers, averaged across listeners. Baseline conversational intelligibility scores are listed in parentheses. Errorbars at right indicate standard error of talker difference scores in each condition.

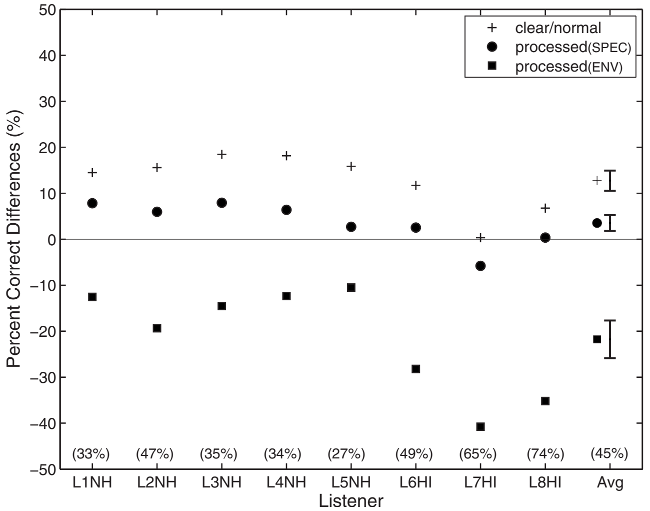


FIG. 5. Percent correct difference scores for each speaking condition (relative to conversational speech) obtained by individual listeners (NH designates normal-hearing listeners and HI designates hearing-impaired listeners), averaged across talkers. Baseline conversational intelligibility scores are listed in parentheses. Errorbars at right indicate standard error of listener difference scores in each condition.

Figure 4 shows that the effects of condition were largely consistent across individual talkers. For example, all talkers achieved a benefit with clear/normal speech, and the benefit was generally sizable (>10 points for three of the four talkers; note that this result was expected since the talkers were selected because they had previously demonstrated intelligibility improvements with clear/normal speech). Similarly, Transformation ENV was detrimental to intelligibility for all talkers: processed(ENV) speech was substantially less intelligible (19–27 points) than conversational speech. For Transformation SPEC, however, the benefit was not consistent across talkers. Instead, processed(SPEC) speech showed improved intelligibility for two talkers (T3: 16 points; T4: 7 points) and reduced intelligibility (relative to conversational speech) for the other two talkers (T1: −5 points; T5: −5 points).

Figure 5 shows the effects of condition across individual listeners. Again, the effects of clear/normal speech and processed(ENV) speech were generally consistent. For clear/normal speech, seven of eight listeners received an intelligibility benefit, with the average benefit across talkers ranging from 8 to 19 percentage points. The effect of Transformation ENV on listeners was similarly robust, although in the opposite direction. Processed(ENV) speech was less intelligible than conversational speech for all listeners, with differences in intelligibility ranging from −11 to −41 percentage points on average across talkers. The effect of Transformation SPEC, on the other hand, was less consistent. Although processed(SPEC) speech was more intelligible than conversational speech on average for six of eight listeners, these listeners did not receive a benefit from the processed(SPEC) speech of all talkers. All listeners received large benefits from processed(SPEC) speech for T3 (5–25 points), but none received any benefit for T1, and only about half received benefits of any size for T4 and T5, respectively. As a result, Fig. 5 shows that the size of the average intelligibility benefit that each listener received from processed(SPEC) speech was considerably smaller (3–8 points) than the benefit received from clear/normal speech.

With the exception of the processed(SPEC) speech of T4 and T5, the relative intelligibility of speaking conditions for any given talker was qualitatively very similar across individual listeners. For example, the benefit of clear/normal speech was so robust that the seven listeners who received a benefit from clear/normal speech on average across talkers (Fig. 5) also received an intelligibility benefit from each individual talker, and the remaining listener (L7HI) received a comparable benefit for two of the four talkers (T3 and T4) as well. Thus, clear/normal speech improved intelligibility for nearly all (30) of the 32 combinations of individual talkers and listeners, in most cases (24) by a substantial margin (>5 percentage points). Similarly, individual data confirm that Transformation ENV consistently decreased intelligibility; with only one exception (L5NH for T5), processed(ENV) speech was substantially less intelligible (<−5 percentage points) than conversational speech for each of the 32 combinations of individual talkers and listeners. Such consistency across four talkers and eight listeners (particularly listeners who differ considerably in age and audiometric profile) clearly shows that (1) Transformation ENV is detrimental to intelligibility, and (2) the benefit from clear/normal speech is considerably larger and more robust than the benefit of Transformation SPEC.

A. Hearing-impaired listeners

Although the relative intelligibility of conditions was qualitatively similar across listeners in general, a few differences were observed between hearing-impaired and normal-hearing listeners. In particular, the benefit of clear/normal speech was smaller for hearing-impaired listeners on average (7 points) than for normal-hearing listeners (17 points), and the detrimental effect of Transformation ENV was larger

(−35 vs −14 points). In addition, all three hearing-impaired listeners exhibited substantial decreases in intelligibility for the processed(SPEC) speech of T1 (−7 points on average) and T5 (−19 points) that were not typical of the normal-hearing listeners, who merely exhibited little to no benefit (T1: −3 points; T5: 4 points). As a result, Transformation SPEC did not improve intelligibility on average for hearing-impaired listeners but did provide a 6-point improvement for normal-hearing listeners, even though the processing provided both groups with comparable intelligibility improvements for T3 (17 points for NH listeners vs 14 points for HI listeners) and T4 (6 points vs 8 points). Whether any of these trends could reflect true difference(s) between populations cannot be determined from this study, because listeners were recruited without regard to audiometric profile. While these listeners provided valuable information regarding the relative intelligibility of the two signal transformations that were developed, more listeners (both hearing-impaired and normal-hearing) would be required to detect differences in performance between groups of listeners with different audiometric profiles.

B. Processing artifacts

To assess whether the potential benefits of either signal transformation may have been reduced or obscured by digital-signal processing artifacts, additional listeners were employed to evaluate speech that had been processed twice. For each signal transformation, twice-processed speech was obtained by first altering the specified acoustic parameter and then restoring the parameter to its original value. Thus, any reduction in intelligibility between the original (unprocessed) speech and the restored (twice-processed) speech would reflect only those deleterious effects on intelligibility specifically caused by processing artifacts.

For each talker, five conditions were tested, one conversational and four processed conditions: processed(SPEC), processed(ENV), restored(SPEC), and restored(ENV). Four additional normal-hearing listeners (all males; age range: 21–27 years old) each heard the speech of one talker in all five conditions, and each listener was presented speech from a different talker. The presentation setup was the same as described above, with one small difference: A SNR of 0 dB was used to avoid floor effects, since scores obtained in initial intelligibility tests for processed(ENV) speech were fairly low (e.g., 14% for T3). If processing artifacts were wholly responsible for these low scores, further reductions in intelligibility would be expected for restored(ENV) speech.

Average scores for the restored(SPEC) condition (45%) were the same as average scores for unprocessed conversational speech (45%), suggesting that processing artifacts associated with Transformation SPEC were negligible. In contrast, processing artifacts were a substantial issue for Transformation ENV, as the restored(ENV) condition was 19 points less intelligible than conversational speech. Notably, processed(ENV) speech showed some benefit relative to the restored(ENV) condition for two talkers (7 points for T3 and 15 points for T5). Therefore, it may be possible to achieve intelligibility improvements by altering the temporal envelope, if processing artifacts can be avoided. Also noteworthy is that while processed(SPEC) speech provided some benefit for three of four talkers in the initial experiment at SNR = −1.8 dB, the benefit was evident for only one of those talkers (T5) at SNR = 0 dB. Although this difference between experiments may have occurred by chance (particularly given that only one listener per talker was used at SNR = 0 dB), it is also possible that the benefit of Transformation SPEC for normal-hearing listeners diminishes as SNRs improve. This possibility will be discussed further in Sec. V.

V. DISCUSSION

Despite the fact that both signal transformations were based on the acoustics of clear/normal speech, results of intelligibility tests showed that neither transformation provided robust intelligibility improvements over unprocessed conversational speech. Transformation SPEC, which increased energy near second and third formant frequencies, improved intelligibility for some talkers and listeners, but the benefit was inconsistent and averaged just 4% overall. Transformation ENV, which enhanced low-frequency (<3–4 Hz) modulations of the intensity envelope, decreased intelligibility for all talkers and listeners, most likely because of the detrimental effects of processing artifacts associated with the transformation.

Although previous clear speech studies found intelligibility results for hearing-impaired listeners to be consistent with results for normal-hearing listeners in noise (Payton *et al.*, 1994; Uchanski *et al.*, 1996), some differences between these two populations were observed in this study. Most notably, Transformation SPEC improved intelligibility for normal-hearing listeners by 6 points on average but did not provide any benefit at all to hearing-impaired listeners on average. While both types of listeners did receive large benefits from processed(SPEC) speech relative to conversational speech for T3 (14 points for hearing-impaired listeners and 17 points for normal-hearing listeners), all three hearing-impaired listeners exhibited substantial *decreases* in intelligibility for the processed(SPEC) speech of T1 (−7 points on average) and T5 (−19 points on average) that were not typical of the normal-hearing listeners. These differences suggest that the benefit of Transformation SPEC may be associated with formant audibility for hearing-impaired listeners. That is, increasing the energy near F2 and F3 may improve the intelligibility of some talkers (e.g., T3), whose formants are not consistently audible to hearing-impaired listeners with NAL-R amplification, but may not improve the intelligibility of talkers whose formants are consistently audible (e.g., T1 and T5). Instead, listeners may perceive the processed sentences of these talkers as unnecessarily loud and respond by decreasing the overall level of amplification, thereby reducing the level of other frequency components and potentially decreasing sentence intelligibility.

In this case, it would seem likely that the benefit of formant processing for normal-hearing listeners would also be largely associated with formant audibility. To examine this possibility, band-dependent SNRs for each talker were calculated from the speech stimuli used in the intelligibility tests. Within each of the five third-octave bands over which formant modification occurred (center frequencies ranging from 1260 to 3175 Hz), a band-dependent SNR was measured by comparing the rms level of the talker's speech within that band to the rms level of speech-shaped noise within that band. These measurements confirm that T3 had the poorest band-dependent SNRs in this frequency region (−5.3 dB on average) of all talkers, while T1 had the highest (−0.3 dB). Thus, it is not surprising that all five normal-hearing listeners received large intelligibility improvements from the processed(SPEC) speech of T3—for whom raising the level of the formants could substantially increase the percentage of formants that were audible over the noise, but none received a benefit greater than 1% from the processed(SPEC) speech of T1—for whom a high percentage of formants in conversational speech were probably already above the level of the noise. Inspection of spectrograms for each talker is consistent with this explanation: A much higher percentage of second and third formants are at levels well above the level of the noise for T1 than for T3. Following this reasoning, a higher percentage of formants would also be audible for intelligibility tests conducted at higher SNRs, which also explains why only one of four listeners obtained a benefit from processed(SPEC) speech during the follow-up intelligibility tests that examined processing artifacts, which were conducted at SNR=0 dB.

That formant audibility would play a role in the intelligibility benefit of Transformation SPEC is not surprising, given its similarities to high-frequency spectral emphasis. Such similarities also suggest that any intelligibility improvement provided by Transformation SPEC should not be large; altering the spectral slope of frequency-gain characteristics used to present sentences in noise has little effect on sentence intelligibility, even for hearing-impaired listeners (van Dijkhuizen *et al.*, 1987, 1989). Unlike Transformation SPEC, which alters only the speech signal, however, the frequency-gain characteristic affects both the speech signal and the background noise, thus preserving band-dependent SNRs. In contrast, decreased spectral tilt can occur naturally when talkers produce speech in noisy environments (Summers *et al.*, 1988), leaving background noise unchanged. Although typically associated with large improvements in intelligibility, the decreased spectral tilt occurs in this circumstance in conjunction with several other acoustic changes, similar to those observed in clear speech (Summers *et al.*, 1988). In light of the intelligibility results for Transformation SPEC, it seems likely that one or more of those acoustic changes provides the bulk of that intelligibility benefit.

While the relative intelligibility of the other conditions in this study was qualitatively similar for hearing-impaired listeners and normal-hearing listeners in noise, a second difference observed between these populations is that hearing-impaired listeners received a smaller benefit from clear/normal speech (7 vs 17 points) and a larger detriment from processed(ENV) speech (−35 vs −14 points) than their normal-hearing counterparts. Given that only three hearing-impaired listeners were tested, these differences could certainly have occurred by chance. Furthermore, there is a possibility that speaking styles were differentially affected by the NAL-R amplification provided to hearing-impaired listeners, thereby altering the relative intelligibility differences between conditions for these listeners. In other words, it is possible that the smaller benefit from clear/normal speech occurred for hearing-impaired listeners because the NAL-R amplification improved the intelligibility of conversational speech relatively more than it improved clear/normal speech. However, previous examination of these issues in clear speech produced at slower rates suggests that this possibility is not likely; that is, the benefit of clear speech is typically independent of frequency-gain characteristic (Picheny *et al.*, 1985). Therefore, further investigation regarding the benefits of clear/normal speech for hearing-impaired listeners is warranted to determine whether the trend observed in this study reflects a true difference between populations.

If such a difference between the populations exists, one possibility is that the benefits of clear/normal speech may be related to age, since the hearing-impaired listeners in this study were older (40–65 years) than the normal-hearing listeners (19–43 years). Another possibility is that an individual's audiometric characteristics may be a factor in whether clear/normal speech can be of benefit. A close inspection of interactions between hearing-impaired listener and talker reveals that L7HI received little or no benefit from clear/normal speech, except when listening to T4's speech, while each of the other two listeners experienced moderate to large intelligibility gains from clear/normal speech for all talkers. Since L7HI also had the most precipitous hearing loss, it is possible that listeners with this type of audiometric configuration may not be as likely to benefit from clear/normal speech as other hearing-impaired listeners. If so, clear/normal speech would differ in this respect from clear speech at slow rates, which provides roughly the same amount of benefit to listeners with various audiometric configurations (Uchanski *et al.*, 1996). To address the question of whether age and/or audiometric characteristics are linked to an individual's ability to benefit from clear/normal speech, additional intelligibility tests targeting younger and older groups of listeners with various configurations and severity of hearing loss would be required.

VI. CONCLUSION

Of the two processing schemes examined in this study, only Transformation SPEC, the transformation associated with modification of formant frequencies, provided an intelligibility advantage over conversational speech. However, this benefit appeared to be largely a function of formant audibility: The transformation was more likely (1) to improve intelligibility for talkers with second and third formants that were relatively low in amplitude prior to processing and (2) to provide benefits for normal-hearing listeners in noise, who could take advantage of improvements in band-dependent SNRs associated with processing. Even so, the benefit that normal-hearing listeners in noise received from processed(SPEC) speech (6 percentage points on average) was less than half what they received from clear/normal speech (17 percentage points), suggesting that increased energy near second formant and third formants is not the only factor responsible for the intelligibility advantage of clear/normal speech for these listeners.

Another factor that is likely to contribute to the improved intelligibility of clear speech (Krause and Braida, 2004; Liu *et al.*, 2004), at least at high SNRs (Liu and Zeng, 2006), is increased depth of low-frequency modulations of the intensity envelope. Although Transformation ENV successfully increased the depth of these modulations in the speech signal, processing artifacts made it difficult to determine the extent to which, if at all, this factor accounts for the clear speech advantage. While processed(ENV) speech was less intelligible than unprocessed conversational speech, it was more intelligible than the restored(ENV) condition for two of four talkers, suggesting that intelligibility improvements associated with altering the temporal intensity envelope may be possible, if processing artifacts can be minimized. Given that Transformation ENV manipulated all low-frequency modulations uniformly, it is also possible that an unnatural prosodic structure imposed by the transformation is the source of the processing artifacts. If so, it may be helpful to develop a processing tool that allows for nonuniform alteration of the intensity envelope. With such a tool, it may be possible to enhance low-frequency modulations while maintaining the general prosodic structure of the speech. Speech manipulated in this manner could then be evaluated with further intelligibility tests in order to provide a better understanding of how increases in intensity envelope modulations are related to the clear speech benefit.

Although further research is needed, the results of the present study are an essential first step toward quantifying the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. By independently manipulating these acoustic parameters and systematically evaluating the corresponding effects on intelligibility, two important findings have been established. First, an increase in energy between 1000 and 3000 Hz does not fully account for the intelligibility benefit of clear/normal speech. One or more other acoustic factors must also play a role. Second, simple filtering of the intensity envelope to achieve increased depth of modulation is generally detrimental to intelligibility, even though this acoustic property is considered likely to be at least partly responsible for the intelligibility benefit of clear speech (Krause and Braida, 2004; Liu *et al.*, 2004). Therefore, future research investigating nonuniform alterations of the intensity envelope is required to isolate the effects of this factor on intelligibility. In addition, signal transformations and intelligibility tests aimed at identifying the role of other acoustic properties of clear/normal speech (Krause and Braida, 2004) are also needed. Such tests would provide additional information regarding the mechanisms responsible for the intelligibility benefit of clear speech and could ultimately lead to improved signal processing approaches for digital hearing aids.

ACKNOWLEDGMENTS

The authors wish to thank Jae S. Lim, Joseph S. Perkell, Kenneth N. Stevens, and Rosalie M. Uchanski for many helpful technical discussions. In addition, they thank J. Desloge for his assistance in scoring sentences. Financial support for this work was provided by a grant from the National Institute on Deafness and Other Communication Disorders (NIH Grant No. 5 R01 DC 00117).

Bradlow, A. R., and Bent, T. (**2002**). “The clear speech effect for non-native listeners,” J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). “Speaking clearly for children with learning disabilities: Sentences perception in noise,” J. Speech Lang. Hear. Res. **46**, 80–97.

Byrne, D., and Dillon, H. (**1986**). “The national acoustic laboratories new procedure for selecting the gain and frequency response of a hearing aid,” Ear Hear. **7**, 257–265.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). “Effect of reducing slow temporal modulations on speech reception,” J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). “Effect of temporal envelope smearing on speech reception,” J. Acoust. Soc. Am. **95**, 1053–1064.

Ferguson, S. H., and Kewley-Port, D. (**2002**). “Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. **112**, 259–271.

Griffin, D. W., and Lim, J. S. (**1984**). “Signal estimation from modified short-time Fourier transform,” IEEE Trans. Acoust., Speech, Signal Process. **32**, 236–243.

Helfer, K. S. (**1998**). “Auditory and auditory-visual recognition of clear and conversational speech by older adults,” J. Am. Acad. Audiol **9**, 234–242.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). “Acoustic characteristics of american english vowels,” J. Acoust. Soc. Am. **97**, 3099–3111.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., and Benson, R. W. (**1952**). “Development of materials for speech audiometry,” J. Speech Hear Disord. **17**, 321–337.

Houtgast, T., and Steeneken, H. (**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.

Krause, J. C., and Braida, L. D. (**2002**). “Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility,” J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2003**). “Effects of listening environment on intelligibility of clear speech at normal speaking rate,” Iran. Audiol. **2**, 39–47.

Krause, J. C., and Braida, L. D. (**2004**). “Acoustic properties of naturally produced clear speech at normal speaking rates,” J. Acoust. Soc. Am. **115**, 362–378.

Liu, S., and Zeng, F.-G. (**2006**). “Temporal properties in clear speech perception,” J. Acoust. Soc. Am. **120**, 424–432.

Liu, S., Rio, E. D., and Zeng, F.-G. (**2004**). “Clear speech perception in acoustic and electric hearing,” J. Acoust. Soc. Am. **116**, 2374–2383.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). “Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am. **95**, 1085–1099.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). “Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,” J. Acoust. Soc. Am. **95**, 1581–1592.

Peterson, G., and Barney, H. (**1952**). “Control methods used in a study of the vowels,” J. Acoust. Soc. Am. **24**, 175–184.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). “Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech,” J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). “Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,” J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). “Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech,” J. Speech Hear. Res. **32**, 600–603.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). “Effects of noise on speech production: Acoustic and perceptual analyses,” J. Acoust. Soc. Am. **84**, 917–928.

Tillman, T., and Carhart, R. (**1966**). “An expanded test for speech discrimination utilizing cnc monosyllabic words: Northwestern university auditory test no. 6," Technical Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, TX.

Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.

van Dijkhuizen, J., Anema, P., and Plomp, R. (**1987**). "The effect of varying the slope of the amplitude-frequency response on the masked speech-reception threshold of sentences," J. Acoust. Soc. Am. **81**, 465–469.

van Dijkhuizen, J. N., Festen, J. M., and Plomp, R. (**1989**). "The effect of varying the slope of the amplitude-frequency response on the masked speech-reception threshold of sentences for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 621–628.

Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners

Joshua G. W. Bernstein[a] and Ken W. Grant
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

(Received 8 October 2008; revised 10 March 2009; accepted 10 March 2009)

Speech intelligibility for audio-alone and audiovisual (AV) sentences was estimated as a function of signal-to-noise ratio (SNR) for a female target talker presented in a stationary noise, an interfering male talker, or a speech-modulated noise background, for eight hearing-impaired (HI) and five normal-hearing (NH) listeners. At the 50% keywords-correct performance level, HI listeners showed 7–12 dB less fluctuating-masker benefit (FMB) than NH listeners, consistent with previous results. Both groups showed significantly more FMB under AV than audio-alone conditions. When compared at the same stationary-noise SNR, FMB differences between listener groups and modalities were substantially smaller, suggesting that most of the FMB differences at the 50% performance level may reflect a SNR dependence of the FMB. Still, 1–5 dB of the FMB difference between listener groups remained, indicating a possible role for reduced audibility, limited spectral or temporal resolution, or an inability to use auditory source-segregation cues, in directly limiting the ability to listen in the dips of a fluctuating masker. A modified version of the extended speech-intelligibility index that predicts a larger FMB at less favorable SNRs accounted for most of the FMB differences between listener groups and modalities. Overall, these data suggest that HI listeners retain more of an ability to listen in the dips of a fluctuating masker than previously thought. Instead, the fluctuating-masker difficulties exhibited by HI listeners may derive from the reduced FMB associated with the more favorable SNRs they require to identify a reasonable proportion of the target speech. [DOI: 10.1121/1.3110132]

PACS number(s): 43.71.Ky, 43.71.An, 43.71.Gv, 43.71.Rt [BCM] Pages: 3358–3372

I. INTRODUCTION

Listeners with sensorineural hearing loss experience difficulty understanding speech in the presence of a competing talker or a modulated noise. Unlike normal-hearing (NH) listeners, hearing-impaired (HI) listeners generally show little or no improvement in speech intelligibility for such fluctuating maskers relative to stationary noise when both maskers are presented at the same long-term sound pressure level (SPL) (e.g., Carhart and Tillman, 1970; Festen and Plomp, 1990; Eisenberg *et al.*, 1995; Bacon *et al.*, 1998; Peters *et al.*, 1998; Dubno *et al.*, 2003; George *et al.*, 2006; Jin and Nelson, 2006; Wilson *et al.*, 2007). The fluctuating-masker benefit (FMB) observed in NH individuals is generally interpreted in terms of "dip listening," whereby a listener is able to make use of momentary spectral and temporal dips in the level of the fluctuating masker (Alcántara and Moore, 1995).

One possible explanation for this deficit is that HI listeners lack certain hearing acuities, related to their sensorineural hearing loss, that are thought to be important for dip listening, such as spectral (e.g., Glasberg and Moore, 1986) or temporal (e.g., Oxenham and Moore, 1997; Nelson *et al.*, 2001) resolution, or sensitivity to low-level sounds. According to this line of reasoning, a listener's limited resolution in frequency and time limits the ability to pull target and masker apart from one another because they occupy the same spectral filters or temporal integration windows (Festen and Plomp, 1990). In addition, elevated absolute thresholds might limit the audibility of the target signal within the spectral and temporal gaps, potentially reducing the benefits of listening in fluctuating backgrounds. However, these types of psychophysical deficits have not been shown to entirely account for the inability of HI listeners to benefit from masker modulations. Although reduced audibility can limit FMB to some extent (Summers and Molis, 2004), NH listeners still show substantial FMB when elevated thresholds are simulated via the introduction of masking noise (Eisenberg *et al.*, 1995; Dubno *et al.*, 2003; George *et al.*, 2006). Simulated spectral-resolution impairment has been shown to reduce the FMB for NH listeners (ter Keurs *et al.*, 1993; Baer and Moore, 1994). However, there is little evidence of a relationship between frequency selectivity and the FMB for HI listeners (e.g., George *et al.*, 2006). Although significant correlations have been found between measures of temporal resolution and FMB (Dubno *et al.*, 2003; George *et al.*, 2006), this is only for relatively fast masker modulations of 16–50 Hz, much faster than the 1–10-Hz frequencies that dominate speech and speech-masker modulation spectra (Steeneken and Houtgast, 1980). Dubno *et al.* (2003) reported that the correlation strength between temporal resolution and FMB was reduced at these lower rates.

[a] Author to whom correspondence should be addressed. Electronic mail: joshua.bernstein@amedd.army.mil

A second possible explanation for the reduced FMB in HI listeners is that hearing loss may result in suprathreshold distortions that disrupt the cues normally available to identify the talker and the masker as separate sources and allow their perceptual segregation. For example, reduced frequency selectivity (Bernstein and Oxenham, 2006) or inefficient use of temporal fine structure information (i.e., fast fluctuations in the stimulus waveform; Moore *et al.*, 2006) may underlie poor pitch discrimination abilities in HI listeners, thereby reducing cues needed for separating simultaneous speech sources (e.g., Darwin and Hukin, 2000). In NH listeners, perceptual similarity between masker and target (e.g., same-gender or same-talker interferers) can sometimes lead to performance deficits relative to the stationary-noise case (Brungart, 2001; Freyman *et al.*, 2004) rather than a benefit. This deficit, termed informational masking, can occur even when target and competing speech are filtered into non-overlapping bands (Arbogast *et al.*, 2002), suggesting that it is not peripheral in origin. HI listeners might be susceptible to informational masking, even in situations where NH listeners show little difficulty in separating sources, due to a loss of peripheral cues to mark differences between target and masker. Kwon and Turner (2001) demonstrated that under situations involving degraded stimuli and therefore reduced redundancy of speech information, modulated maskers could actually impair speech intelligibility in NH listeners. They suggested that this could reflect an inability to distinguish target and masker due to the fact that the two signals contain related modulations.

A third possible reason that HI listeners do not receive as much benefit from masker modulations as NH listeners is that the amount of benefit may be dependent on the baseline signal-to-noise ratio (SNR) at which the FMB is estimated. Oxenham and Simonson (2009) measured speech intelligibility as a function of SNR for low- and high-pass filtered speech using NH listeners. In both cases, listeners received the greatest benefit in the interfering-talker condition (relative to the stationary-noise condition) for the lowest stationary SNRs tested (−6 dB), with the magnitude of the benefit diminishing with increasing SNR, disappearing completely above 0 dB. HI listeners often require a more favorable SNR than NH listeners for equivalent performance in a stationary masker. If the FMB is dependent on this baseline SNR, differences in FMB between NH and HI listeners could be confounded by differences in the stationary-noise SNR among subjects. The magnitude of the benefit might be more similar between NH and HI listeners if they were tested at comparable SNRs. If so, this would suggest that suprathreshold deficits discussed above (e.g., impaired spectral or temporal resolution or an inability to use temporal fine-structure information) only indirectly impact the FMB for HI listeners. A general reduction in intelligibility caused by such deficits would increase the SNR required for HI listeners to attain a given level of speech understanding, thereby reducing the FMB.

The current study sought to differentiate between these possible explanations for the reduced benefit to speech intelligibility from fluctuating maskers received by HI listeners. This was done in two ways. First, speech-intelligibility performance in stationary and fluctuating maskers was measured across a range of SNRs, thereby allowing an examination of the influence of changes in SNR on the results. Second, visual speech cues were introduced to improve performance in the stationary-noise condition. Visual cues can improve speech-intelligibility performance in two ways: (1) by providing additional phonetic information about the target speech and (2) by providing cues for simultaneous source-segregation. In quiet and relatively simple noise backgrounds (e.g., stationary noise), simultaneously-presented video of the talker's face improves speech intelligibility by providing information about the target speech itself (Walden *et al.*, 1981; Bernstein *et al.*, 2000; Auer and Bernstein, 2007), especially in the form of consonant place-of-articulation cues (Braida, 1991; Grant and Walden, 1996). Visual cues should therefore reduce the SNR required to achieve a given level of performance, allowing an estimate of the FMB at less favorable SNRs where a larger FMB is likely to be observed.

Visual speech stimuli can also provide cues for auditory-source segregation (Driver, 1996; Helfer and Freyman, 2005; Wightman *et al.*, 2006). This allowed a test of the hypothesis that a reduced ability to segregate speech from a modulated masker is responsible (at least in part) for the reduced benefit from masker fluctuations associated with hearing loss. Visual cues improve speech intelligibility more for a same-gender interfering talker than for a stationary-noise masker (Helfer and Freyman, 2005; Wightman *et al.*, 2006) and can even provide a benefit to the intelligibility of an audio target when the visual cues associated with the interfering talker are presented (Driver, 1996). A likely basis for this source-segregation benefit is the comodulation of visible facial movements with the envelope of the target acoustical signal, which has been shown to facilitate the detection of auditory speech signals in noise (Grant and Seitz, 2000; Grant, 2001). If a reduction in the effectiveness of auditory cues to mark differences between target and masker signals contributes to the lack of benefit from fluctuating maskers in HI listeners, then the introduction of visual source-segregation cues should increase the FMB for these listeners.

The current study estimated speech intelligibility for NH and HI listeners for audio-alone and AV speech presented in a stationary noise or a fluctuating masker. Two different fluctuating maskers were used (an opposite-gender interfering talker and a speech-modulated noise), with the idea that source-segregation might be more problematic for the HI listeners in the interfering-talker case. Although Hygge *et al.* (1992) also examined the role of visual cues on the intelligibility of speech presented in stationary-noise and interfering-talker maskers for NH and HI listeners, they only measured performance at a single performance level for each group and condition. Measuring performance across a range of SNRs for each condition made it possible to determine whether any increase in the magnitude of the FMB with the availability of visual cues could be attributed to SNR-dependent FMB changes. Any additional FMB under AV conditions in HI listeners, beyond that predicted due to stationary-noise SNR differences, would suggest that visual cues enhanced source separation. Furthermore, if HI listeners are suffering from a lack of auditory segregation cues to a

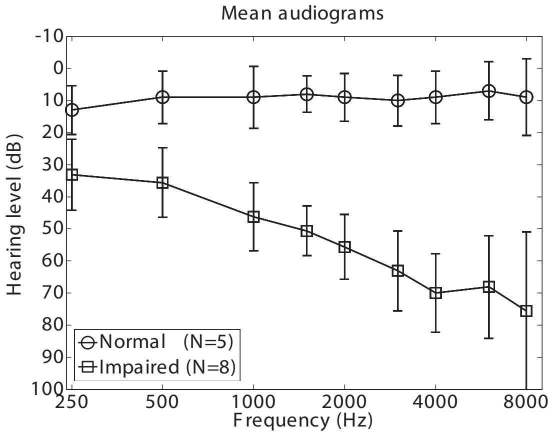


FIG. 1. Mean audiograms ± one standard deviation for the NH and HI listener groups.

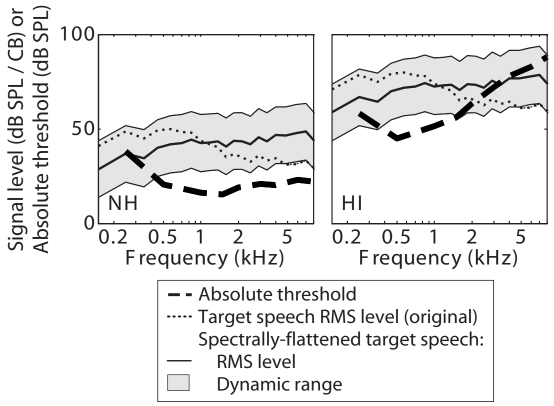


FIG. 2. Long-term rms power spectra (solid curves) for the spectrally flattened target speech and the associated ±15-dB dynamic range (shaded areas) presented to NH at 57 dB SPL (left panel) and to HI listeners at 87 dB SPL (right panel). The original speech power spectra before spectral shaping (dotted curves) are shown for comparison. Target signal audibility in quiet is represented by the proportion of the shaded area falling above the absolute hearing threshold (dashed curves).

greater extent than NH listeners, the introduction of visual segregation cues should lead to a greater increase in FMB for HI listeners than for NH listeners. Finally, if the FMB for HI listeners is still reduced relative to that for NH listeners after SNR differences are controlled, this would suggest that psychophysical deficits such as limited audibility or spectral or temporal resolution are likely involved in limiting the FMB for these listeners.

II. EXPERIMENT: AUDITORY AND AUDITORY-VISUAL SPEECH INTELLIGIBILITY IN STATIONARY AND FLUCTUATING MASKERS

A. Methods

1. Listeners

Five NH (two male, age range 30–58 years, mean 44.4 years) and eight HI listeners (all male, age range 49–80 years, mean 66.0 years) took part in the study. All listeners had near normal or corrected-normal vision (equal to or better than 20/50 as measured with a Snellen chart at 6 m). Mean audiometric thresholds for each listener group (± one standard deviation) are shown in Fig. 1. With two exceptions, the NH listeners had pure-tone audiometric thresholds in the test ear of 20 dB hearing level (HL) or better (ANSI, 2004) at octave frequencies between 250 and 8000 Hz as well as the inter-octave frequencies of 1500, 3000, and 6000 Hz. The exceptions were one NH listener with a threshold of 25 dB HL at 1000 Hz, and another with a threshold of 25 dB HL at 8000 Hz. A third NH listener had a slight high-frequency hearing loss in the non-test ear, with thresholds of 30–35 dB HL at 3000, 4000, and 6000 Hz.

The HI listeners had sensorineural hearing loss, defined by an absence of air-bone gaps greater than 10 dB for octave frequencies between 500 and 4000 Hz. Their air-conduction audiometric thresholds were between 45 and 85 dB HL at 2000, 4000, 6000, and 8000 Hz in the test ear, with the following exceptions: one listener had a threshold of 35 dB HL at 2000 Hz (but a threshold of 50 dB HL at 1000 Hz), a second listener had a threshold of 90 dB HL at 4000 Hz, and two other listeners were unable to detect an 8000-Hz tone presented at 100 dB HL. All eight HI listeners reported a history of occupational noise exposure. One of the HI listeners had also suffered additional sensorineural hearing loss (mostly in the non-test ear) in an acute barotraumatic accident, but had symmetric bilateral hearing loss before the accident.

2. Stimuli

Target stimuli consisted of the IEEE (1969) sentences spoken by a female native talker of American English. The IEEE sentences were selected because they are a phonetically balanced set of materials whose properties are well known (e.g., low-context, grammatically constrained) and because a prerecorded set of AV stimuli was available for the experiment. The speech was recorded audiovisually using a three-tube Ikegami color camera and stored on optical disk (Panasonic TQ-3031F). The audio portion of each production was digitized at a sampling rate of 20 kHz with 16-bit amplitude resolution. The digitized samples were ramped on and off (50-ms raised cosine), lowpass-filtered at 8.5 kHz, normalized in level so that all stimuli had the same average root-mean-squared (rms) amplitude, and upsampled to a sampling rate of 24.424 kHz. To reduce the influence of audibility on speech intelligibility in the HI listeners, the speech was spectrally flattened by applying a length-37 finite impulse response (FIR) filter that amplified the high frequencies (where the HI listeners tended to have more hearing loss and where the amplitude of the original speech signal was lower). The spectrally shaped speech signals were presented at fixed level of 57 (NH listeners) or 87 dB SPL (HI listeners). To illustrate the effects of these stimulus manipulations on audibility, long-term-average power spectra and associated dynamic ranges for the target signals are shown in Fig. 2, along with mean absolute hearing thresholds for the NH (left panel) and HI listeners (right panel). The average power spectra for the spectrally flattened speech [dB SPL per critical band (CB), as defined by ANSI, 1997] are plotted as thick solid curves, with the associated 30-dB effective dynamic ranges (±15 dB, ANSI, 1997) denoted by the shaded areas. The average power spectra for the original speech (before spectral shaping) are plotted for comparison (dotted curves). Group-mean absolute hearing thresholds are indicated by dashed lines. The proportion of the dynamic range audible (on average) to each group of listeners is represented by the proportion of the shaded area falling above the absolute hearing threshold. Although the spectral shaping and higher overall signal level improved audibility for the HI listeners, a substantial portion of the dynamic range was inaudible for frequencies above 2 kHz for this group. The speech-intelligibility index (SII) (ANSI, 1997) for speech in quiet, calculated for each group based on the mean audiogram, was 0.95 for NH listeners, but only 0.63 for HI listeners. Thus, despite the elevated signal level and spectral shaping, reduced audibility likely played a role in limiting speech intelligibility for the HI listeners.

Three different masker conditions were evaluated, with each masker spectrally shaped to match the long-term power spectrum of the 720 spectrally flattened target sentences. The maskers consisted of a Gaussian stationary noise, a male interfering talker, and a one-talker speech-modulated noise. The interfering-talker and modulated-noise conditions were chosen as they have been shown in previous studies to yield a reduced FMB in HI compared to NH listeners (e.g., Festen and Plomp, 1990; Peters *et al.*, 1998; Versfeld and Dreschler, 2002). For the interfering-talker condition, a long-duration masker was generated by concatenating the first half of the hearing in noise test (HINT) (Nilsson *et al.*, 1994) sentences spoken by a single male talker. Silences between sentences were removed by excluding the initial and final portions of each sentence that fell below −40 dB re the peak sentence level. The modulated-noise masker was generated as described by Festen and Plomp (1990). The first 12 sentences of the female-talker IEEE materials used as target sentences were concatenated in the same manner as for the HINT materials used for the interfering-talker condition. For each target sentence, a segment of the female-talker IEEE speech was selected and filtered (sixth-order Chebychev) into two bands (above and below 1 kHz). Envelopes were derived from the signal in each band via half-wave rectification and low-pass filtering (fourth-order Butterworth with a 40-Hz cutoff frequency) and were used to modulate a speech-shaped Gaussian noise filtered into the same two bands. The resulting signals were rescaled to the original long-term-average noise levels for each band (i.e., before the modulation was applied), and then summed together.

A segment of this long-duration masker was chosen at random for each target sentence and individual listener and was ramped on and off with 30-ms raised-cosine ramps. The masker started before and ended after the target for all sentences and conditions. The masker always terminated 250 ms after the target audio stimulus, while the onset delay of the target audio signal relative to the masker varied from sentence to sentence. This was because the masker audio signal started synchronously with the target video presentation in the AV conditions, and the video-audio onset delay varied across sentences. In selecting the video segments from a longer recording, the first frame of the video presentation was selected such that the talker was initially in a resting position. Since there was a segment of silence between the first frame of the video and the onset of the audio, this yielded an auditory masker-target onset delay that ranged from 68 to 1041 ms (mean=620 ms; standard deviation =134 ms) across the IEEE sentence database, but was fixed for each sentence token. In the audio-alone conditions, the masker-target delay was set as if the video signal had been present, yielding the same variable audio masker-target onset asynchrony. Although uncertainty in the masker-target onset delay may have negatively affected performance relative to a situation with a fixed-duration onset asynchrony, this effect was minimized because the asynchrony was the same for each successive presentation of a given sentence. Furthermore, the variation in the asynchrony was the same across all conditions and listeners.

For each sentence presentation, the target and masker were combined, then processed with a length-128 minimum-phase FIR filter to compensate (in the 125–8000-Hz frequency range) for the non-flat headphone magnitude response. The resulting digital signal was sent to an enhanced real-time processor (TDT RP2.1) where they were stored in a buffer. The control PC sent a serial command to the optical disk player to initiate the video playback, which in turn triggered the TDT RP2.1 to initiate the digital-to-analog conversion of the audio signal. This process ensured synchronization of the audio and video stimuli to within ±2 ms. The audio signal was passed though a headphone buffer (TDT HB7) before being presented to the listener through one earpiece of a Sennheiser HD580 headset. To prevent detection of the target speech signal in the contralateral ear via acoustic or electric crosstalk, a speech-shaped stationary-noise background (uncorrelated with the masking noise, where applicable) with a level 25 dB below that of the target speech was presented to the non-test ear. For those conditions where only an audio stimulus was desired, the same process was used with blank frames selected for video playback. The listener was seated inside a double-walled sound-attenuating chamber, approximately 1 m in front of a 21 in. color video monitor (Panasonic CT2082Y).

3. Procedure

Several considerations were taken into account in designing the testing procedure to measure performance across a range of SNRs. First, an adaptive procedure was implemented, rather than a method using a set of fixed SNRs, as it was difficult to predict an appropriate range of SNRs to yield a range of performance levels for each individual listener across the variety of conditions tested. Second, the IEEE stimulus set is limited, with only 72 lists of ten target sentences available, and it was desirable to limit the number of sentences presented to each listener, thereby enabling each listener to participate in future studies involving the IEEE stimuli. To address these issues, we used an adaptive paradigm whereby each sentence was presented several times with increasing (improving) SNR on each successive trial until at least four (out of five) keywords were identified correctly (Summers and Leek, 1998). This limited the number of unique sentences presented to each listener over the course of the experiment and identified the correct range of SNRs due to the adaptive nature of the procedure. This procedure was used instead of a traditional adaptive speech reception threshold (SRT) tracking procedure (e.g., Plomp and Mimpen, 1979) that estimates the SNR required for a specified level of performance because it was desirable to distribute the trials across a wide range of SNRs. The SRT method would have concentrated the majority of trials at SNRs near the specified performance level. For the interfering-talker and modulated-noise conditions, the masker segment was identical for each presentation of a given sentence, except that its level was adjusted from trial to trial in order to yield the desired SNR. This was done rather than selecting a new masker on every trial to prevent different parts of the target signal from being unmasked by valleys in the fluctuating masker on subsequent trials. For the stationary-noise condition, a new noise was generated on each signal trial.

The SNR was defined as the ratio between the rms levels of the target speech and masker. To achieve a given SNR, the target speech level was generally held constant and the masker level was adjusted. This was done, rather than adjusting the speech level for a fixed masker level, to ensure that any effects of SNR on speech intelligibility were not due to changes in the influence of absolute hearing thresholds on the audibility of the target speech. For the HI listeners, the adaptive track occasionally required a SNR below −12 dB, which would have required uncomfortable masker levels greater than 99 dB SPL. For these infrequently occurring cases, the SNR was achieved by fixing the masker level at 99 dB SPL and reducing the target speech level below 87 dB SPL. For the NH listeners, the lower fixed target speech level (57 dB SPL) allowed SNRs as low as −42 dB without the masker level exceeding 99 dB SPL.

For each run, the SNR was initially set to −15 dB (NH listeners) or −6 dB (HI listeners). These starting SNRs were selected in pilot tests to be 3 dB lower than the SNR needed to yield any measurable intelligibility for the audio-alone stationary-noise condition. After each stimulus presentation, the listener responded verbally by repeating back as much of the sentence as possible. The experimenter then entered into the computer the number (out of five) of keywords that were correctly identified. If the number of keywords correct was 3 or fewer, the SNR was increased (improved) by 3 dB and the sentence was repeated. If the listener reached or exceeded four keywords correct, the correct answer was displayed orthographically on the screen, the SNR was reduced (worsened) by 9, 12, or 15 dB (selected at random, each with probability 1/3), and the process repeated with a new sentence. To reduce testing time by preventing spurious jumps into a range of SNRs that tended to yield performance at floor levels, the SNR was not allowed to go below −39 dB (NH listeners) or −18 dB (HI listeners). The SNR was not allowed to go above +15 dB (both groups) to address the infrequent occurrence where a listener would become fixated on a particular (wrong) answer for a given sentence. If the +15-dB SNR was reached without the listener achieving four keywords correct, orthographic feedback was given, and the next sentence was presented beginning at a SNR randomly selected to be either −3 or 0 dB, each with probability 0.5.

This experimental procedure differed from more traditional methods of measuring speech intelligibility in that the same stimulus was presented repeatedly with increasing SNR. Theoretically this could have increased the estimated slope of the psychometric function, as a listener could have additional information about a particular sentence as a result of having identified portions of the same sentence on a previous trial. However, the potential contribution of any such *a priori* knowledge to intelligibility was limited by several factors. First, the low-context nature of the IEEE sentence set (Rabinowitz *et al.*, 1992) limited the opportunity to identify individual words in a given sentence based on contextual cues. Second, the same fluctuating maskers were used for each trial involving the same target sentence, such that different time segments of the target speech would not have been unmasked in subsequent trials. Third, each subsequent trial involved a higher SNR, such that any acoustic information received in previous trials would also have been available on the trial in question, even if the previous trials had not been presented. Finally, the only feedback given on any non-terminal trial for a given sentence was the information that fewer than four keywords were correctly identified. No additional feedback was given as to which or how many keywords were correctly identified.

A run usually consisted of six sentences. If the listener achieved four or more keywords correct on the first presentation of a given sentence, an additional sentence was added to the run. Each listener completed three runs for each masker condition and modality (audio-alone or AV) for a total of 18 runs. All of the trial-by-trial data were included in the data analysis, including sentences where the track was terminated at +15 dB with fewer than four keywords correctly identified, or where four or more keywords were correctly identified on the first trial. Each listener received at least 30 min of practice before data collection began. The practice consisted of completing several experimental runs for the different masker conditions and modalities, using different IEEE target sentences than those used in the actual experiment.

B. Results

Psychometric functions (Fig. 3) were derived across a range of SNRs for each condition. The total proportion of keywords correct for a given SNR was calculated across all of the sentences presented in the three runs for each condition. In deriving the performance functions from the trial-by-trial data in this way, a rolloff in performance was observed at the highest SNRs (not shown). The nature of the adaptive track was such that the highest SNRs for a given condition would have been tested for only those sentences that a listener was having particular difficulty identifying. Thus, the performance estimate at higher SNRs is likely to be biased below the actual performance level for the entire sentence set. (This was not an issue at the lowest SNRs because the starting SNR for each upward moving track was randomized.) To offset this effect, whenever a listener obtained all five keywords correct for a particular sentence, it was assumed that the listener would also have obtained five key-

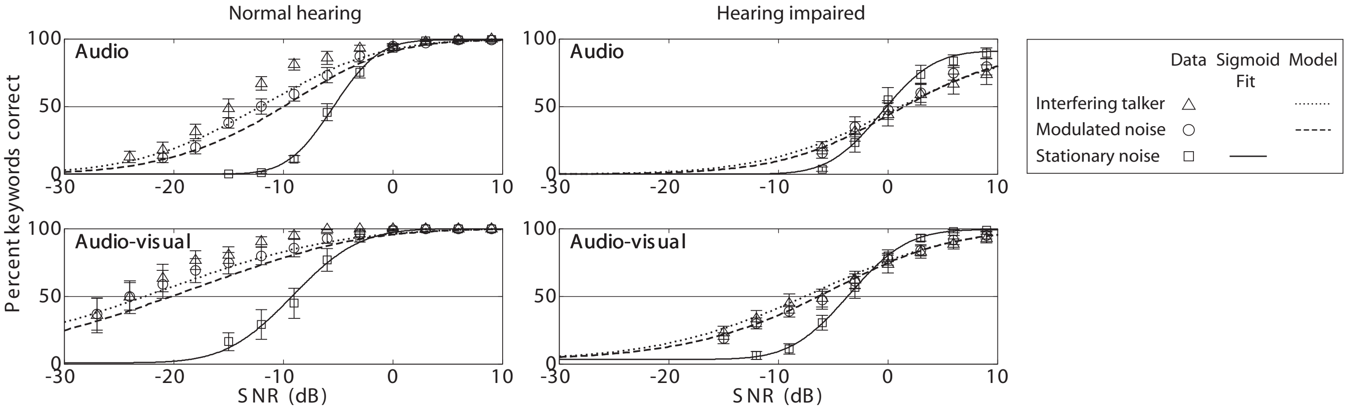


FIG. 3. Mean performance in identifying sentence keywords for target speech presented in three types of masker as a function of SNR. Symbols indicate the experimental data, with error bars indicating ± one standard error across listeners. Solid curves indicate sigmoid fits to the stationary-noise data. Dashed and dotted curves indicate ESII predictions for the modulated-noise and interfering-talker conditions, respectively, calculated using the SNR-dependent fluctuating-masker IIFs depicted in Fig. 8.

words correct at each higher SNR (up to a SNR of +15 dB in 3-dB steps), although these SNRs were not actually presented to the listener for that sentence.

Mean performance functions across the listeners in each group are plotted in Fig. 3. Results for the NH and HI listeners are shown in the left and right panels, respectively. Results for the audio-alone and AV modalities are shown in the upper and lower panels, respectively. Because of the adaptive nature of the SNR selection, some SNR values were tested only a very small number of times for a given listener and condition. For each condition and listener group, performance is plotted only for those SNRs where each listener in the group was presented with at least three sentences (15 words). Error bars indicate ± one standard error across listeners. The horizontal solid line denotes the 50% correct performance level. The other curves represent modeling results, discussed later in Sec. III.

The audio-alone data (top panels) are generally consistent with previous results (e.g., Festen and Plomp, 1990; Peters *et al.*, 1998; George *et al.*, 2006). NH listeners benefited substantially from masker fluctuations (the fluctuating-masker data fall to the left of the stationary-noise curve), whereas HI listeners showed less FMB (the fluctuating and stationary functions are closer together and largely overlapping). For the AV conditions, NH listeners still received more benefit from masker fluctuations than the HI listeners, although in this case the HI listeners also showed some FMB for low SNRs. While visual cues provided a benefit to speech intelligibility in all conditions (the performance functions cross the 50% correct line at a lower SNR for the AV conditions in all cases), the benefit of visual cues was larger for the fluctuating maskers than the stationary-noise masker, thereby increasing the FMB.

Most studies investigating speech intelligibility in fluctuating maskers have quantified the FMB as the difference in SRT between the stationary-noise and a fluctuating-masker condition. The SRT was derived for each listener and condition by fitting a sigmoid curve to the performance-SNR function and estimating the SNR required for 50% keywords-correct performance. (The solid curves in Fig. 3 denote the sigmoid curve fit to the stationary-noise data, but the sigmoid fits to the modulated-noise and interfering-talker data are not shown. The dashed and dotted curves represent the results of a model simulation described in Sec. III B.) The resulting mean SRTs across listeners are plotted for each masker and modality in the upper panel of Fig. 4, with more negative SRTs indicating better speech intelligibility. The mean FMBs (the difference between the SRT for each fluctuating-masker and the stationary-noise SRT) are shown in the lower panel of Fig. 4. Here, more positive values indicate that listeners received more benefit from masker fluctuations. Error bars in both panels represent ± one standard error across listeners.

FMB results (Fig. 4, lower panel) were subjected to a repeated-measures analysis of variance with two within-subject factors (fluctuating-masker type and modality) and

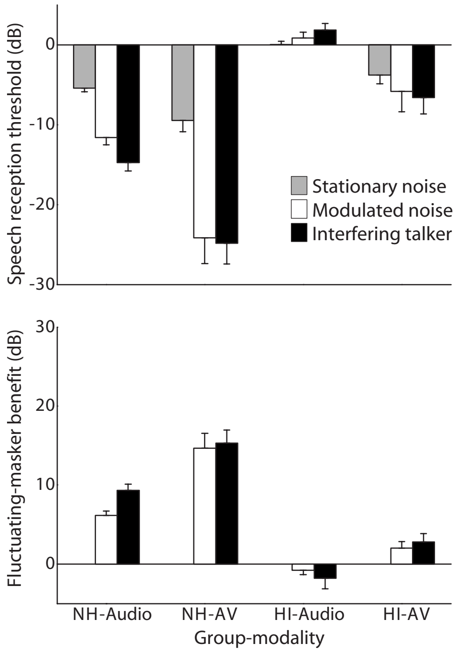


FIG. 4. (Upper panel) Mean SRTs are plotted for each listener group, modality, and masker. (Lower panel) The mean FMB, defined as the difference between the stationary and fluctuating-masker SRTs, is plotted for each listener group, modality, and fluctuating masker. Error bars indicate ± one standard error across listeners.

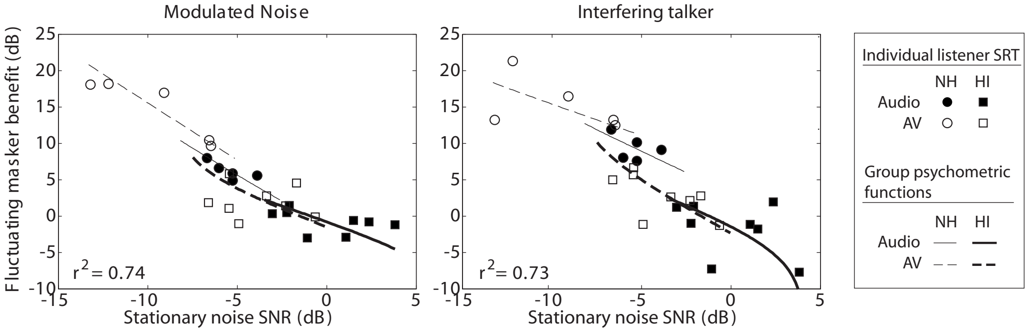


FIG. 5. Experimental results for individual listeners (symbols), showing an inverse relationship between the magnitude of the FMB and the stationary noise SRT across listener groups and audio-alone and AV presentation modalities. Individual curves indicate a similar inverse relationship across a range of stationary-noise SNRs within each listener group and modality. The vertical separation between pairs of curves indicates FMB differences that remain once baseline stationary-noise SNR differences were controlled.

one between-subjects factor (hearing status). The FMB was generally greater for the AV conditions than for the audio-alone conditions. There was a significant main effect of modality $[F(1,11)=121,\ p<0.0005]$, with *post-hoc* paired-sample *t*-tests confirming an effect of visual cues for all combinations of fluctuating-masker type and hearing status $(p<0.01)$. The effect of visual cues on the magnitude of the FMB was generally greater for NH listeners than for HI listeners, with a significant interaction between modality and hearing status $[F(1,11)=12.6,\ p<0.01]$. *Post-hoc t*-tests indicated that this was only the case for the modulated-noise masker $(p<0.05)$, with no significant difference between listener groups in the effect of visual cues on the FMB for the speech masker $(p=0.26)$.

The magnitude of the FMB was greater for NH listeners than for HI listeners, consistent with previous results. There was a significant main effect of hearing status $[F(1,11)=63.9,\ p<0.0005]$, with *post-hoc* independent-sample *t*-tests confirming an effect of hearing status for all combinations of fluctuating masker and modality $(p<0.0005)$. The magnitude of the FMB was not generally different between the interfering-talker and modulated-noise maskers, with no significant main effect of masker type $(p=0.17)$, or two-way interactions between masker type and modality $(p=0.70)$ or hearing status $(p=0.12)$. However, there was a significant three-way interaction between masker type, modality, and hearing status $[F(1,11)=5.9,\ p<0.05]$. *Post-hoc t*-tests showed a significantly larger FMB for an interfering talker than for modulated noise for the audio-alone stimuli presented to NH listeners $(p<0.05)$. This result is consistent with previous findings indicating that an opposite-gender interferer provides more masking release than a speech-modulated noise (e.g., Festen and Plomp, 1990; Qin and Oxenham, 2003). This could be due to the existence of spectral dips for the interfering talker which are available in addition to the temporal dips that are present for both maskers.

As discussed in the Introduction, the magnitude of the FMB has been shown to be dependent on the stationary-noise SNR (Oxenham and Simonson, 2009). The upper panel of Fig. 4 indicates that the stationary-noise SRT occurred at a lower SNR for the AV than for the audio-alone conditions and at a lower SNR for NH listeners than for HI listeners. The magnitude of the FMB (Fig. 4, lower panel) therefore might have been influenced by these stationary-noise SRT differences. The individual data were examined to investigate the relationship between the magnitude of the FMB and the stationary-noise SRT across listeners and modalities. Figure 5 shows the amount of FMB (in dB) plotted as a function of the stationary-noise SRT for individual NH (circles) and HI listeners (squares) in the audio-alone (filled symbols) and AV (open symbols) conditions. The left and right panels of Fig. 5 represent the modulated-noise and interfering-talker conditions, respectively. For both fluctuating-masker conditions, the FMB was negatively correlated with the stationary SRT, with statistically significant $(p<0.0001)$ r^2 values exceeding 0.7 in each case.[1] This is consistent with the idea that the differences in FMB between groups and modalities (Fig. 4) may be related to differences in the SNR needed to yield 50% performance in the stationary-noise conditions.

If the differences in the FMB between listener groups and modalities mainly reflect differences in the stationary-noise SRT, one would expect differences in the FMB to be reduced when listener groups or modality conditions are compared at the same baseline stationary-noise SNR. To investigate this possibility, the magnitude of the FMB was estimated across a range of stationary-noise SNRs for all combinations of listener group, modality, and fluctuating-masker type. Sigmoid functions were fitted to the performance-intensity functions for each masker type in Fig. 3 (fitted functions are shown only for the stationary case). The relationship between SNR and the FMB was estimated by taking the horizontal distance (in dB) between these curves for performance levels ranging from 20% to 80% keywords correct. The resulting benefit-vs-SNR functions are plotted as solid (audio) and dashed (AV) lines in Fig. 5, with the thick and thin lines representing HI and NH listeners, respectively.

For both masker types, the variation in the group-mean functions across stationary SNR (solid and dashed curves) generally resembles the variation across stationary SRTs for individual listeners (circles and squares). This suggests that most of the differences in FMB between groups and modalities can be attributed to differences in the stationary SNR. The lower panel of Fig. 4 indicates that HI listeners experienced a reduction in the FMB (relative to NH listeners) of 7 dB (audio-alone) to 12 dB (AV) for the modulated-noise condition (white bars), and a reduction of 11 dB (audio-alone) to 12 dB (AV) for the interfering-talker case (black bars). The vertical separation between pairs of curves in Fig. 5 provides an estimate of the portion of these FMB differences that cannot be explained by differences in the baseline stationary-noise SRT. The SNR ranges (horizontal dimension) for the four curves in each panel of Fig. 5 are largely non-overlapping due to differences between listener groups and modalities in the SNRs that yielded performance in the 20%–80% correct range. However, comparisons between curves can be made for a small range of stationary-noise SNRs (approximately −8–0 dB) where there is overlap in the SNR ranges between listener groups and modalities. For the modulated-noise condition (left panel), the vertical separation between listener groups ranged from 1 dB (audio-alone HI vs NH, solid curves) to 4 dB (AV HI vs NH, dashed curves). This suggests that 1–4 dB of the FMB difference between NH and HI listeners can be attributed to a reduction in the ability to make use of dips in the fluctuating-masker level due to factors such as audibility, spectral and temporal resolution, or source-segregation deficits. The remaining 6 dB (audio-alone) to 8 dB (AV) of the total 7–12-dB reduction in FMB is attributable to an increase in the SNR required to achieve the 50% correct performance level. The suprathreshold deficits described above might be indirectly responsible for this portion of the reduced FMB, but only via a general negative impact on speech intelligibility resulting in the increased-SNR requirement.

For the interfering-talker condition (Fig. 5, right panel), the trends are largely similar to those for the modulated-noise condition, except that the vertical separation between the NH and corresponding HI curves is somewhat larger than for the modulated-noise case. Comparing the NH (thin lines) and HI curves (thick lines) for stationary-noise SNRs of −8 to −3 dB, there was about a 5-dB separation between listener groups in the interfering-talker condition (right panel), but only a 1-dB (audio-alone) to 4-dB difference (AV) in the modulated-noise case (left panel). This suggests that after stationary-noise SNR differences were controlled, HI listeners had more residual difficulty (relative to NH listeners) taking advantage of the momentary dips in the masker level for the interfering-talker than for the modulated-noise condition. As for the modulated-noise case, SNR differences accounted for the remaining 6 dB (audio-alone) or 7 dB of the total 11–12-dB reduction in FMB for HI listeners in the interfering-talker condition.

With the addition of visual cues, NH listeners showed a larger increase in the FMB (6–8 dB) than HI listeners (3–5 dB) relative to the audio-alone case for both fluctuating maskers (compare the audio-alone and AV conditions in the lower panel of Fig. 4). In Fig. 5, the vertical separation between curves for the audio-alone (solid curves) and AV conditions (dashed curves) indicates the portion of this additional FMB under AV conditions that cannot be explained by the decrease in the stationary-noise SRT. This portion of the increase in FMB may reflect the influence of visual cues for source-segregation. For both fluctuating maskers, the vertical separation between curves for audio-alone and AV modalities was about 2 dB for the NH listeners (thin curves), but only 0–1 dB for the HI listeners (thick curves). This suggests that visual source-segregation cues provided only limited benefit to speech intelligibility in these conditions, perhaps because the target and masker were qualitatively different thereby yielding little informational masking even in the audio-alone case. Most of the increase in FMB for AV stimuli (the remaining 4–6 dB for NH or 2–4 dB for HI listeners) was attributable to a decrease in the stationary-noise SRT as a result of the additional speech information provided by visual cues.

C. Discussion

When comparing performance at the 50% keywords-correct performance level (Fig. 4), the magnitude of the FMB was larger for NH listeners than HI listeners, and larger under AV than audio-alone conditions. On the surface, the reduced FMB in HI listeners, also observed in previous studies, is consistent with the idea that HI listeners may have a reduced ability to make use of momentary dips in the masker level to extract additional speech information. Similarly, the enhanced FMB under AV conditions is consistent with the idea that the visual stimulus provided source-segregation cues in the fluctuating-masker conditions, in addition to the phonetic speech information provided in all masker conditions.

However, strong correlations (Fig. 5) between the magnitude of the FMB and the stationary SRT suggest that these apparent effects of hearing loss and modality on the FMB may be confounded by stationary-noise SRT differences across listeners and conditions. The relationship between the FMB and the stationary-noise SRT observed across listeners is consistent with previous audio-alone studies that have tested large numbers of NH and HI listeners (Versfeld and Dreschler, 2002; George *et al.*, 2006; Wilson *et al.*, 2007) and NH listeners with a simulated hearing loss (George *et al.*, 2006). Like the current study (Fig. 3), listeners in these previous studies with a stationary-noise SRT near 0 dB received no benefit from fluctuating maskers, and the amount of benefit systematically increased as the stationary-noise SRT became more negative. The current results extend those findings to AV conditions, demonstrating that improvement in the stationary-noise SRT with the addition of visual cues yields a larger FMB for NH and HI listeners. A similar relationship was observed in the performance-SNR functions for each subject group and modality. In Fig. 3, the slopes of the performance functions were generally steeper for the stationary noise than for each of the fluctuating maskers, such that listeners received more benefit (in dB) at more negative SNRs. This trend is consistent with that observed in NH listeners for broadband (Festen and Plomp, 1990) and for filtered speech (Oxenham and Simonson, 2009).

When the FMB was compared between listener groups and modalities at the same stationary-noise SNR (curves in Fig. 5) rather than at the stationary-noise SRT that varied across listeners and conditions, FMB differences were greatly reduced. This suggests that HI listeners are less impaired in their ability to listen in the dips than would be concluded based on FMB estimates calculated at the SRT. Still, some effects of hearing loss and visual cues on the

FMB remained once SNR differences were controlled. The remaining 1–5-dB decrease in the FMB for HI listeners may be attributable to a reduced ability to listen in the dips due to psychoacoustic factors discussed in the Introduction such as reduced audibility, spectral and temporal resolution, or cues for source-segregation. For example, reduced audibility of the target speech for HI listeners (Fig. 2) may have rendered inaudible portions of the speech dynamic range that would have otherwise been unmasked by momentary dips in the masker level. Interestingly, the effect of hearing loss on the FMB was larger for the interfering talker than for the modulated noise. This could be because the speech masker contained both spectral and temporal fluctuations, whereas the modulated-noise background contained mainly temporal fluctuations. Limited frequency resolution in HI listeners (e.g., Glasberg and Moore, 1986) might be expected to have a greater negative impact in the interfering-talker condition containing spectral fluctuations than the modulated-noise case. Another possibility is that informational masking (e.g., Brungart, 2001) may play a larger role in the interfering-talker case, with HI listeners being more susceptible to target-masker confusions than NH listeners.

In contrast to this informational-masking interpretation, the comparison of the FMB between audio-alone and AV conditions does not support the notion that HI listeners lack speech-source segregation cues relative to NH listeners in the audio-alone conditions. If that were the case, and both groups of listeners were able to utilize the visual information as a cue for source-segregation, one might expect that the availability of visual cues would provide more benefit to HI listeners, who are missing audio segregation cues, than to NH listeners, who are presumably already able to successfully segregate target and masker in the audio-alone conditions. This hypothesis was based on previous findings for NH listeners, whereby auditory spatial cues (Freyman *et al.*, 1999; 2001) and visual cues (Helfer and Freyman, 2005) provided much more benefit in situations with substantial target-masker confusion than in situations where there was not, such as in stationary noise. In fact, the opposite occurred: the increase in the FMB with the introduction of visual cues was slightly less for HI than for NH listeners, even when compared at the same stationary-noise SNR. One possible explanation for this result is that HI listeners could be impaired in their ability to use visual cues for source-segregation. The use of visual speech information as a segregation cue likely requires the visually perceived motion of the lips, tongue, and jaw to be integrated with dynamic changes in the target audio signal. If the internal representation of the signal is distorted due to hearing loss, the strength of association between visually perceived articulation and acoustic dynamics may be weakened, reducing the FMB.

The strong relationship between the stationary-noise SNR and the magnitude of the FMB across listeners and modalities (Fig. 5) suggests that a substantial portion (6–8 dB) of the total 7–12-dB reduction in benefit for HI listeners may be accounted for by the fact that these listeners require a more favorable SNR in the stationary-noise case. Suprathreshold deficits may underlie the impaired speech intelligibility that leads to the more favorable SNRs required for HI listeners, thereby indirectly reducing the FMB. However, these results suggest that such deficits may have less of a direct impact on the ability to listen in the dips of a fluctuating masker than is suggested by previous studies where appropriate controls for baseline SNR differences were not considered. Section III discusses a possible theoretical basis for the relationship between SNR and the FMB in terms of a level-distributed model of speech intelligibility based on the extended speech-intelligibility index (ESII) of Rhebergen and Versfeld (2005) and Rhebergen *et al.* (2006).

III. MODELING THE FLUCTUATING-MASKER BENEFIT

A. A theoretical basis for a SNR-dependent benefit

The SII (ANSI, 1997) provides a method of estimating speech intelligibility for a given set of speech materials, speech spectrum, stationary-noise masker spectrum, and audiometric thresholds. Rhebergen and Versfeld (2005) modified the original SII to allow estimates of speech intelligibility in fluctuating maskers. Speech intelligibility is computed in successive short time frames, then averaged across time frames to estimate the average intelligibility for a fluctuating masker. The resulting ESII was successful in predicting the FMB observed for NH listeners across a variety of fluctuating maskers, including speech-modulated noise, interrupted noise, sinusoidally intensity-modulated noise (Rhebergen and Versfeld, 2005), and real-life maskers such as animal and machine sounds (Rhebergen *et al.*, 2008). The inclusion of a forward-masking function (Ludvigsen, 1985) improved the accuracy of speech-intelligibility predictions as a function of the modulation rate of the fluctuating masker (Rhebergen *et al.*, 2006). Notably for the current results, the ESII predicts a general increase in the FMB as the stationary-noise SNR decreases (Rhebergen and Versfeld, 2005).

Freyman *et al.* (2008) suggested that the SNR dependence of the FMB could reflect the offsetting effects of speech information masked and unmasked by peaks and valleys in the level of a fluctuating masker. Because speech information is distributed across a range of levels, the relative magnitudes of these offsetting effects will depend on the shape of the distribution of speech information across the dynamic range [i.e., the intensity importance function (IIF); Boothroyd, 1990; Studebaker and Sherbecoe, 2002] and on the SNR operating point within this range. An illustration of this concept is presented in Fig. 6. Note that this is a simplified schematic to illustrate the possible relationship between the distribution of speech information and the range of fluctuating-masker levels, and is not intended to be quantitatively accurate.

An IIF based on the measurements of Studebaker and Sherbecoe (2002) is plotted in both panels of Fig. 6. Circles denote the IIF derived from intelligibility estimates for narrowband single-syllable words presented in stationary noise as a function of SNR in 4-dB increments. The continuous function (solid curve) was estimated via piecewise cubic spline interpolation of the discrete IIF.[2] In each panel of Fig. 6, the rms level of the stationary noise is represented by the position along the horizontal axis of the vertical dashed line.

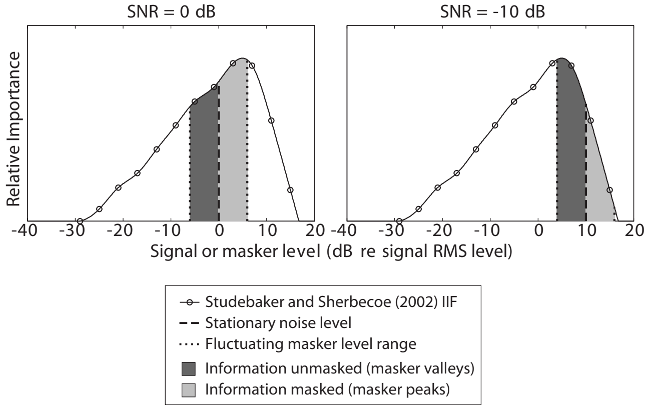


FIG. 6. A schematic drawing describing how a SNR-dependent FMB might arise. The functions in each panel represent an interpolated version of the IIF derived by Studebaker and Sherbecoe (2002), indicating the distribution of speech information across the dynamic range. Fluctuating maskers contain both peaks and valleys, which mask (light shaded area) and unmask (dark shaded area) speech information relative to the stationary-noise condition. For an average SNR of 0 dB (left panel), the masking and unmasking of speech information are approximately equal, yielding no net FMB. For an average SNR of −10 dB (right panel) the amount of speech information unmasked in the masker valleys is greater than the information masked by the masker peaks, yielding a net FMB benefit.

The upper and lower extremes of the range of levels for a hypothetical fluctuating masker are depicted by the two vertical dotted lines (at ±6 dB re the rms masker level). The shaded areas represent the offsetting quantities of speech information released by masker valleys (dark shading) and masked by masker peaks (light shading) relative to the stationary-noise case. The two panels depict average SNRs of 0 dB (left panel) and −10 dB (right panel), with the masker levels 10 dB higher in the right panel. At a SNR of 0 dB (left panel), the masking and unmasking effects are roughly equal in magnitude, thereby producing no net benefit of the fluctuating masker. At a negative SNR (right panel), the masker peaks mask a portion of the dynamic range (light shaded area) with less importance than the portion unmasked during the masker valleys (dark shaded area), yielding a net benefit from masker fluctuations.

B. The extended speech-intelligibility index

The ESII (Rhebergen *et al.*, 2006) was tested to assess its ability to account for variation in the FMB across NH and HI listeners and audio-alone and AV conditions (Fig. 5). The idea was to determine whether the effects observed (more FMB for NH than HI listeners, and more FMB under AV than audio-alone conditions) could be accounted for in terms of differences in the stationary-noise SRT. The ESII calculation methodology was described by Rhebergen *et al.* (2006). The simulation used 5-s samples of the various masker signals that were presented to listeners. First, excitation patterns as a function of time were generated for each masker sample by (1) filtering the sample into 21 CBs, (2) introducing an interpolated audiometric-threshold function modeled as a minimum excitation level in each band, and (3) applying a forward-masking function (Ludvigsen, 1985) to the output of each filter. The masker excitation pattern for each 4-ms time frame was calculated to be the greater of (a) the absolute hearing threshold or (b) the temporally smeared masker's rms level. Next, the level of the target speech in each CB was estimated in the same manner, substituting a speech-spectrum shaped noise in place of the actual speech signal. Then, the SII was computed for each 4-ms time frame and CB based on the local SNR by integrating the audible portion of one of the three versions of the IIF discussed below. Also included in the modeling procedure was a level distortion factor, dependent on the level of the target speech in that CB, that reduced the SII for speech presented at a high level (ANSI, 1997).[3] Finally, the ESII was calculated by averaging the SII in each CB over time and computing a frequency-weighted average of the SIIs across CBs. The FMB estimate was derived by comparing the SNR in a fluctuating-masker condition needed to achieve the same ESII as for the stationary-noise condition at a given SNR. Because the FMB estimate was not derived from a percent-correct score (which would have required a transformation specific to the IEEE sentence set), the FMB predictions are also likely to be independent of the particular stimulus set chosen and therefore applicable to other speech materials.

Because the relative amount of information masked and unmasked by the masker peaks and valleys is determined by the shape of the distribution of speech information across the dynamic range (Fig. 6), the choice of the IIF affects the ESII predictions. ESII-based FMB predictions generated using three possible IIF alternatives are shown in Fig. 7. Because differences in audibility and high-level distortion could also have contributed to the difference between FMB observed

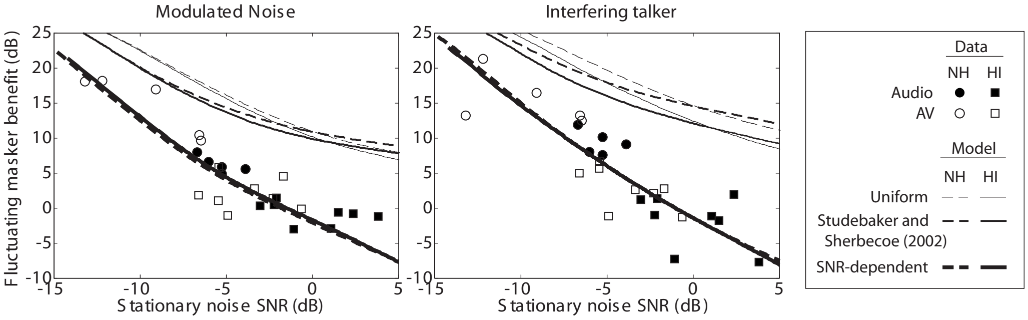


FIG. 7. ESII model FMB predictions as a function of the stationary-noise SRT, with audiometric thresholds and signal levels set to represent the average NH (dashed curves) and HI (solid curves) listener, with three assumed IIFs: (thin curves) a uniform distribution (ANSI, 1997; Rhebergen *et al.*, 2006); (medium curves) the Studebaker and Sherbecoe (2002) IIF estimate; and (thick curves) a SNR-dependent fluctuating-masker IIF fit to the experimental data. Data for individual listeners (symbols) are replotted from Fig. 5.

for NH and HI listeners, separate predictions were generated for each listener group based on the mean audiometric thresholds shown in Fig. 1 and the respective signal levels used in the experiment. Differences in the ESII predictions between average NH (Fig. 7, dashed curves) and HI (solid curves) simulations reflect differences in the audibility of the speech dynamic range for frequencies above 2 kHz (Fig. 2).

The first set of FMB estimates (thin curves in Fig. 7) was generated using the assumption (ANSI, 1997; Rhebergen *et al.*, 2006) that speech information is uniformly distributed between ±15 dB re the long-term rms level. The second set of estimates (medium-thickness curves in Fig. 7) implemented the function (Fig. 6) derived by interpolating the Studebaker–Sherbecoe (2002) IIF estimates, as described in Sec. III A. For both IIF alternatives, the ESII estimates failed to account quantitatively for the experimental results, yielding poor fits to the variation in FMB observed across listeners and modalities. Nevertheless, the ESII predictions were qualitatively consistent with the trend observed in the data across listeners and modalities whereby the FMB decreased with increasing stationary-noise SRTs. The ESII also predicted that for a given stationary-noise SNR, NH listeners should show more FMB than HI listeners (compare thin vs thick dotted or dashed curves), as a result of the audibility differences between the two groups. This effect is qualitatively consistent with the observation (curves in Fig. 5) that HI listeners received 1–5 dB less benefit than NH listeners at the same SNR, depending on the masker type and modality.

The third IIF alternative assumed that the relative importance for different portions of the speech dynamic range is different for a fluctuating masker than for a stationary-noise masker. As implemented in the ESII, the IIF represents the contribution to overall intelligibility yielded by each portion of the dynamic range that is audible above absolute threshold and is not masked by the masker. Because the uniform and Studebaker-Sherbecoe (2002) IIFs predicted a larger FMB than was observed in the data, we reasoned that the lowest levels of the dynamic range might not contribute to intelligibility for speech presented in a fluctuating masker, even if they are rendered audible by dips in the masker level. One possible reason for this is that a minimum duration might be required for an audible portion of a signal to convey speech information. Lorenzi *et al.* (2006b) showed that the modulation-rate dependence of the FMB for consonant identification varied across consonant feature (voicing, manner, and place). This suggests that in addition to psychophysical forward-masking effects already included in the ESII (Rhebergen *et al.*, 2006), durational constraints for the successful relaying of speech information may also limit the FMB. We further reasoned that this IIF limitation would only apply to the fluctuating-masker condition and not the stationary-noise case. Because the level of the stationary-noise masker is approximately constant over time, any portion of the dynamic range audible above the noise would remain so for a sufficient duration to fully contribute to intelligibility. In contrast, very short segments of speech would be rendered audible at the base of a dip of a fluctuating masker and would therefore be subject to the putative durational constraint that limits their contribution to intelligibility.

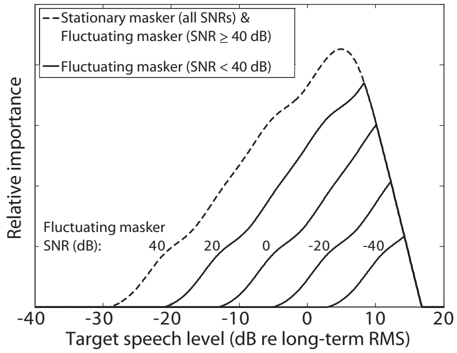


FIG. 8. The Studebaker and Sherbecoe (2002) stationary-noise IIF (dashed curve) and a series of SNR-dependent fluctuating-masker IIFs (solid curves) used in the ESII simulations to produce the FMB predictions of Fig. 7.

For the stationary-noise case, the IIF was assumed to be distributed according to the Studebaker and Sherbecoe (2002) estimates (dashed curve in Fig. 8) that were derived from stationary-noise speech-intelligibility measurements. Hypothetical fluctuating-masker IIFs were generated by reducing the assumed contribution that low-level portions of the dynamic range make to overall speech intelligibility. For a fluctuating masker, the shortest audible time segments will occur near the base of the valleys in the masker level. Therefore, the contributions of the low-intensity portions of the speech signal would be affected most by these putative durational constraints. By this logic, the shape of the IIF for speech presented in a fluctuating masker should vary as a function of SNR. At very high SNRs, only the very lowest portions of the dynamic range would be affected by the duration limitation, and the IIF should resemble that for the stationary-noise case. As SNR decreases, increasingly higher signals levels would be subject to durational constraints, thereby narrowing the effective dynamic range.

A SNR-dependent fluctuating-masker IIF was implemented by shifting the lower tail of the Studebaker–Sherbecoe (2002) IIF for stationary noise in the horizontal (dB) dimension, but not allowing the IIF value to exceed the stationary-noise IIF value at any level within the dynamic range (Fig. 8). The modified functions were not renormalized to sum to one because the IIF modification was meant to reduce the contribution of low-level portions of the dynamic range without increasing the contributions of higher-level portions. Two free parameters described the shift of the fluctuating-masker IIF as a function of SNR. The first parameter described the upper SNR limit at which the fluctuating-masker IIF was equal to the stationary-noise IIF (best-fitting value=40 dB). The second parameter defined the ratio between the magnitude of the IIF shift and the change in SNR (best-fitting value=0.4 dB/dB). Fluctuating-masker IIFs associated with these best-fitting parameters are shown as solid curves in Fig. 8 for a range of SNRs. This set of fluctuating-masker IIFs yielded good quantitative fits (thick curves) to the NH and HI individual-listener FMB data (Fig. 7). Note that in contrast to the predictions based on the uniform (thin curves) and Studebaker–Sherbecoe (2002) IIFs (medium curves), the NH (thick solid curves) and HI (thick dashed curves) predictions are largely similar for the SNR-dependent family of IIFs fit to the experimental data (solid lines). This is because differences in audibility between the two listener groups occurred for signal levels that would contribute little to overall intelligibility for the fluctuating-masker IIFs depicted in Fig. 8.

To further investigate the extent to which SNR is the main predictor of the FMB, the ESII predictions were applied to the psychometric functions for each subject group and modality (Fig. 3). In each panel of Fig. 3, the stationary-noise data were fitted with a sigmoid function (solid lines). Predictions for the fluctuating-masker conditions were generated by horizontally shifting the stationary-noise curve at each percent-correct level by an amount equal to the predicted FMB (in dB) at the stationary-noise SNR associated with that performance level. The model generally produced a reasonable fit to the fluctuating-masker psychometric functions for all groups, modalities, and fluctuating-masker conditions. Specifically, for all listener groups and conditions, the magnitude of the predicted FMB was largest at very low SNRs and decreased to zero at a SNR of about 0 dB.

Additionally, the model accounted for the crossover effect observed in the data, whereby fluctuating maskers actually impaired performance for positive SNRs. (This effect was clearly observed in the HI data, but not for the NH listeners who performed at ceiling levels for positive SNRs.) The model was able to account for this effect because the masker peaks tend to mask more speech information than is uncovered by the masker valleys. This crossover effect has also been observed in studies that have distorted the speech signal to allow below-ceiling performance at positive SNRs in NH listeners (e.g., Qin and Oxenham, 2003; Oxenham and Simonson, 2009), although this effect has been previously interpreted in terms of modulation interference (e.g., Kwon and Turner, 2001) or a limited ability to segregate sources. However, positive stationary-noise SNRs are not always associated with a fluctuating-masker detriment. For example, in a recent study by Stuart (2008), children and adults received a benefit from interrupted relative to stationary noise for sentence materials presented at a positive SNR. The magnitude of the FMB is likely to depend on the particular statistics of the masking signal. The interrupted noise employed by Stuart (2008) allows full audibility of the target signal during the quiet periods. It may be that the benefit to speech intelligibility yielded by these quiet periods outweighed any excessive masking of the target during the on cycles of the interrupted noise, even at positive SNRs.

Overall, this analysis demonstrates that by making reasonable assumptions about a fluctuating-masker IIF that depends on the SNR, many of the differences in FMB across listeners, hearing status, and modality can be explained based on differences in the stationary-noise SRT due to overall poorer speech intelligibility, rather than a specific deficit in the ability to listen in the dips of a fluctuating masker. IIF estimates for speech presented in fluctuating-masker backgrounds have not been established in the literature, nor is it known whether such IIFs would differ from the stationary-noise case or as a function of SNR. Further work is needed to determine the shape of the IIF in fluctuating maskers and as a function of SNR and to determine whether this modeling approach can account for other examples in the literature of reduced FMB in HI and simulated-HI listeners.

IV. GENERAL DISCUSSION

The experimental results (Sec. II) and the results of the modeling simulation (Sec. III) indicate that a large proportion of the total difference in FMB between NH and HI listeners and between audio-alone and AV modalities in the current study may stem from differences in the SNR required to achieve comparable levels of performance in stationary noise. The same explanation may hold for other results in the literature that have shown reduced benefit from fluctuating maskers in HI listeners (e.g., Festen and Plomp, 1990; Versfeld and Dreschler, 2002; George *et al.*, 2006; Lorenzi *et al.*, 2006b; Wilson *et al.*, 2007) and NH listeners presented with speech distorted by spectral smearing (ter Keurs *et al.*, 1993; Baer and Moore, 1994) or removal of temporal fine structure (Qin and Oxenham, 2003; Hopkins *et al.*, 2008). In all of these cases, those listeners or conditions that yielded the least benefit from fluctuating maskers also had stationary-noise SRTs at the highest (least negative) SNRs. In particular, Versfeld and Dreschler (2002), George *et al.* (2006), and Wilson *et al.* (2007) measured SRTs in stationary and fluctuating maskers for a large number of listeners and showed a strong relationship between the FMB and the stationary-noise SRT across listeners. For studies that estimated the magnitude of the FMB but did not report the baseline stationary-noise SRT (e.g., Peters *et al.*, 1998; Lorenzi *et al.*, 2006a), the possible confounding role of stationary-noise SRT differences is more difficult to quantify.

Nevertheless, it is important to note that stationary-noise SRT differences do not explain all the reduced FMB observed for HI listeners in the literature. For example, Jin and Nelson (2006) found substantial intelligibility differences between NH and HI listeners for speech presented in interrupted noise, even though there were no measurable differences between NH and HI listeners in the stationary SRT. Audibility differences between the two groups might be more likely to reduce the ability to listen in the dips for interrupted noise than for the speech-modulated noise or speech maskers used in the current study because signals are effectively presented in quiet during the off cycle of the interrupted noise. In any case, the possible confounding influence of SNR differences on the FMB estimate strongly suggests that these differences should be controlled for in studies examining the possible role of audibility and suprathreshold acuity on the ability to make use of dips in the level of a fluctuating masker.

That FMB differences between NH and HI listeners were reduced by 6–8 dB when performance was compared at similar SNRs suggests that the HI listeners may not be as impaired in the ability to listen in the dips of a fluctuating masker as previously thought. Instead, the generally poor performance exhibited by HI listeners in fluctuating noise might be ascribed to a general speech-intelligibility deficit.

As a result of reduced audibility or suprathreshold deficits, HI listeners are forced to listen at more favorable SNRs to achieve a reasonable level of performance. Dips in the level of a fluctuating masker yield less of a benefit to speech intelligibility at these higher SNRs, regardless of hearing status. Nevertheless, a reduced ability to listen in the dips as a direct result of the psychophysical impairments associated with hearing loss could underlie the remaining 1–5-dB reduction in FMB observed in HI listeners in the current study. Similarly, Takahashi and Bacon (1992) showed that older listeners received less benefit from fluctuating maskers than young listeners even when performance was compared at the same SNR. Psychophysical deficits associated with hearing loss may also play a more significant role in listening situations different from those tested here. For example, deficits in temporal resolution may become an important factor for maskers with higher modulation rates (Dubno *et al.*, 2003; George *et al.*, 2006) than those associated with the speech and speech-modulated maskers employed in the current study. Furthermore, the ability to use temporal fine-structure cues may play an important role in providing pitch cues for target-masker segregation in situations involving more similar target and masker signals.

Although differences in SNR in the stationary-noise condition may underlie a substantial proportion of the FMB differences between NH and HI listeners, this explanation does not address the underlying causes for the poorer performance by HI listeners in stationary noise. Suprathreshold distortions such as impaired spectral (e.g., Tyler *et al.*, 1982; Glasberg and Moore, 1989) or temporal resolution (Dubno *et al.*, 2003) or a reduced ability to use temporal fine-structure information (Buss *et al.*, 2004; Lorenzi *et al.*, 2006a) may play a significant role in reducing speech intelligibility in noise, generally. If so, these factors could be indirectly implicated in the reduced FMB observed in HI listeners as a result of increasing the stationary-noise SNR required for these listeners to achieve a given performance level.

In the current study, fluctuating-masker and stationary-noise conditions were compared by equating their long-term rms levels, yielding fluctuating-masker levels both above and below that of the stationary noise. The FMB was therefore influenced by the offsetting effects of masker peaks and dips. In a situation where the maximum level of the fluctuating masker is equated with the rms level of the stationary noise (e.g., Dubno *et al.*, 2003), any benefit from dips in the masker level would not be offset by the detrimental effects of masker peaks. Although this experimental paradigm might reduce the confounding effects of SNR differences, the amount of speech information released by the masker valleys is still likely to be SNR-dependent due to variation in the importance of different portions of the speech dynamic range. This is illustrated in Fig. 6, where the amount of speech information unmasked during dips in the masker level (dark shaded areas) is larger for a SNR of 0 dB (left panel) than for a SNR of −10 dB (right panel).

The large differences observed between NH and HI listeners in understanding speech in a fluctuating background have led to the suggestion that such measures may help to differentiate between listeners that show relatively small differences in a stationary-noise background (Festen and Plomp, 1990). The present results suggest the possibility that a substantial portion of the fluctuating-masker deficits experienced by HI listeners and NH listeners listening to distorted speech can be predicted based on differences in the SNR at which performance is tested. If performance for a fluctuating masker can be accurately predicted based on stationary-noise performance, SNR, and the statistical characteristics of the masker, one could argue that it may not be worthwhile to test listeners using fluctuating maskers, as little new information regarding an individual would be garnered from such a test. On the other hand, estimates of speech-intelligibility performance in a fluctuating background and/or with visual cues may yield the greatest differences between individual listeners, and therefore be a more sensitive test to differentiate speech reception performance across listeners. Our results indicate that the range of SRT estimates across the NH and HI listeners was only 9.8 dB for the stationary audio-alone condition, but was 30.7 dB for the modulated-noise AV condition. By the same logic, the AV fluctuating-masker conditions may be expected to yield greater, and therefore more measurable, within-group differences across signal-processing algorithms aimed toward improving speech understanding for HI listeners. Thus, before ruling out a particular algorithm because it only yields a very small benefit to speech intelligibility in stationary noise (e.g., spectral enhancement, Simpson *et al.*, 1990), it may be worthwhile to determine whether the benefit would be larger under AV and/or fluctuating-masker conditions.

V. CONCLUSIONS

HI listeners showed 7–12 dB less FMB than NH listeners when performance was compared at the 50% correct performance level. This result, consistent with previous studies, has been interpreted as an indication that HI listeners have difficulty taking advantage of momentary dips in a fluctuating masker. With the introduction of visual cues, NH and HI listeners showed 3–8 dB more FMB than that obtained under audio-alone conditions. One interpretation of these results is that visual cues provided two sources of benefit in the fluctuating-masker conditions (segregation cues and phonetic information), but primarily phonetic information in the stationary-noise case.

However, when the FMB was compared at the same stationary-noise SNR, rather than the same speech reception performance level (50% keywords correct), FMB differences between audio-alone and AV modalities, and between the two listener groups, were reduced. This suggests that a substantial proportion of the FMB differences between listener groups and modalities was due to differences in the baseline stationary-noise SNR required to achieve the 50% correct performance level. Nevertheless, HI listeners still showed 1–5 dB of reduced FMB once SNR differences were controlled. This suggests that audibility, spectral or temporal resolution, or reduced cues for source-segregation may still have played a role in directly limiting the ability to listen in the dips of a fluctuating masker. Similarly, visual cues yielded an additional 1–2 dB of FMB over the audio-alone case for NH listeners, which could be attributable to source-segregation cues provided by the visual stimuli. The ESII (Rhebergen *et al.*, 2006), with the effective dynamic range of speech assumed to be different for fluctuating-masker conditions than for stationary noise, accounted for most of the variation in the FMB across individual listeners and modalities and across the performance-SNR function within each listener group and modality. Overall, the results of the experiment and model simulation highlight the importance of controlling for SNR differences between listener groups or stimulus conditions before drawing conclusions regarding the mechanisms underlying the FMB.

ACKNOWLEDGMENTS

This work was supported by a grant from the Oticon Foundation. We thank Van Summers, Chris Darwin, an anonymous reviewer, and the Associate Editor, Brian Moore, for their helpful comments on an earlier version of this manuscript, and Koen Rhebergen for providing the ESII code. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army, Department of Defense, or U.S. government.

[1] A positive correlation between the stationary-noise SRT and the FMB would be expected even if the stationary-noise and fluctuating-masker SRTs were uncorrelated because the FMB calculation (i.e., the difference between the stationary-noise SRT and the fluctuating-masker SRT) includes the stationary-noise SRT as the first term in the equation. However, in this case, the relationship between these two quantities would be in the opposite direction (i.e., the benefit *decreasing* with decreasing stationary-noise SRT) from that observed experimentally.

[2] The Studebaker–Sherbecoe (2002) IIF contained negative values for very high SNRs. This was likely due to a rollover effect (e.g., Fletcher, 1922; Summers and Molis, 2004), whereby intelligibility decreased with increasing SNR at very high signal levels (up to 91 dB SPL). In the current study, the signal level was held fixed and the noise level varied to produce the desired SNR, such that any negative impact of high stimulus level was likely to have been constant across SNR. Therefore, negative values in the Studebaker–Sherbecoe (2002) IIF were set to zero before interpolation.

[3] Although the level distortion factor reduced the absolute ESII values for the HI listeners (who were tested at a higher signal level than the NH listeners), it did not affect the model's predictions of the amount of FMB because the same factor was applied to each of the masker conditions.

Alcántara, J. I., and Moore, B. C. J. (**1995**). "The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency," J. Acoust. Soc. Am. **97**, 3813–3824.

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

ANSI (**2004**). *Specification for Audiometers, S3.6* (American National Standards Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Auer, E. T. Jr., and Bernstein, L. E. (**2007**). "Enhanced visual speech perception in individuals with early-onset hearing impairment," J. Speech Lang. Hear. Res. **50**, 1157–1165.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**2000**). "Speech perception without hearing," Percept. Psychophys. **62**, 233–252.

Bernstein, J. G. W., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.

Boothroyd, A. (**1990**). "Articulation index: Importance function in the intensity domain (A)," J. Acoust. Soc. Am. **88**, S31.

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. **43**, 647–677.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Carhart, R. C., and Tillman, T. W. (**1970**). "Interaction of competing speech signals with hearing losses," Arch. Otolaryngol. **91**, 273–279.

Darwin, C. J., and Hukin, R. W. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970–977.

Driver, J. (**1996**). "Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading," Nature (London) **381**, 66–68.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with impaired hearing," J. Acoust. Soc. Am. **113**, 2084–2094.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Fletcher, H. (**1922**). "The nature of speech and its interpretation," Bell Syst. Tech. J. **1**, 129–144.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2008**). "Spatial release from masking with noise-vocoded speech," J. Acoust. Soc. Am. **124**, 1627–1637.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Grant, K. W. (**2001**). "The effect of speechreading on masked detection thresholds for filtered speech," J. Acoust. Soc. Am. **109**, 2272–2275.

Grant, K. W., and Seitz, P. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415–2424.

Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Hygge, S., Rönnberg, J., Larsby, B., and Arlinger, S. (**1992**). "Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds," J. Speech Hear. Res. **35**, 208–215.

IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measures* (Institute of Electrical and Electronic Engineers, New York).

Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: The effects of temporal resolution," J. Acoust. Soc. Am. **119**, 3097–3108.

Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006a**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006b**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception," Int. J. Audiol. **45**, 487–495.

Ludvigsen, C. (**1985**). "Relations among some psychoacoustic parameters in normal and cochlearly impaired listeners," J. Acoust. Soc. Am. **78**, 1271–1280.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

Nilsson, M., Soli, S., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Oxenham, A. J., and Moore, B. C. J. (**1997**). "Modeling the effects of peripheral nonlinearity in normal and impaired hearing," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 273–288.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–53.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). "Relations among different measures of speech reception in subjects using a cochlear implant," J. Acoust. Soc. Am. **92**, 1869–1881.

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2008**). "Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing," Ear Hear. **29**, 169–175.

Simpson, A. M., Moore, B. C. J., and Glasberg, B. R. (**1990**). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," Acta Oto-Laryngol., Suppl. **469**, 101–107.

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

Stuart, A. (**2008**). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children," J. Am. Acad. Audiol **19**, 135–146.

Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–1436.

Summers, V., and Leek, M. R. (**1998**). "Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners," Hear. Res. **118**, 139–150.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception II," J. Acoust. Soc. Am. **93**, 1547–1552.

Tyler, R. S., Summerfield, A. Q., Wood, E. J., and Fernandes, M. A. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **72**, 740–752.

Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.

Walden, B. E., Erdman, S. A., Montgomery, A. A., Schwartz, D. M., and Prosek, R. A. (**1981**). "Some effects of training on speech recognition by hearing-impaired adults," J. Speech Hear. Res. **24**, 207–216.

Wightman, F., Kistler, D., and Brungart, D. (**2006**). "Informational masking of speech in children: Auditory-visual integration," J. Acoust. Soc. Am. **119**, 3940–3949.

Wilson, R. H., Carnell, C. S., and Cleghorn, A. L. (**2007**). "The words-in-noise (WIN) test with multitalker babble and speech-spectrum noise maskers," J. Am. Acad. Audiol **18**, 522–529.

The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean

Sahyang Kim
Department of English Education, Hongik University, Seoul 121-791, Korea

Taehong Cho[a)]
Department of English Language and Literature, Hanyang University, Seoul 133-791, Korea

(Received 4 May 2008; revised 14 February 2009; accepted 20 February 2009)

This study investigated the role of phrase-level prosodic boundary information in word segmentation in Korean with two word-spotting experiments. In experiment 1, it was found that intonational cues alone helped listeners with lexical segmentation. Listeners paid more attention to local intonational cues (…H#L…) across the prosodic boundary than the intonational information within a prosodic phrase. The results imply that intonation patterns with high frequency are used, though not exclusively, in lexical segmentation. In experiment 2, final lengthening was added to see how multiple prosodic cues influence lexical segmentation. The results showed that listeners did not necessarily benefit from the presence of both intonational and final lengthening cues: Their performance was improved only when intonational information contained infrequent tonal patterns for boundary marking, showing only partially cumulative effects of prosodic cues. When the intonational information was optimal (frequent) for boundary marking, however, poorer performance was observed with final lengthening. This is arguably because the phrase-initial segmental allophonic cues for the accentual phrase were not matched with the prosodic cues for the intonational phrase. It is proposed that the asymmetrical use of multiple cues was due to interaction between prosodic and segmental information that are computed in parallel in lexical segmentation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097777]

PACS number(s): 43.71.Sy, 43.71.Es, 43.66.Lj [AJ] Pages: 3373–3386

I. INTRODUCTION

In order to comprehend spoken language successfully, listeners must be able to segment the stream of speech into individual words. The lexical segmentation process, however, is by no means trivial. Not only is there no invariant acoustic cue that consistently signals word boundaries, but there also exist multiple layers of phonetic and phonological variation within and across words which add complexity to the process of word boundary search. A long-standing question in the field of speech comprehension has therefore been how listeners find word boundaries successfully, given lack of consistent acoustic cues (see McQueen, 2005 for a review). One approach to lexical segmentation is to consider the process as a consequence of lexical competition (e.g., Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Norris, 1994). In lexical competition, a set of cohort competitors whose acoustic onsets are matched with the input is initially activated. Competitors are then inhibited as soon as they mismatch the input, eventually leaving a single candidate as the winner in the competition. As lexical competition ends, a search for the word boundary is also finalized, and so is lexical segmentation. Lexical competition mechanisms relying only on phonemic representations, however, could result in ambiguous parsing for a given speech stream (e.g., *I scream* vs *ice cream*, Lehiste, 1960), especially when no other semantic and/or pragmatic context is available.

A large body of recent psycholinguistic research has shown that such ambiguity can be resolved by fine-grained phonetic information in the speech input, suggesting that lexical segmentation is modulated by subphonemic information. A well-known case is the use of subtle durational difference in lexical processing. For example, ambiguous Dutch sequences due to resyllabification (e.g., *diep # in* vs *die # pin*, Quené, 1993) and lexically ambiguous sequences interpretable as two words or as one word (e.g., *two lips* vs *tulips*, Gow and Gordon, 1995) are both reliably differentiated by subtle durational cues. Similarly, temporary ambiguity that arises due to an initially embedded word in a longer word (e.g., *cap* in *captain*) is also resolved by the word's duration (Salverda *et al.*, 2003). Further evidence for the exploitation of phonetic details can be found in the use of word-internal coarticulatory information (e.g., Dahan *et al.*, 2001), assimilatory information (e.g., Gow, 2002), and phonetic differences due to the syllable structure (e.g., Tabossi *et al.*, 2000). Although these studies have successfully demonstrated that lexical segmentation is modulated by fine-grained phonetic cues that signal lexical boundaries, their focus of attention has been only on the use of the phonetic cues that are mainly associated with low-level linguistic structures of an utterance, especially in the syllable or the word levels.

[a)] Author to whom correspondence should be addressed. Electronic mail: tcho@hanyang.ac.kr

Some studies have suggested that prosodic information about lexical stress, which is generally expressed by prosodic cues such as pitch, duration, and amplitude (Lehiste, 1970), is exploited in lexical segmentation. English listeners, for instance, tend to segment words based on the strong-weak lexical stress pattern, treating the strong (stressed) syllable as the beginning of a word with a stressed syllable (Cutler and Butterfield, 1992; Cutler and Norris, 1988). Fragment priming experiments in Spanish (Soto *et al.*, 2001) showed that when the stress of the spoken fragment (e.g., *prin*) is matched with the stress of the visual target (e.g., *PRINcipe* "prince" vs *prinCIpio* "beginning"), recognition of the target word is facilitated, but inhibition occurs when stress is mismatched. Note that these studies are also confined to word-level prosodic effects on lexical processing.

Our understanding has therefore been limited with respect to how phonetic information of higher-level structure is used in lexical segmentation. In particular, although it has been well established that an utterance is produced with phonetic markers of high-level prosodic structure (e.g., Beckman, 1996; Keating and Shattuck-Hufnagel, 2002), possible roles of its acoustic consequences in lexical processing have not been fully understood. Recent studies have therefore attempted to expand their scope of investigation of lexical segmentation cues, exploring how high-level prosodic information of a given utterance influences lexical segmentation process (Cho *et al.*, 2007; Christophe *et al.*, 2004; Shukla *et al.*, 2007; Welby, 2007).

A. Prosodic structure and its importance for lexical processing

A tenet of prosodic phonology is that an utterance is produced with prosodic structure which is assumed to be organized in such a way that prosodic constituents of different sizes are hierarchically nested (see Shattuck-Hufnagel and Turk, 1996 for a review). According to a model of prosodic organization in English (Beckman and Pierrehumbert, 1986), for example, the prosodic structure consists of the syllable, the prosodic word (PW), the intermediate phrase (ip), and the intonational phrase (IP) (see Nespor and Vogel, 1986; Selkirk, 1984 for similar prosodic structural views). These prosodic domains are assumed to be strictly layered, such that a prosodic domain of one level is exhaustively parsed into constituents of the immediately next-lower level (Selkirk, 1984). The prosodic structure is known to be marked by various prosodic cues such as pause (e.g., Gee and Grosjean, 1983; Krivokapić, 2007), phrase-final lengthening (e.g., Edwards *et al.*, 1991; Wightman *et al.*, 1992), intonation (e.g., Beckman and Pierrehumbert, 1986; Ladd, 1996), and domain-initial strengthening (e.g., Cho and Keating, 2001; Fougeron and Keating, 1997; Keating *et al.*, 2003).

The structural view of prosody generally assumes that prosodic structure is a crucial element of speech production and comprehension processes (e.g., Beckman, 1996). A general hypothesis is that if the speaker produces an utterance based on prosodic structure generated online (Keating and Shattuck-Hufnagel, 2002), its acoustic consequence should be exploited by listeners in speech comprehension. Christophe *et al.* (2004), for example, demonstrated that lexical segmentation is modulated by prosodic structure. In their word-monitoring experiments, a local ambiguity (e.g., [d'un **chat grin** cheux] "of a grumpy cat," ambiguous with *chagrin*) slowed down the detection of a target word (e.g., *chat*, "cat"), but the ambiguity effect disappeared when the ambiguity-creating sequence (e.g., *chat grin*) spanned a prosodic phrase boundary. Shukla *et al.* (2007) also showed that Italian listeners were better at recognizing words that were internal to a prosodic phrase than the same syllable sequences spanning a phrase boundary. Strong articulation of segments after a prosodic boundary (i.e., domain-initial strengthening) is also known to help listeners to recognize the word before the prosodic boundary as the domain-initial strengthening serves as a cue to the beginning of a new word (Cho *et al.*, 2007).

These empirical findings demonstrate that prosodic structure, as phonetically manifested in the speech input by combination of various prosodic cues such as pause, intonation, and duration, plays an important role in lexical segmentation. What is not yet clear, however, is exactly what kind of phonetic information of the prosodic structure is exploited by the listener in lexical processing. The present study therefore explores effects of two major prosodic cues, intonation and duration, in two word-spotting experiments to see how independently or collectively these cues are used in lexical segmentation in Korean.

The prosodic model of Seoul Korean that we adopt for our study is the one proposed by Jun (1993, 1995, 2000), which is by far the most widely adopted model in the literature. The prosodic hierarchy consists of the syllable, the PW, the accentual phrase (AP), and the IP. It also assumes the strict layer hypothesis (Selkirk, 1984), so that the edges of IP always coincide with the edges of APs, which in turn coincide with the edges of PW. The Korean prosodic model differs from the English prosodic model discussed above in that it assumes the AP between the IP and the PW. Unlike the English IP, the AP is not marked by a noticeable phrase-final lengthening (Jun, 1993, 1995; cf. Cho and Keating, 2001), and its intonational structure is independent of word-level prosody, as Seoul Korean does not have any word-level prosody such as lexical stress, pitch accent, and tone (Jun, 1993).

The Korean AP is intonationally defined, having default initial and final rising intonation patterns at the edge (i.e., #LH…LH#, where "#" refers to an AP boundary). However, the intonation system interacts with AP-initial segmental information, such that an AP that starts with an aspirated or a tense consonant is associated with #HH (but otherwise with #LH, including AP that starts with a vowel). As the AP is produced with no discernible final lengthening at the end, substantial final lengthening is associated only with the IP (Jun, 1993; Chung *et al.* 1996). With respect to boundary tones that mark the end of an IP, various intonation patterns have been identified such as L%, H%, LH%, and HL% ("%" refers to an IP-final tone), all of which occur in the IP-final syllable (Jun, 2000). When an AP is located IP-finally, the AP-final tone is overridden by the IP-final boundary tone.

AP-initial tones, however, are preserved regardless of the position of an AP within an IP since the IP does not have any initial boundary tone.

Some researchers have suggested that AP-internal intonational structure in Seoul Korean may be further constrained by the structure of the AP-initial syllable (with or without a coda consonant) and the vowel quantity (long vs short) (e.g., Lim and de Jong, 1999; Park, 2004). The vowel quantity distinction, however, is not maintained anymore in Seoul Korean, and underlying intonational structure of AP assumed by Jun (1993) is still supported by the statistics of corpus studies (Jun and Fougeron, 2000; Kim, 2004). Kim's (2004) study, which transcribed Korean intonation patterns in read speech and radio drama, showed that when AP-initial consonants were not aspirated or tense, about 88% of AP-initial multisyllabic content words started with a rising (LH) tone and about 85% of APs ended with a final H tone. Crucially, this frequency pattern was observed regardless of the AP-initial syllable structure. Other tonal patterns such as #LL, #HL, or #HH did occur, but with a very low frequency (9% for #LL, 2% for HH, and 1% for #HL). The notion of AP which is intonationally-defined as a prosodic unit has been further supported by segmental phonology—i.e., AP serves as an application domain of some phonological rules. Jun (1993), for example, showed that the lenis stop intervocalic voicing rule (where a lenis stop becomes voiced between vowels) applies within an AP, although voiced variants do occur sometimes in AP-initial position (cf. Cho and Keating, 2001).[1] Other phonological rules operating within an AP include post-obstruent tensing (a lenis stop becomes tense after an obstruent: Jun, 1998), lateralization (/n/ becomes [l] after /l/: Kim, 2000) and *n*-insertion (/n/ is inserted stem-initially in stems that begin with /i/ or /j/ when it is preceded by a stem or a prefix ending with a consonant: Kim, 2000).

To recap, we selected Seoul Korean as our target language because it is prosodically interesting in two ways. First, it does not have any word-level prosody (Jun, 1995), so the intonation structure of an utterance is determined solely at the phrase level without any influence from word-level prosody. Second, the medium-sized phrase AP is not accompanied by substantial final lengthening at the end (Jun, 1993, 1995). These unique properties of Korean prosody allow us to observe the role of intonational cues of AP in lexical segmentation without any confounding effects from word-level prosody and other domain-edge phenomena.

There have, in fact, been attempts to understand the role of AP-initial (postboundary) tones in lexical segmentation. For example, Warner *et al.* (2009) and Kim (2004) showed that speakers of Japanese and Korean, respectively, can use the AP-initial rising intonational cue in online lexical processing. In both studies, however, potential prosodic cues (domain-initial strengthening and boundary-adjacent lengthening) other than AP-initial intonational cues were not completely eliminated. The present study is therefore the first that controls the experimental condition in such a way that the sole effect of intonation patterns of AP in online segmentation can be observed without any confounding effects that would otherwise stem from other possible phonetic cues available at prosodic boundaries. In Sec. I B, we will discuss specific research questions that are to be addressed in the present study.

B. Research questions

In the present study, two word-spotting experiments are carried out to examine the use of prosodic information in lexical segmentation. In experiment 1, we explore how intonational cues alone influence lexical segmentation and in experiment 2, we add phrase-final lengthening as an additional cue to understand how single vs multiple prosodic cues are processed by listeners in lexical segmentation.

In experiment 1, both preboundary and postboundary AP intonational sequences are considered. As for AP-initial (postboundary) tones, we compare the effects of four different intonation patterns superimposed upon target words: frequent #LH vs infrequent #LL, #HH, and #HL. (They are used for disyllabic target words; see below for tonal descriptions of trisyllabic target words.) As for AP-final (preboundary) tones before the target word, two intonation patterns, frequent H# vs infrequent L#, are used. Crucially, all the consonants used as initial segments of the target words are either lenis stops or nasals, which are extracted from AP-medial position in order to eliminate other potential prosodic cues such as domain-initial strengthening cues that might affect the listener's performance (Cho *et al.*, 2007).

We specifically ask how the frequency of intonational cues influences listeners' performance in lexical segmentation. It is well established that frequency influences lexical access in terms of word frequency (Forster and Chambers, 1973; Norris, 1986) and sequential probabilities (e.g., Saffran, *et al.*, 1996; Vitevitch, *et al.*, 1997). The statistics of stress patterns within the vocabulary (i.e., stress tends to fall on initial syllables in English and Dutch: Cutler and Carter, 1987; Schreuder and Baayen, 1994) also influences lexical access—i.e., listeners tend to put a word boundary before a stressed syllable (Cooper *et al.*, 2002; Cutler and Butterfield, 1992; Vroomen and de Gelder, 1995). It is therefore reasonable to hypothesize that the frequency of intonation patterns also influences detection of a phrase boundary which coincides with a lexical boundary, such that listeners will perform better with target detection with frequent intonation patterns for AP than less frequent intonation patterns.

Another important question is whether the preboundary and the postboundary intonation patterns are exploited independently by the listener. Previous studies have focused on the role of within-phrase prosodic cues to a prosodic boundary. For instance, Welby (2007) and Warner *et al.* (2009) examined the role of phrase-initial prosodic cues, and the work of Christophe *et al.* (2004) was based on the assumption that listeners use within-phrase prosodic information to terminate their lexical search before a prosodic boundary. It is then possible to posit that intonation patterns within each prosodic phrase may play an independent role in speech processing. Under this hypothesis, the detection of target words is expected to be more facilitated with the frequent tonal pattern (#LH) than with the infrequent patterns (#LL, #HH, and #HL), regardless of whether the preboundary tone is H#

or L#. Likewise, the frequent AP-final rising tone (with H#) is expected to signal the end of AP, which, at the same time, indicates that what is coming up is the beginning of another AP. The presence of the frequent AP final tone (H#) before the boundary is therefore expected to help listeners' recognition of the postboundary target word, regardless of the postboundary tones. Alternatively, however, if one of the goals of prosodic structuring in speech production were to mark prosodic boundaries, which would be eventually available to the listener in speech comprehension, the crucial prosodic information might be present locally at the prosodic boundary (just before and after it), with H#L being the most frequent pattern. If so, there would be an interaction between the preboundary and the postboundary intonational effects, in such a way that target detection would be facilitated as long as the local boundary condition (H#L) is met. It is then expected that #LH and #LL, which differ in frequency but both meet the locality condition, would show similar facilitatory effects on the target detection as long as the preboundary tone is H#.

Examining intonational effects on lexical processing in Korean also allows us to address the general vs language-specific perceptual role of a high (H) pitch element in lexical segmentation. Intonational cues are often associated with an H pitch element or an F0 rise. Warner *et al.* (2009) therefore suggested that F0 rise is perceptually salient, and it would facilitate detection of syllables marked by it. Under this assumption, an H tone on the initial syllable of the target word would facilitate lexical segmentation regardless of listeners' linguistic background. Interestingly, however, Korean has an H tone, but it frequently falls on the second syllable of AP, providing a counter example to this assumption. Based on the results of the present study, we will address this issue, especially in terms of how the language-specific distribution of H in Korean bears on the issue of language-specific vs cross-linguistic use of the perceptually salient F0 rise in lexical processing.

In experiment 2, a final-lengthening factor is added. As in experiment 1, two preboundary tones (H# vs L#) at the preboundary syllables are used. For the postboundary tones, for the sake of simplicity, just two extreme conditions from experiment 1 are used: #LH (the most frequent) and #HL (the least frequent).

As phrase-final lengthening is known to be another important boundary-marking phonetic event that co-occurs with boundary-marking tones (e.g., Edwards *et al.*, 1991; Wightman *et al.*, 1992), some researchers have suggested that listeners make use of subtle phonetic differences and compute prosodic structure of an utterance during online word recognition (e.g., Christophe *et al.*, 2004; Salverda *et al.*, 2003; Shatzman and McQueen, 2006). We therefore test how intonational effects interact with final lengthening cues. Important questions to be addressed are how cumulative multiple prosodic cues facilitate lexical processing and how the cumulative effect is constrained by a mismatch between segmental and prosodic cues.

Spitzer *et al.* (2007) showed that the level of intelligibility gets lower when more cues of lexical stress are missing in the speech signal, suggesting that available cues are used in a cumulative way. It has been also proposed that listeners make immediate use of any available cues in order to modulate the activation of lexical competitors (Donselaar *et al.*, 2005), and all the available information is used (Norris *et al.*, 1997). Cumulative boundary cues could therefore be a very effective tool to modulate lexical search: Listeners would be able to guess the location of the end of a prosodic phrase with more certainty when the final lengthening cue is added to the intonational cue. It is therefore hypothesized that the addition of substantial final lengthening will augment the effect of intonational cues, facilitating listeners' target detection in a cumulative way.

The presence of substantial lengthening at the end of a phrase would, in fact, give rise to a percept of an IP boundary before the target word. However, adding the lengthening cue to the speech input would not make a seamless IP boundary percept in the current experiment because a mismatch would arise between segmental and prosodic information. Recall that in the present study, we use speech materials extracted from phrase-internal (AP-medial) position in order to eliminate potential prosodic cues at domain edges. IP-initially, consonant durations (including VOTs for lenis stops) are longer and lenis stops are always voiceless (no application of lenis stop voicing rule across an IP boundary). Yet, in experiment 2, initial consonants of the target words lack such domain-initial strengthening cues, including lenis stop voicing cues. That is, the consonants that listeners hear are not consistent with IP-initial, but compatible with IP-medial position. Such a mismatch between prosodic and segmental information might hinder the detection of the target words (e.g., Cho *et al.*, 2007; Salverda *et al.*, 2003).

There are therefore two competing hypotheses. On the one hand, if the cumulative effect takes precedence in lexical segmentation process, the presence of an additional phrase-final lengthening would augment listener's performance with the target detection. On the other hand, if the effect of the mismatch between prosodic and segmental cues comes into play, it would at least suppress the cumulative effect (showing no further facilitation with the presence of phrase-final lengthening) or under its strongest influence, it could override effects of both intonational and lengthening cues.

In experiments 1 and 2, we therefore test various hypotheses in order to understand the relationship between high-level prosodic information and lexical segmentation. Building on our knowledge about the relationship will also have an important implication for existing models of speech segmentation such as TRACE (McClelland and Elman, 1986), shortlist (e.g., Norris, 1994; Norris and McQueen, 2008), the distributed cohort model (Gaskell and Marslen-Wilson, 1997), and the hierarchical model (Mattys *et al.*, 2005) because these models do not take high-level prosodic information into account.

II. EXPERIMENT

In experiment 1, we carried out a word-spotting experiment in Korean in order to explore how prosodic structural information manifested in intonation influences lexical seg-

TABLE I. Intonation patterns of carrier strings and target words. (# indicates an assumed phrase boundary; underlined syllables in bold indicate target words.)

(a) A seven-syllable carrier string with a disyllabic target word

First two syllables		Preboundary syllable	Disyllabic target word (postboundary word)	Last two syllables	
σ1	σ2	σ3	**σ4 σ5**	σ6	σ7
L	L	L #	**LH**	L	L
		H #	**LL**		
			HH		
			HL		

(b) An eight-syllable carrier string with a trisyllabic target word

First two syllables		Preboundary syllable	Trisyllabic target word (postboundary word)	Last two syllables	
σ1	σ2	σ3	**σ4 σ5 σ6**	σ7	σ8
L	L	L #	**LLH**	L	L
		H #	**LLL**		
			HHL		
			HLH		

mentation. We tested whether intonation patterns that are frequently associated with AP boundaries would facilitate detection of lexical boundaries.

A. Method

1. Participants

Ninety-six student participants from Hanyang University in Seoul were paid for their participation. They were all native speakers of Seoul Korean who were born and raised in the Seoul metropolitan area. They were divided into eight groups of 12, according to experimental conditions that will be described below.

2. Materials

24 disyllabic and 24 trisyllabic Korean words were selected and inserted, respectively, in seven- and eight-syllable nonsense carrier strings. Other than the target word, no consecutive syllables in a carrier string formed a word in Korean. Target words and target-bearing carrier strings were composed of open syllables (CV) only. 14 out of 24 disyllabic words and 15 out of 24 trisyllabic words started with an oral lenis stop (/p/, /t/, or /k/), and the rest started with a nasal stop (/m/ or /n/). Lenis stops and nasals were chosen because they are known to be associated with the AP-initial rise tone (#LH) (Jun, 1993, 2000), which was confirmed by Kim's (2004) corpus study. The list of target words and carrier strings is given in the Appendix.

As shown in Table I, the initial syllable of a target word was always the fourth syllable of carrier strings, such that each target word was preceded by three syllables and followed by two syllables. We used two- and three-syllable target words to avoid any potential performance bias that comes from a fixed number of syllables of the target word. This also allows us to see potential effects of the syllable count, as the number of syllables in target words was found to affect listeners' word segmentation (Kim, 2004). The carrier strings with a disyllabic target word had seven syllables [Table I(a)] and those with a trisyllabic target word had eight syllables in total [Table I(b)]. Each syllable in a carrier string was associated with one tone, which was either H or L. The first two syllables and the last two syllables of carrier strings were controlled with an L tone. In order to observe preboundary tone effects on word segmentation, the syllable that immediately preceded the target word in each carrier string (viz., σ3 #, where # refers to the intended phrase boundary) had a preboundary tone of either H# (frequent) or L# (infrequent).

For intonation patterns on target words, one of the four postboundary tones was used. The frequent pattern for disyllabic target words was a rising tone with #LH and the infrequent patterns were #LL, #HH, and #HL. When the target word was trisyllabic, three tonal elements were needed for each word. The four intonation patterns employed were #LLH (frequent), #LLL, #HHL, and #HLL (infrequent). They were then matched with the infrequent disyllabic intonation patterns in the first two tonal elements: #LL vs #LLL, #HH vs #HHL, and #HL vs HLL. For the sake of simplicity, we will use the disyllabic intonation patterns (#LH, #LL, #HH, and #HL) throughout the paper when we refer to the intonation patterns of the target words. An important note here concerns the predicted effect of the infrequent #LL for disyllabic targets vs the frequent #LLH for trisyllabic targets. One of the hypotheses for the frequency effect was that the frequent AP-internal tonal pattern #LLH for trisyllabic targets would help listeners to detect the targets. However, given that the frequent #LLH shares the same initial #LL sequence with the infrequent #LL for a disyllabic target, one might wonder how listeners are expected to treat #LL as an infrequent cue for disyllabic targets but as part of the frequent cue (#LLH) for a trisyllabic target. In the experiment, an equal number of disyllabic and trisyllabic targets was employed. Listeners, therefore, did not know *a priori* whether the target word would be disyllabic or trisyllabic. As the second L tone of #LL is being heard, the likelihood for the target to be disyllabic would become weakened because #LL as a whole is not a frequent tone for disyllabic words. As the third tone H is being heard, however, it is matched with the frequent #LLH pattern for trisyllabic words, and therefore the likelihood for the target to be trisyllabic would increase. In this way, if the frequency of the phrase-internal tonal patterns would play a role in lexical segmentation, the LL portion would not work in favor of disyllabic targets, but it would help detect trisyllabic targets in the form of #LLH, even if listeners did not know beforehand the number of syllables for the targets.

Since each target word appeared in eight different intonational conditions (2 preboundary tones × 4 postboundary tones), there were eight experimental lists. In addition to the experimental items, 36 disyllabic and 36 trisyllabic Korean words were selected as fillers and were included in the experimental lists. The location of a filler word in filler-bearing strings, however, differed from that of a target word in experimental target-bearing strings in order to avoid the potential bias caused by the fixed location of both the experimental target and filler words. As with the target-bearing strings, two preboundary tones and four postboundary tones were

TABLE II. Mean error rates (%) in experiment 1 with three factors, syllable count, preboundary tone, and postboundary (target word) tone. (Standard errors are included in parentheses.)

No. of syllables in target words	Preboundary tone	Postboundary (target word) tone			
		#LH	#LL	#HH	#HL
2	H#	38.5	45.1	54.5	62.1
		(2.1)	(2.2)	(2.5)	(2.5)
	L#	48.6	54.8	49.6	62.1
		(2.5)	(2.4)	(2.6)	(2.8)
3	H#	15.2	19	24.6	28.8
		(1.9)	(2)	(2.2)	(2.4)
	L#	22.2	23.2	23.6	31.2
		(2.2)	(2.2)	(2.1)	(2.6)

evenly distributed among filler-bearing strings. Additional eight words were selected and inserted in carrier strings for practice items.

Eight experimental lists were arranged such that each subject heard every word just once in one of the eight intonational conditions. Each list contained 48 target-bearing strings and 72 filler-bearing strings in a pseudo-random order, and the filler strings appeared in the same order in all eight lists. There were no two stimuli with the same intonation pattern presented in a row.

To ensure that the carrier string did not have any other confounding prosodic factors on crucial syllables, such as phrase-final lengthening on the preboundary syllable ($\sigma3\#$) and acoustic consequences of domain-initial strengthening on the postboundary syllable (the initial syllable of target words; $\#\sigma4$), the following recording and splicing procedure was performed.

Each seven-syllable string (with a disyllabic target word) was divided into three four-syllable chunks ($[\sigma1\text{-}\sigma2\text{-}\sigma3\text{-}\sigma4]$, $[\sigma3\text{-}\underline{\boldsymbol{\sigma4\text{-}\sigma5}}\text{-}\sigma6]$, and $[\sigma4\text{-}\sigma5\text{-}\sigma6\text{-}\sigma7]$). (Here the underlined syllable symbols in bold were the ones that were actually used for target words.) A female native speaker of Seoul Korean who was naive about the experiment's purpose produced each chunk separately multiple times with flat intonation and with as consistent speaking rate as possible. The recording was made in a sound-proof booth onto a TASCAM HD-P2 digital recorder at the sampling rate of 44 kHz. Among the multiple tokens, the best tokens were selected that did not contain any unintended prosody, agreed by transcriptions of the two authors. The recorded strings were then spliced to be used for the actual speech input string as follows. The first three syllables of a seven-syllable carrier string were spliced from $[\sigma1\text{-}\sigma2\text{-}\sigma3\text{-}\sigma4]$ so that the preboundary syllable ($\sigma3\#$) of the actual input string ($\sigma1\sigma2\sigma3\#\underline{\boldsymbol{\sigma4\sigma5}}\sigma6\sigma7$) did not involve phrase-final lengthening. Target words ($\sigma4\sigma5$) were spliced from $[\sigma3\text{-}\underline{\boldsymbol{\sigma4\text{-}\sigma5}}\text{-}\sigma6]$ so that the initial syllable ($\#\sigma4$) of target words did not have characteristics of domain-initial strengthening.[2] Finally, the last two syllables of the actual input string were the last two syllables of $[\sigma4\text{-}\sigma5\text{-}\sigma6\text{-}\sigma7]$. A similar procedure was employed to build eight-syllable strings with trisyllabic target words. Filler-bearing strings and practice items were also recorded and spliced in the same manner. The splicing was made at zero crossings, using PRAAT.

After splicing was completed, the pitch of each carrier string was manipulated using the pitch-synchronous overlap and add technique with PRAAT software. The speaker's flat intonation was regarded as a default L tone (143 Hz), and pitch was raised when H tone was required in the experimental setting. F0 minima for an L tone and F0 maxima for an H tone were aligned with the midpoint of the vowel of the target syllables. The rates of pitch rising were 1.23 times at the preboundary syllable ($\sigma3\#$) and 1.16 times at the target-word final syllable (i.e., $\sigma5$ in disyllabic words and $\sigma6$ in trisyllabic words). The rates were determined by the rate of AP-final and AP-initial pitch rising from other recordings of the speaker with natural sentences.

3. Procedures

Subjects were tested individually in a sound-attenuated room. Stimuli presentation and data collection were performed by NESU software and a button box (www.mpi.nl/world/tg/ experiments/nesu.html). Subjects heard the stimuli on a PC through a pair of headphones at a comfortable volume. They were told that they would hear a list of nonsense strings and were instructed that they should spot a real Korean word in each string. They were asked to press a button with their preferred hand as quickly and as accurately as possible when they spotted a real word, and then to say the word aloud. Each subject was presented with eight practice items and was then given one of the eight experimental lists. There were 120 strings in each list. Subjects heard 70 strings (for approximately 12 min), took a 1-min break, and then continued with the rest of the strings (for approximately 8 min). An experimenter was always in the room with a subject during the experimental session and monitored the subject's missing or incorrect responses. Subjects had to detect words in both target-bearing and filler-bearing strings, but their responses for filler-bearing strings were not analyzed.

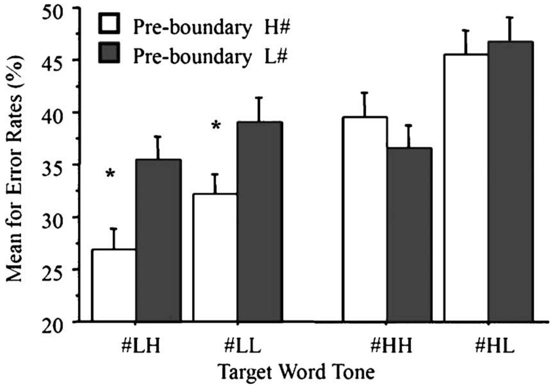


FIG. 1. Error rates in different tone conditions with an interaction between preboundary tone and postboundary (target word) tone. (* refers to $p < 0.05$ in *posthoc* analyses.) The postboundary tones are #LH (#LLH), #LL (#LLL), #HH (#HHL), and #HL (#HLL), where non-parenthetical tones are for disyllabic words and parenthetical tones are for trisyllabic words. Note that the syllable count factor did not interact with either preboundary tone or postboundary tone.

B. Results

Reaction time (RT) was the duration between the offset of the target word and button press. Missing items, incorrect responses, and RT over 1500 ms were treated as errors. Mean error rates are summarized in Table II.

The error rates were very high especially when target words were disyllabic, as shown in Table II. Over 50% of target words (51.9% error rates on average) were missed in the disyllabic conditions and 23.5% in the trisyllabic conditions. Because overall error rates were very high, the latency analyses were not entirely reliable, although the latency results were by and large comparable with the accuracy results. In the present study, we will therefore report just results of the accuracy analyses. For the accuracy analyses, the error rates were submitted to repeated measures analyses of variance (ANOVAs) with the factors syllable count (number of syllables in target words; 2 vs 3), preboundary tone (frequent H# vs infrequent L#), and postboundary (=target word) tone (frequent #LH vs infrequent #LL, #HH, and #HL).

All three factors showed significant main effects. Subjects were more accurate when target words were trisyllabic than when they were disyllabic ($F1[1,88]=442.25$, $p < 0.001$; $F2[1,46]=14.56$, $p < 0.001$). Target words were detected with significantly lower error rates when preboundary tone was H# (frequent) than when it was L# (infrequent) ($F1[1,88]=9.84$, $p < 0.005$; $F2[1,46]=5.12$, $p < 0.05$). The postboundary tone effect was significant ($F1[3,264]=29.11$, $p < 0.001$; $F2[3,138]=15.01$, $p < 0.001$), and Bonferroni *posthoc* tests showed an overall pattern of #LH<(#LL=#HH)<#HL. #LH was significantly different from the other three conditions ($p < 0.05$ both by-subjects and by-items), showing the lowest error rates. The two infrequent intonation patterns #LL and #HH were also significantly different from the least frequent pattern #HL ($p < 0.05$ both by-subjects and by-items), while #LL and #HH were not significantly different from each other.

There was an interaction between preboundary tone and postboundary tone ($F1[3,264]=5.73$, $p < 0.005$; $F2[3,138]=5.52$, $p < 0.005$), as illustrated in Fig. 1. The effect of preboundary tone was significant only when postboundary tone was #LH or #LL (with #LH, $F1[1,95]=9.73$, $p < 0.005$, $F2[1,47]=18.81$, $p < 0.001$; with #LL, $F1[1,95]=6.84$, $p < 0.05$, $F2[1,47]=8.86$, $p < 0.01$), showing that listeners' better performance with preboundary tone H# was reliable only when postboundary tone started with #L (viz., #LH or #LL). Likewise, the effect of postboundary tone was reliable only when preboundary tone was H#. In other words, when preboundary tone was H#, the order of error rates by postboundary tone was (#LH=#LL)<(#HH=#HL) ($p < 0.01$ by-subjects, $p < 0.05$ by-items), but when preboundary tone was L#, the order of error rates by postboundary tone was (#LH=#LL=#HH)<#HL ($p < 0.01$ by-subjects, $p < 0.01$ by-items), with #LL being no different from #HL. There was no difference between #LH and #LL regardless of preboundary tone ($p > 0.05$ both by-subjects and by-items).

C. Summary and discussion

In experiment 1, a general finding was that the preboundary and postboundary tones that are used most frequently to mark AP boundaries in speech production indeed help listeners to recognize words: The detection accuracy was higher when the preboundary tone ended with the frequent H# (vs L#) and when the postboundary tone started with the frequent #L (#LH, #LL vs #HH, #HL). This supports the basic hypothesis that the frequency of intonation patterns for AP is exploited by listeners, such that frequent intonation patterns facilitate lexical segmentation.

The results, however, showed interactions between the preboundary and the postboundary tones. The frequent preboundary H# improved listeners' word-spotting performance, but this effect held only when the postboundary tone (on the target word) also started with the frequent intonational element #L, but with no difference between #LH (frequent) and #LL (infrequent). Likewise, the most frequently occurring postboundary tone #L was found to be useful only when the preceding (preboundary) tone was also most frequent (i.e., with H#), but again with no difference between #LH (frequent) and #LL (infrequent). That is, as long as the first element of the postboundary tone was #L, listeners' performance was not influenced by the second tonal element of the postboundary intonational sequence (#LH, #LL). This supports the hypothesis that what is important is whether adjacent intonational elements form an H#L sequence locally across the boundary, rather than the individual patterns of preboundary and postboundary intonational sequences. Insofar as the boundary-spanning local information (H#L) was available to the listener, what comes after that does not seem to influence listeners' boundary detection in a noticeable way.

One might wonder why the second element of the postboundary intonational sequence (i.e., H in the frequent #LH) does not contribute to lexical segmentation, especially when the local intonational condition (an H#L sequence) is met. One possible explanation for this is as follows. Upon hearing the preboundary H#, listeners would start to entertain the possibility of a phrase boundary, and the subsequent postboundary #L would confirm their boundary decision (as #L is the typical AP-initial tone). Lexical search would then be initiated and cohort competitors would be simultaneously activated from the point when listeners make a decision about where to put a boundary. Whichever tone pattern follows the postboundary #L (e.g., #LH vs #LL), the same members of cohort are already activated, and therefore the effect of the second intonational element would be the same on all the cohort competitors. This is indeed in line with the explanation of Cho *et al.* (2007). Domain-initial strengthening cues immediately after a prosodic boundary (e.g., lengthened VOTs) would not necessarily facilitate the recognition of the postboundary word because the heightened acoustic clarity of domain-initial position would be beneficial to not only the target word but also all other cohort competitors that share the same initial consonant.

Finally, results showed that listeners were more accurate in segmenting words when they were trisyllabic than when they were disyllabic. As in Kim (2004), this can be accounted for by the different neighborhood density of target words (e.g., Luce, 1986; Luce and Pisoni, 1998)—i.e., trisyllabic words used in the present study had less neighboring words to compete against than disyllabic words had. A *posthoc* analysis on the neighborhood density supports this account. The number of phonological neighbors of 48 target words was calculated based on Kim and Kang (2004)'s corpus, which had 550,000 listed words. A phonological neighbor (i.e., a phonologically similar word) was defined by an addition, deletion, or substitution of a segment to a target word regardless of the location of a segment within a word. The calculated result showed that disyllabic targets had about 15 times more phonological neighbors than trisyllabic targets (mean, 49, s.d., 22.2 vs mean, 3.3, s.d., 2.1, respectively). The syllable count effect did not interact with intonational effects, suggesting its independence.

III. EXPERIMENT 2

In experiment 2, we tested how intonational effects observed in experiment 1 would interact with final lengthening cues. Important questions were whether multiple prosodic cues (intonation and duration) would cumulatively facilitate lexical processing, and how the cumulative effect would be constrained by a mismatch between segmental and prosodic cues—i.e., an IP boundary percept created just by final-lengthening and intonational cues at the boundary mismatched with domain-initial strengthening cues (including lenis stop's allophonic cues).

A. Method

1. Participants

Ninety-six student participants from Hanyang University were paid for their participation. They were all native speakers of Seoul Korean born and raised in the Seoul metropolitan area and had not participated in experiment 1. They were divided into eight groups of 12, according to the experimental conditions that will be explained below.

2. Materials

We employed two degrees of lengthening (lengthening vs no lengthening) and two preboundary tones (H# vs L#) at the preboundary syllables. For the postboundary tones, for the sake of simplicity, we included just two extreme conditions from experiment 1: #LH, which showed the most robust effect, and #HL, which showed the smallest effect. For the lengthening condition, the vowels of the preboundary syllables were lengthened, resulting in about 1.7 times of the original syllable's vowel length. The range of syllable durations was 178–259 ms in the no lengthening condition and 256–386 ms in the lengthening condition. The rate of lengthening was determined based on multiple speakers' mean values reported by Chung *et al.* (1996)—i.e., IP-final syllables are about 1.7 times longer than non-final syllables. (Note that unlike this lengthening manipulation, the pitch manipulation in experiment 1 was based on the talker's own recordings as pitch range is talker-specific, generally constrained by the talker's physiological characteristics.) PRAAT was used for duration manipulation. The two authors, as trained Korean intonation transcribers, agreed that the lengthened versions of H# and L# gave IP-final percepts as H% and L%, respectively. Target words (both disyllabic and trisyllabic), filler words, and segmental order of carrier strings were the same as those used for experiment 1. In total, there were two levels of lengthening (lengthening vs no lengthening), two preboundary tones (H# vs L#), and two postboundary tones [#LH vs #HL: Recall that tonal markings of IP-initial APs are the same as those of IP-internal APs (Jun 1993, 2000)]. Thus, each target word appeared in eight different combinations of these factors (2×2×2), yielding eight experimental lists.

As in experiment 1, eight experimental lists were arranged such that each subject heard every word just once, in one of the eight conditions. Each list contained the same number of target- and filler-bearing strings, and they were presented in the same pseudo-random order as in experiment 1. No two stimuli with the same prosodic condition were presented in a row.

3. Procedures

The procedure of experiment 2 was the same as that of experiment 1.

B. Results

Missing items, incorrect responses, and RTs over 1500 ms were treated as errors. Mean error rates are summarized in Table III. As was the case with experiment 1, only the results of the accuracy analyses are reported here, as the RT data were not entirely reliable due to high error rates, especially in disyllabic conditions (mean error rate, 48.7%).

The error rates were submitted to repeated measures ANOVAs with the factors syllable count (2 vs 3), preboundary tone (frequent H# vs infrequent L#), postboundary (=target word) tone (frequent #LH vs infrequent #HL), and lengthening (lengthening vs no lengthening).

TABLE III. Mean error rates (%) in experiment 2 with four factors, syllable count, preboundary tone, postboundary (target word) tone, and lengthening. (Standard errors are included in parentheses.)

No. of syllables	Preboundary tone	Postboundary #LH		Postboundary #HL	
		Lengthening	No lengthening	Lengthening	No lengthening
2	H	39.2	30.5	53.4	59.3
		(2.4)	(2.6)	(2.6)	(2.4)
	L	39.9	44.4	57.6	65.2
		(2.5)	(2.4)	(2.4)	(2.2)
3	H	13.8	11.8	15.9	23.2
		(1.7)	(1.3)	(1.7)	(2.2)
	L	11.1	18.7	23.9	29.5
		(1.5)	(1.9)	(2)	(2.3)

All four factors showed significant main effects. Error rates were significantly lower in the trisyllabic condition than in the disyllabic condition ($F1[1,88]=614.74$, $p<0.001$; $F2[1,46]=19.84$, $p<0.001$); when preboundary tone was H# than when it was L# ($F1[1,88]=26.41$, $p<0.001$; $F2[1,46]=20.96$, $p<0.001$); and when postboundary tone was #LH than when it was #HL ($F1[1,88]=167.70$, $p<0.001$; $F2[1,46]=37.86$, $p<0.001$). The effects of the three factors (syllable count, preboundary tone, and postboundary tone) were therefore in line with the results found in experiment 1. The lengthening effect was also significant ($F1[1,88]=11.54$, $p<0.005$; $F2[1,46]=5.169$, $p<0.05$), showing that listeners detected target words more accurately when the preboundary syllable was lengthened than when it was not.

There was an interaction between syllable count and postboundary tone ($F1[1,88]=24.14$, $p<0.001$; $F2[1,46]=5.3$, $p<0.001$). *Posthoc* analyses showed that the interaction stemmed from the differential effect size of postboundary tone depending on the syllable count. The error rates were lower for #LH than for #HL in both disyllabic ($F1[1,95]=95.92$, $p<0.001$; $F2[1,23]=24.57$, $p<0.001$) and trisyllabic conditions ($F1[1,95]=38.57$, $p<0.001$; $F2[1,23]=13.61$, $p=0.001$), but the effect was greater for disyllabic target words (mean difference: 20.4%, eta^2=0.502) than for trisyllabic ones (mean difference: 9.3%, eta^2=0.289) at $p<0.01$ both by-subjects and by-items.

There were also interactions between the lengthening factor and the intonational factors. Lengthening interacted with preboundary tone ($F1[1,88]=7.14$, $p<0.01$; $F2[1,46]=6.92$, $p<0.05$) and with postboundary tone ($F1[1,88]=8.08$, $p<0.01$; $F2[1,46]=5.42$, $p<0.05$). *Poshoc* analyses showed that the lengthening effect was reliable only when the preboundary tone was the infrequent L# ($F1[1,95]=13$, $p<0.001$; $F2[1,47]=10.96$, $p<0.005$) and when the target word tone was the infrequent #HL ($F1[1,95]=15.88$, $p<0.001$; $F2[1,47]=8.7$, $p<0.01$). However, there was also a significant three-way interaction ($F1[1,88]=8.09$, $p<0.01$; $F2[1,46]=6.29$, $p<0.05$). As shown in Fig. 2, the interaction came from the fact that, although lengthening reduced error rates in L#HL, H#HL, and L#LH (all at $p<0.05$ both by-subjects and by-items), the opposite was true for H#LH ($p<0.05$ both by-subjects and by-items): While the presence of final lengthening was generally helpful in most cases, its presence made it harder for listeners to detect target words with the frequent intonation pattern H#LH.

C. Summary and discussion

The results of experiment 2 confirmed the effects of syllable count, preboundary tone, and postboundary tone, consistent with those in experiment 1. In addition, the main effect of lengthening suggests that substantial phrase-final lengthening can also serve as a helpful segmentation cue for Korean listeners. The interaction between lengthening and preboundary tone, however, revealed that the presence of phrase-final lengthening was not useful when the preboundary tone was a frequent AP-final marking tone (H#). It was useful only when the preboundary tone was infrequent L# for AP-final marking. It appears that, when the preboundary L# gives rise to misparsing errors (as L# is likely to be perceived as the AP-initial tone) as observed in experiment 1, final lengthening becomes operative and helps listeners to reduce such misparsing errors.

The three-way interaction in the error rate analysis showed a more detailed interaction between lengthening and intonational cues. In the less frequent tone conditions (H#HL, L#LH, and L#HL), the detection performance was

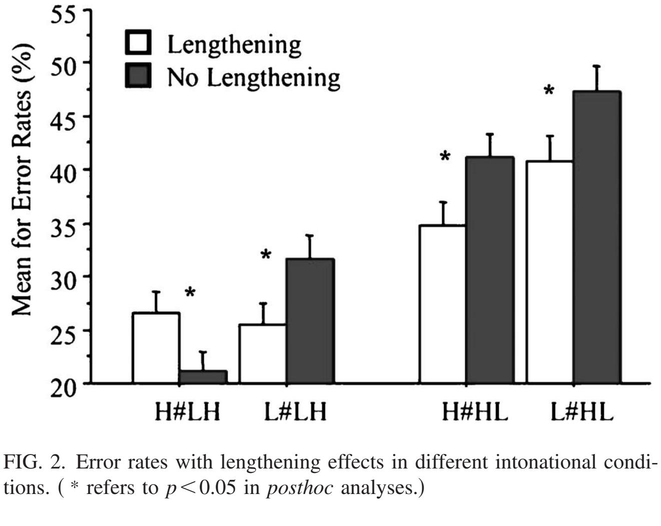


FIG. 2. Error rates with lengthening effects in different intonational conditions. (* refers to $p<0.05$ in *posthoc* analyses.)

better with than without lengthening. But the opposite was true for the most frequent tone condition, H#LH. In this condition, listeners' accuracy was significantly better without lengthening. This result works against our initial hypothesis that listeners would make use of available prosodic cues in a cumulative way in detecting a prosodic boundary, but it appears to be in favor of the alternative hypothesis that the mismatch between prosodic and segmentation information would take precedence over the advantage of the additional final-lengthening cue (see Sec. IV for further discussion on this point).

Finally, it is worth addressing an issue about detection error rates that were found to be very high in both experiments, especially with disyllabic target. Although error rates are generally high in word-spotting tasks (e.g., McQueen, 1996), one might be concerned that such high error rates could reduce the interpretability of findings of the present study. Statistical analyses, however, suggest that our results are still reliably interpretable. The main findings in detection accuracy were statistically robust in both by-subjects and by-items analyses. More crucially, error rates with trisyllabic targets were far lower than those with disyllabic targets, but there were no interactions at all between crucial prosodic factors and the syllable count. That is, patterns found with disyllabic targets were statistically the same as those with trisyllabic targets, indicating that findings of the present study were not biased due to high error rates with disyllabic targets.

IV. GENERAL DISCUSSION

This study examined the role of prosodic cues in online word segmentation of Korean. The prosodic cues under investigation in two word-spotting experiments were language-specific intonational cues and a phrase-final lengthening cue.

The results of experiment 1 showed that listeners detected the target word better when accompanied by intonation patterns frequently associated with AP boundaries (forming an H#L sequence locally at AP boundary). They suggest that intonation patterns which are frequently used in marking an AP-boundary in speech production are indeed exploited by listeners in processing the speech signal. The results also revealed an interesting interaction between the preboundary and the postboundary intonation patterns. The facilitatory frequency effect of the postboundary #L (for both #LH and #LL) was robust only when the preceding tone was also frequent (with H#). Likewise, the facilitatory frequency effect of the preboundary H# was reliable only when the following (postboundary) tone was also frequent (with #L). Moreover, although #LH is far more frequent than #LL, the difference was not significant insofar as the local H#L boundary condition was met. This suggests that listeners process the intonational cues across the prosodic boundary rather than focusing on tonal intonation in preboundary and postboundary positions independently. It also indicates that the local intonation pattern (H#L) across the prosodic boundary is the most crucial information that the listener makes use of in detecting a possible word boundary, but when such a local condition is not met, listeners make further use of other available intonational information (e.g., the second element of the postboundary intonational sequence).

It should be noted that in the present study, we used target words that started with either a lenis stop or a nasal because these consonants (among many others, including vowels) are known to be characteristically associated with an AP-initial rising tone (#LH) (Jun, 1993, 2000). (Recall that the initial tone becomes #HH only when the initial segment is an aspirated or a tense consonant.) Although the tone-segment interaction is thought to be a phonological process (Jun, 1993), it is not inviolable: Other tonal variants do occur, though with low frequency (Jun, 2000; Jun and Fougeron; Kim, 2004). Still, any observed effects of AP-initial tones may not be seen as purely intonational, to the extent that the intonation effect cannot be separated from the effect of the tone-segment interaction. However, purely intonational effects do exist with the preboundary (AP-final) tone, which is not constrained by any segmental information. Taken together, the observed effects can be viewed "intonational" insofar as listeners make use of available intonational information in lexical segmentation, regardless of whether the information is purely intonational or attributable to a result of the complex tone-segment interaction.

In experiment 2, we explored how the intonational cues for AP would interact with the substantial final-lengthening cue that might arise with IP. The results showed that Korean listeners' lexical segmentation of postboundary words is robustly influenced by lengthening and AP-boundary marking intonational cues, but the effects are not entirely cumulative: Preboundary lengthening helps the detection of the following word across the boundary, but only when boundary-adjacent intonational cues are not frequent. When final lengthening was combined with the most frequent boundary-adjacent intonation pattern (H#LH), however, a poorer performance was observed. This suggests that an additional prosodic cue (in this case, phrase-final lengthening) does not always operate in favor of lexical segmentation. Then, a question arises: Why does the presence of final lengthening cue yield asymmetrical results, especially showing the unexpected pattern when intonational cues are frequent for AP boundary percept?

We propose that this is due to a mismatch between segmental and prosodic information in our stimuli. While processing the incoming speech signal, listeners are likely to be able to predict what comes next, based on what they have already heard. Preboundary (phrase-final) lengthening should give a phrase boundary percept to listeners, just as preboundary H# does. When an IP boundary is hypothesized due to substantial (IP-induced) final lengthening, the listener would expect that what comes after the assumed IP boundary would be another IP. If the forthcoming segmental materials across the boundary are perceived as containing phonetic information appropriate for that IP-initial position (e.g., IP-initial strengthening cues), the listener's initial boundary decision would eventually be confirmed.

As discussed in the Introduction, however, appropriate domain-initial strengthening cues for IP-initial position were not included in our speech materials to avoid any potential confounding effects from other possible prosodic cues. In particular, allophonic variation in lenis stops was not matched with IP: Lenis stops are always voiceless when in IP-initial position (a type of domain-initial strengthening) (Cho and Keating, 2001; Jun, 1993), but our speech materials contained the lenis stops with voicing as appropriate for IP-medial rather than IP-initial position. The substantial final lengthening cue which would give rise to an IP-final boundary percept is not matched with AP-induced segmental cues, lacking domain-initial strengthening cues (including allophonic voicing cues) for IP-initial position. Such a mismatch may therefore hinder the segmentation process. This idea can be further explained in terms of the prosody analyzer account of speech perception proposed by Cho *et al.* (2007) and Salverda *et al.* (2003). The prosody analyzer account assumes that suprasegmental and segmental information are processed in parallel. When the computed prosodic boundary based on suprasegmental information is matched with prosodically-driven segmental information for that boundary, lexical segmentation is facilitated, whereas a mismatch between them is predicted to hinder lexical segmentation.

This also has another implication for the use of multiple phrase-level prosodic cues in lexical segmentation process. Previous studies have suggested that available lexical segmentation cues are used immediately in an exhaustive and cumulative way (e.g., Spitzer *et al.*, 2007; Donselaar *et al.*, 2005; Norris *et al.*, 1997). However, our results suggest that not all the available segmentation cues are necessarily exploited by the listener in a straightforwardly cumulative way. Instead, in detecting prosodic boundaries, listeners appear to use available prosodic cues with differential weighting: When the local boundary-marking information is not sufficient to finalize the decision about the location of a prosodic boundary (as with L#L), additional information (such as preboundary lengthening and the second element of the postboundary intonational sequence, as in #LH) is further utilized by the listener. The relative use of available phrase-level prosodic cues is reminiscent of the hierarchical model of speech segmentation proposed by Mattys *et al.* (2005). It assumes that available segmentation cues are weighted, such that higher order knowledge (e.g., lexical information and contextual information) takes precedence over sublexical information. The sublexical cues then become operative in a non-optimal communicative condition so that segmental cues are first called upon and word-internal metrical prosodic cues are used as a last resort. Although the model does not deal with how phrase-level prosodic cues are utilized in lexical segmentation, multiple phrase-level prosodic cues may well be used with relative weights. For example, the local boundary-marking intonational cues (H#L) may well be weighted above other boundary-adjacent prosodic cues.

We are now left with a question about how the use of phrase-level prosodic information in lexical segmentation may be incorporated into existing models of speech segmentation such as the hierarchical model (Mattys *et al.*, 2005), or the shortlist model (Norris, 1994; Norris and McQueen, 2008). While addressing this issue is beyond the scope of the present study, one possible way is to allow a module such as the prosody analyzer whose function is to compute prosodic boundaries using multiple prosodic cues, along with their relative weights. The prosodic structure computed in this way can then serve as part of the high order knowledge in the hierarchical model. Or, it can be checked against computed lexical boundaries in a model like the shortlist model, such that lexical competition is further modulated by the alignment between the computed prosodic structure and the potential lexical boundary.

Finally, the results of the present study have implications for the language-specific vs cross-linguistic use of perceptual salient rising F0 (or high pitch) in lexical segmentation. Given that an F0 rise is a potential prosodic cue in lexical segmentation in various languages, Warner *et al.* (2009) suggested that F0 rise is perceptually salient, and the presence of F0 rise word-initially is likely to facilitate lexical segmentation cross-linguistically. Our results, however, showed the opposite: Korean listeners found it difficult to detect the target word when it starts with H surrounded by L tones as in L#HL.

We propose that although substantial F0 rise may be universally perceptually salient, how such an intonational element is aligned with segmental content is determined by language-specific intonational phonology. The intonational element of H in Korean, for example, is aligned with either the second or the final syllable of AP (i.e., #LH···LH#), and thus H on the first syllable of the target word hinders lexical segmentation process. But in French, H is phonologically specified to be aligned with the first syllable of the content word in AP (Welby, 2007), and thus it facilitates lexical segmentation process in a way that conforms to French intonational phonology. Shukla *et al.* (2007) showed that Italian listeners were able to exploit Japanese IP boundaries in word segmentation and claimed that prosody contains universal cues for lexical segmentation. Again, what is "universal" here may be that some prosodic cues are used cross-linguistically: Italian listeners are likely to have exploited Japanese IP boundaries not because the particular prosodic cues that were available to them were universal, but because Italian and Japanese IP happen to share some prosodic cues. Our interpretations are in line with previous studies on word segmentation in general: The set of segmentation cues is language-universal, but the detailed manifestations of individual segmentation cues that listeners exploit are language-specific. It is well known that listeners use phonotactic, allophonic, and various other cues for word segmentation, but the exploitation of those cues is sensitive to the phonological constraints and structure of a given language (Cutler *et al.*, 2002; Cutler and Norris, 1988; Cutler and Otake, 1994; Mehler *et al.*, 1981; Sebastián-Gallés *et al.*, 1992; Weber, 2001). Likewise, the claim that prosodic boundaries constrain online lexical search (Christophe *et al.*, 2004; Shukla *et al.*, 2007) may be applicable cross-linguistically, but the way the prosodic boundary is phonetically manifested in speech production and the way that listeners exploit the prosodic boundary cues in speech comprehension must be language-specific.

V. CONCLUSION

The present study investigated the role of phrase-level prosodic cues in word segmentation of Korean. In experiment 1, we found that listeners make use of a local intonation pattern at a prosodic (AP) boundary (i.e., H#L) in online lexical segmentation, even when other important boundary-marking cues (such as domain-initial strengthening and final lengthening) are not present in the speech signal. The locality condition of H#L at the boundary suggests that listeners generally pay more attention to information straddling the prosodic boundary rather than the global intonational contour within a phrase, indicating that the boundary detection is crucial in lexical segmentation. In experiment 2, an additional final lengthening cue was found to help listeners with lexical segmentation when intonation patterns are not frequent for marking a prosodic boundary. However, when the lengthening cue was combined with the most frequent intonation pattern of H#L, creating a percept of IP, the mismatch between prosodic cues (appropriate for IP) and domain-initial strengthening cues (including allophonic cues of lenis stops, appropriate for IP-medial position) appears to take precedence over the cumulative effect. The hypothesized cumulative effect would work only if the computed prosodic boundary is matched with expected segmental allophonic variation. A follow-up study is necessary to corroborate this claim, in order to examine the interaction between effects of prosodic variation and allophonic variation in lexical segmentation. It also remains to be seen how the use of phrase-level prosodic information in spoken word recognition may be implemented within current models of speech segmentation. But following the prosody analyzer account, we propose that prosodic information is computed in parallel with segmental information, and lexical segmentation is modulated by the interaction between the two kinds of information. More generally, our study builds up on the growing body of psycholinguistic research which highlights the important roles that prosody plays in both speech production and speech comprehension: Speakers generate a prosodic structure online in which a given utterance is organized into prosodic units, and its expected acoustic-phonetic cues are in turn exploited by listeners in online lexical segmentation.

ACKNOWLEDGMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (Contract No. KRF-2007-332-A00102). We thank the reviewers and the associate editor for their constructive comments; Adam Albright for his suggestions; and Eunbin Lee, Jiyoung Park, Soomin Yoo, and Yuna Hur for their assistance with data collection.

APPENDIX: LIST OF TARGET WORDS AND CARRIER STRINGS

The following shows the list of disyllabic and trisyllabic target words and carrier strings.

a. Disyllabic Target Words

Target Words (Phonemic Transcription)	Glosses	Carrier Strings (Phonemic Transcription)
/ka.wi/	scissors	/ma.pɛ.tʃo.**ka.wi**.nɛ.li/
/kɛ.mi/	ant	/tʃʌ.ku.ta.**kɛ.mi**.sʌ.tu/
/kʌ.li/	street	/na.so.lɛ.**kʌ.li**. tʃʌ.p^{h}u/
/ko.kɛ/	hill	/ta.sʌ.ni.**ko.kɛ**.tʃu.pi/
/ko.ki/	meat	/tu.tʃɛ.ma.**ko.ki**.na.po/
/ku.tu/	shoes	/po.na.tʃa.**ku.tu**.mi.pɛ/
/ki.to/	prayer	/tʌ.so.nʌ.**ki.to**.ju.ma/
/na.ra/	nation	/nu.jʌ.ko.**na.la**.kɛ.tʃi/
/na.mu/	tree	/tʃo.la.su.**na.mu**.kʌ.to/
/no.lɛ/	song	/pɛ.mi.tʃʌ.**no.lɛ**.pa.ki/
/ta.li/	leg/bridge	/ku.tʃʌ.mo.**ta.li**.pɛ.la/
/to.si/	city	/tsi.ma.pɛ.**to.si**.mʌ.tsɛ/
/tu.pu/	tofu	/tʃo.nu.sɛ.**tu.pu**.nɛ.ma/
/ma.lu/	floor	/tʃu.ka.pʌ.**ma.lu**.ki.ta/
/mʌ.li/	head	/tɛ.tʃu.pi.**mʌ.li**.tʃo.hɛ/
/mo.ki/	mosquito	/ta.mi.tʃʌ.**mo.ki**.pɛ.ki/
/mo.tʃa/	hat	/tʃʌ.pɛ.**mo.tʃa**.so.tu/
/mu.kɛ/	weight	/pa.mʌ.lo.**mu.kɛ**.nʌ.sɛ/
/mi.lɛ/	future	/kjo.tʃʌ.tu.**mi.lɛ**.po.ku/
/mi.so/	smile	/tʃa.pɛ.lo.**mi.so**.nʌ.ku/
/pa.ta/	sea	/pu.sɛ.tʃi.**pa.ta**.mʌ.tʃu/
/pa.tʃi/	pants	/ma.jʌ.ku.**pa.tʃi**.ni.mɛ/
/pu.tʃa/	a rich person	/ti.lɛ.mi.**pu.tʃa**.lʌ.ko/
/pi.nu/	soap	/pa.tɛ.mʌ.**pi.nu**.mɛ.tʃʌ/

b. Trisyllabic Target Words

Target Words (Phonemic Transcription)	Glosses	Carrier Strings (Phonemic Transcription)
/ko.ku.ma/	sweet potato	/no.tʃu.pʌ.**ko.ku.ma**.li.tʃa/
/ku.tʌ.ki/	maggot	/ko.ta.po.**ku.tʌ.ki**.pa.tɛ/
/ki.lʌ.ki/	wild goose	/mu.tʃa.nʌ.**ki.lʌ.ki**.no.kɛ/
/na.nu.ki/	division	/ta.lu.tʃo.**na.nu.ki**.mʌ.no/
/na.tɨ.li/	outing	/pi.ma.ko.**na.tɨ.li**.ku.tʃʌ/
/na.mʌ.tʃi/	remainder	/mɛ.pi.tʃu.**na.mʌ.tʃi**.mi.sʌ/
/no.ta.tʃi/	a gold mine	/tʃa.pʌ.ku.**no.ta.tʃi**.ma.la/
/nu.tʌ.ki/	rag	/mi.ta.po.**nu.tʌ.ki**.tʃu.mi/
/tʌ.tɨ.mi/	antenna	/mo.na.tʃu.**tʌ.tɨ.mi**.pʌ.sɛ/
/to.ka.ni/	pot	/ma.la.pe.**to.ka.ni**.mʌ.ha/
/to.kɛ.pi/	elf	/nu.tʃɛ.**to.k*ɛ.pi**.ma.sʌ/
/to.ka.tʃi/	bellflower	/kʌ.pi.tʃo.**to.la.tʃi**.pʌ.ta/
/to.t^{h}o.li/	acorn	/tʃa.mi.kʌ.**to.t**^{h}**o.li**.pa.tɛ/
/tu.k*ʌ.pi/	toad	/mo.la.tʃɛ.**tu.k*ʌ.pi**.ta.tʃo/
/tu.tʌ.tʃi/	mole	/ni.thɛ.mʌ.**tu.tʌ.tʃi**.la.kɛ/
/tu.lu.mi/	crane	/ka.tʌ.pi.**tu.lu.mi**.po.tʃu/
/ma.nu.la/	wife	/pɛ.li.no.**ma.nu.la**.tʃi.pʌ/
/mɛ.t*u.ki/	grasshopper	/tʃɛ.ni.ka.**mɛ.t*u.ki**.tu.pa/

b. Trisyllabic Target Words (cont.)

Target Words (Phonemic Transcription)	Glosses	Carrier Strings (Phonemic Transcription)
/mjʌ.nɨ.li/	daughter-in-law	/ti.kε.no.**mjʌ.nɨ.li**.pa.tʃu/
/mu.tʌ.ki/	pile	/nu.mi.pa.**mu.tʌ.ki**.sʌ.nε/
/pa.ku.ni/	basket	/pε.ta.tʃo.**pa.ku.ni**.mʌ.la/
/po.t*a.li/	bundle	/ku.mo.sʌ.**po.t*a.li**.tε.mjʌ/
/tʃε.tʃ^{h}ε.ki/	sneeze	/tu.pa.mʌ.**tʃ.tʃ^{h}ε.ki**.pa.tʃu/
/tʃu.mʌ.ni/	pocket	/tʃo.ti.**tʃu.mʌ.ni**.mo.li/

[1] Intervocalic voicing of lenis stops is generally held to apply within the AP; however, even within the AP, it does not apply in every single instance. A reviewer suggested that this might weaken the argument that AP is a phonologically-motivated categorical prosodic unit. However, the phonological nature of AP cannot be determined simply by the observed gradient nature of voicing within an AP. This is because a process may be gradient or probabilistic, even if the necessary pre-conditions for application (i.e., internal to the AP) are categorically present (see Zsiga, 1995; Cohn, 1998; Fourakis and Port, 1986 for discussion on phonetic vs phonological processes). Jun (1995) indeed noted this and suggested that the lenis stop intervocalic voicing rule may be seen as a gradient phonetic process. More importantly, however, whether AP is phonetic or phonological is beyond the scope of the present study. What is critical in the present study is that the most extensively received model (Jun, 1993, 2000) is adopted here as a framework within which effects of intonation and durational cues on lexical segmentation can be tested: Even if the model dividing Korean phrases into a small number of discrete categories such as AP and IP were not theoretically impregnable, the results of our experiments could easily be interpreted in a model that makes use of a more flexible prosodic model since our study actually measures the effects of acoustic properties (lengthening and intonation), not of phrase structures like "AP" or "IP" directly.

[2] The results of splicing were checked by the two authors to ensure that no discernible splicing artifact remained in the speech signal. Nevertheless, as pointed out by a reviewer, it is possible that subtle discontinuities may remain and serve as a segmentation cue (Johnson and Jusczyk, 2001; Mattys, 2004). However, because the same spliced token was used in all test conditions, any observed differences between prosodic conditions (e.g., frequent vs infrequent intonational conditions) must be attributed to the prosodic manipulation, not to splicing artifacts.

Beckman, M. E. (**1996**). "The parsing of prosody," Lang. Cognit. Processes **11**, 17–67.

Beckman, M. E., and Pierrehumbert, J. B. (**1986**). "Intonational structure in Japanese and English," Phonology Yearbook **3**, 255–309.

Cho, T., and Keating, P. (**2001**). "Articulatory and acoustic studies of domain-initial strengthening in Korean," J. Phonetics **29**, 155–190.

Cho, T., McQueen, J. M., and Cox, E. (**2007**). "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," J. Phonetics **35**, 210–243.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (**2004**). "Phonological phrase boundaries constrain lexical access: I. Adult data," J. Mem. Lang. **51**, 523–547.

Chung, K., Chang, S., Choi, J., Nam, S., Lee, M., Chung, S., Koo, H., Kim, K., Kim, J., Lee, C., Han, S., Oh, M., Song, M., Hong, S., and Jee, S. (**1996**). *A Study of Korean Prosody and Discourse for the Development of Speech Synthesis/Recognition System* (KAIST Artificial Intelligence Research Center, Daejon, Korea).

Cohn, A. (**1998**). "The phonetics-phonology interface revisited: Where's phonetics?," in Texas Linguistics Society 1998 Conference Proceedings, pp. 25–40.

Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang Speech **45**, 208–228.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.

Cutler, A., Demuth, K., and McQueen, J. M. (**2002**). "Universality versus language specificity in listening to running speech," Psychol. Sci. **13**, 258–262.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Mem. Lang. **33**, 824–844.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (**2001**). "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition," Lang. Cognit. Processes **16**, 507–534.

Donselaar, W. V., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. **58**, 251–274.

Edwards, J., Beckman, M. E., and Fletcher, J. (**1991**). "Articulatory kinematics of final lengthening," J. Acoust. Soc. Am. **89**, 369–382.

Forster, K., and Chambers, S. (**1973**). "Lexical access and naming time," J. Verbal Learn. Verbal Behav. **12**, 627–635.

Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

Fourakis, M., and Port, R. (**1986**). "Stop epenthesis in English," J. Phonetics **14**, 197–221.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1997**). "Integrating form and meaning: A distributed model of speech perception," Lang. Cognit. Processes **12**, 613–656.

Gee, J. P., and Grosjean, F. (**1983**). "Performance structures: A psycholinguistic and linguistic appraisal," Cogn. Psychol. **15**, 411–458.

Gow, D. W. (**2002**). "Does English coronal assimilation create lexical ambiguity?," J. Exp. Psychol. Hum. Percept. Perform. **28**, 163–179.

Gow, D. W., and Gordon, P. (**1995**). "Lexical and prelexical influences on word segmentation: Evidence from priming," J. Exp. Psychol. Hum. Percept. Perform. **21**, 344–359.

Johnson, E. K., and Jusczyk, P. W. (**2001**). "Word segmentation by 8-month-olds: When speech cues count more than statistics," J. Mem. Lang. **44**, 548–567.

Jun, S.-A. (**1993**). "The phonetics and phonology of Korean prosody," Ph.D. dissertation, The Ohio State University, Columbus, OH.

Jun, S.-A. (**1995**). "Asymmetrical prosodic effects on the laryngeal gesture in Korean," in *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 235–253.

Jun, S.-A. (**1998**). "The accentual phrase in the Korean prosodic hierarchy," Phonology **15**, 189–226.

Jun, S.-A. (**2000**). "K-ToBI (Korean ToBI) labelling conventions (Version 3.1)," http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html (Last accessed 16 March 2009).

Jun, S.-A., and Fougeron, C. (**2000**). "A phonological model of French intonation," in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic, Boston, MA), pp. 209–242.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (**2003**). "Domain-initial strengthening in four languages," in *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 145–163.

Keating, P., and Shattuck-Hufnagel, S. (**2002**). "A prosodic view of word form encoding for speech production," UCLA Working Papers in Phonetics **101**, 112–156, University of California, Los Angeles, CA.

Kim, S.-J. (**2000**). "Accentual effect on segmental phonological rules in Korean," Ph.D. dissertation, University of North Carolina, Chapel Hill, NC.

Kim, S. (**2004**). "The role of prosodic phrasing in Korean word segmentation," Ph.D. dissertation, University of California, Los Angeles, CA.

Kim, H.-G., and Kang, B.-M. (**2004**). *Frequency Analysis of Korean Morpheme and Word Usage II* (Institute of Korean Culture, Korea University, Seoul, Korea).

Krivokapić, J. (**2007**). "Prosodic planning: Effects of phrasal length and complexity on pause duration," J. Phonetics **35**, 162–179.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).

Lehiste, I. (**1960**). "An acoustic-phonetic study of internal open juncture," Phonetica **5**, 1–54.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge, MA).

Lim, B. J., and de Jong, K. J. (**1999**). "Tonal alignment in Seoul Korean," J. Acoust. Soc. Am. **106**(4), 2152.

Luce, P. A. (**1986**). "Neighborhoods of words in the mental lexicon," Research on Speech Perception Technical Report No. 6, National Institutes of Health, Bloomington, IN.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.

Marslen-Wilson, W., and Welsh, A. (**1978**). "Processing interactions during word-recognition in continuous speech," Cogn. Psychol. **10**, 29–63.

Mattys, S. L. (**2004**). "Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation," J. Exp. Psychol. Hum. Percept. Perform. **30**, 397–408.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.

McQueen, J. M. (**1996**). "Word spotting," Lang. Cognit. Processes **11**, 695–699.

McQueen, J. M. (**2005**). "Speech perception," in *The Handbook of Cognition*, edited by K. Lamberts and R. Goldstone (Sage, London), pp. 255–275.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., and Segui, J. (**1981**). "The syllable's role in speech segmentation," J. Verbal Learn. Verbal Behav. **20**, 298–305.

Nespor, M., and Vogel, I. (**1986**). *Prosodic Phonology* (Foris, Dordrecht).

Norris, D. (**1986**). "Word recognition: Context effects without priming," Cognition **22**, 93–136.

Norris, D. (**1994**). "Shortlist: A connectionist model of continuous speech recognition," Cognition **52**, 189–234.

Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: A Bayesian model of continuous speech recognition," Psychol. Rev. **115**, 357–395.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). "The possible-word constraint in the segmentation of continuous speech," Cogn. Psychol. **34**, 191–243.

Park, K. C. (**2004**). "The phrase-initial high in Korean," Studies in Phonetics, Phonology and Morphology **10**, 203–223.

Quené, H. (**1993**). "Segment durations and accent as cues to word segmentation in Dutch," J. Acoust. Soc. Am. **94**, 2027–2035.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996**). "Word segmentation: The role of distributional cues," J. Mem. Lang. **35**, 606–621.

Salverda, A. P., Dahan, D., and McQueen, J. M. (**2003**). "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," Cognition **90**, 51–89.

Schreuder, R., and Baayen, R. H. (**1994**). "Prefix stripping re-revisited," J. Mem. Lang. **33**, 357–375.

Sebastián-Gallés, N., Dupoux, E., Segui, J., and Mehler, J. (**1992**). "Contrasting syllabic effects in Catalan and Spanish," J. Mem. Lang. **31**, 18–32.

Selkirk, E. (**1984**). *Phonology and Syntax: The Relation Between Sound and Structure* (MIT Press, Cambridge).

Shattuck-Hufnagel, S., and Turk, A. (**1996**). "A prosody tutorial for investigators of auditory sentence processing," J. Psycholinguist. Res. **25**, 193–247.

Shatzman, K. B., and McQueen, J. M. (**2006**). "Segment duration as a cue to word boundaries in spoken-word recognition," Percept. Psychophys. **68**, 1–16.

Shukla, M., Nespor, M., and Mehler, J. (**2007**). "An interaction between prosody and statistics in the segmentation of fluent speech," Cogn. Psychol. **54**, 1–32.

Soto, S., Sebastián-Gallés, N., and Cutler, A. (**2001**). "Segmental and suprasegmental mismatch in lexical access," J. Mem. Lang. **45**, 412–432.

Spitzer, S. M., Liss, J. M., and Mattys, S. L. (**2007**). "Acoustic cues to lexical segmentation: A study of resynthesized speech," J. Acoust. Soc. Am. **122**, 3678–3687.

Tabossi, P., Collina, S., Mazzetti, M., and Zoppello, M. (**2000**). "Syllables in the processing of spoken Italian," J. Exp. Psychol. Hum. Percept. Perform. **25**, 758–775.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (**1997**). "Phonotactic and syllable stress: Implications for the processing of spoken nonsense words," Lang Speech **40**, 47–62.

Vroomen, J., and de Gelder, B. (**1995**). "Metrical segmentation and lexical inhibition in spoken word recognition," J. Exp. Psychol. Hum. Percept. Perform. **23**, 710–720.

Warner, N., Otake, T., and Takayuki, A. (**2009**). "Intonational structure as a word boundary cue in Japanese," Lang Speech in press.

Weber, A. (**2001**). "Language-specific listening: The case of phonetic sequences," Ph.D. dissertation, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Welby, P. (**2007**). "The role of early fundamental frequency rises and elbows in French word segmentation," Speech Commun. **49**, 28–48.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. (**1992**). "Segmental durations in the vicinity of prosodic phrase boundaries," J. Acoust. Soc. Am. **91**, 1707–1717.

Zsiga, E. C. (**1995**). "An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 282–302.

Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions

Jianfen Ma[a)]
College of Computer Engineering and Software, Taiyuan University of Technology, Shanxi 030024, China and Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

Yi Hu and Philipos C. Loizou[b)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

(Received 8 August 2008; revised 11 December 2008; accepted 14 February 2009)

The articulation index (AI), speech-transmission index (STI), and coherence-based intelligibility metrics have been evaluated primarily in steady-state noisy conditions and have not been tested extensively in fluctuating noise conditions. The aim of the present work is to evaluate the performance of new speech-based STI measures, modified coherence-based measures, and AI-based measures operating on short-term (30 ms) intervals in realistic noisy conditions. Much emphasis is placed on the design of new band-importance weighting functions which can be used in situations wherein speech is corrupted by fluctuating maskers. The proposed measures were evaluated with intelligibility scores obtained by normal-hearing listeners in 72 noisy conditions involving noise-suppressed speech (consonants and sentences) corrupted by four different maskers (car, babble, train, and street interferences). Of all the measures considered, the modified coherence-based measures and speech-based STI measures incorporating signal-specific band-importance functions yielded the highest correlations ($r=0.89-0.94$). The modified coherence measure, in particular, that only included vowel/consonant transitions and weak consonant information yielded the highest correlation ($r=0.94$) with sentence recognition scores. The results from this study clearly suggest that the traditional AI and STI indices could benefit from the use of the proposed signal- and segment-dependent band-importance functions.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097493]

PACS number(s): 43.72.Ar, 43.72.Dv [DOS] Pages: 3387–3405

I. INTRODUCTION

A number of measures have been proposed to predict speech intelligibility in the presence of background noise. Among these measures, the articulation index (AI) (French and Steinberg, 1947; Fletcher and Galt, 1950; Kryter, 1962a, 1962b) and speech-transmission index (STI) (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985) are by far the most commonly used today for predicting speech intelligibility in noisy conditions. The AI measure was further refined to produce the speech intelligibility index (SII) (ANSI, 1997). The SII measure is based on the idea that the intelligibility of speech depends on the proportion of spectral information that is audible to the listener and is computed by dividing the spectrum into 20 bands (contributing equally to intelligibility) and estimating the weighted average of the signal-to-noise ratios (SNRs) in each band (Kryter, 1962a, 1962b; Pavlovic, 1987; Allen, 1994; ANSI, 1997). The SNRs in each band are weighted by band-importance functions (BIFs) which differ across speech materials (ANSI, 1997). The SII measure has been shown to predict successfully the effects of linear filtering and additive noise on speech intelligibility (e.g., Kryter, 1962a, 1962b). It has, however, a number of limitations. For one, the computation of the SII measure requires as input the levels of speech and masker signals at the eardrum of the listeners, something that might not be available in situations wherein we only have access to recorded (digitized) processed signals. Second, the SII measure has been validated for the most part only for steady (stationary) masking noise since it is based on the long-term average spectra (computed over 125-ms intervals) of the speech and masker signals. As such, it cannot be applied to situations in which speech is embedded in fluctuating maskers (e.g., competing talkers). Several attempts have been made to extend the SII measure to assess speech intelligibility in fluctuating maskers (Rhebergen *et al.*, 2005, 2006; Kates, 1987). Rhebergen *et al.* (2006), for instance, proposed to divide the speech and masker signals into short frames (9–20 ms), evaluate the instantaneous AI value in each frame, and average the computed AI values across all frames to produce a single AI metric. Their extended short-term AI (AI-ST) measure was found to predict speech intelligibility better than the traditional AI measure when evaluated with sentences embedded in artificial masking signals (e.g., periodically interrupted noise) and speech-like maskers, but the predictions with the latter maskers were found to be less accurate (Rhebergen and Versfeld, 2005).

[a)]Work done while Dr. Jianfen Ma visited Professor Loizou's laboratory as a research scholar.

[b)]Author to whom correspondence should be addressed. Electronic mail: loizou@utdallas.edu

Other extensions to the SII measure were proposed by Kates and Arehart (2005) for predicting the intelligibility of peak-clipping and center-clipping distortions in the speech signal, such as those found in hearing aids. The modified index, called the CSII index, used the base form of the SII procedure, but with the SNR estimate replaced by the signal-to-distortion ratio, which was computed using the coherence function between the input and processed signals. While a modest correlation was obtained with the CSII index, a different version was proposed that divided the speech segments into three level regions and computed the CSII index separately for each level region. The three-level CSII index yielded higher correlations for both intelligibility and subjective quality ratings (Arehart *et al.*, 2007) of hearing-aid type of distortions. Further testing of the CSII index is performed in the present study to examine whether it can be used (1) to predict the intelligibility of speech corrupted by fluctuating maskers and (2) to predict the intelligibility of noise-suppressed speech containing different types of non-linear distortions than those introduced by hearing aids.

The STI measure (Steeneken and Houtgast, 1980) is based on the idea that the reduction in intelligibility caused by additive noise or reverberation distortions can be modeled in terms of the reduction in temporal envelope modulations. The STI metric has been shown to predict successfully the effects of reverberation, room acoustics, and additive noise (e.g., Steeneken and Houtgast, 1982; Houtgast and Steeneken, 1985). It has also been validated in several languages (Anderson and Kalb, 1987; Brachmanski, 2004). In its original form (Houtgast and Steeneken, 1971), the STI measure used artificial signals (e.g., sinewave-modulated signals) as probe signals to assess the reduction in signal modulation in a number of frequency bands and for a range of modulation frequencies (0.6–12.5 Hz) known to be important for speech intelligibility. When speech is subjected, however, to non-linear processes such as those introduced by dynamic envelope compression (or expansion) in hearing aids, the STI measure fails to successfully predict speech intelligibility since the processing itself might introduce additional modulations which the STI measure interprets as increased SNR (Hohmann and Kollmeieir, 1995; Ludvigsen *et al.*, 1993; van Buuren *et al.*, 1999; Goldsworthy and Greenberg, 2004). For that reason, several modifications have been proposed to use speech or speech-like signals as probe signals in the computation of the STI measure (Steeneken and Houtgast, 1980; Ludvigsen *et al.*, 1990). Despite these modifications, several studies have reported that the speech-based STI methods fail to predict the intelligibility of nonlinearly-processed speech (van Buuren *et al.*, 1999; Goldsworthy and Greenberg, 2004). Several modifications were made by Goldsworthy and Greenberg (2004) to existing speech-based STI measures but none of these modifications were validated with intelligibility scores obtained with human listeners.

The SII and speech-based STI measures can account for linear distortions introduced by filtering and additive noise, but have not been tested extensively in conditions wherein non-linear distortions might be present, such as when speech is processed via hearing-aid algorithms or noise-suppression algorithms. Some of the noise-suppression algorithms (e.g., spectral subtractive), for instance, can introduce non-linear distortions in the signal and unduly increase the level of modulation in the temporal envelope (e.g., Goldsworthy and Greenberg, 2004). The increased modulation might be interpreted as increased SNR by the STI measure. Hence, it remains unclear whether the speech-based STI measures or the SII measure can account for the type of distortions introduced by noise-suppression algorithms and to what degree they can predict speech intelligibility. It is also not known whether any of the numerous objective measures that have been proposed to predict speech quality (Quackenbush *et al.*, 1988; Loizou, 2007, Chap. 10; Hu and Loizou, 2008) in voice communications applications can be used to predict speech intelligibility. An objective measure that would predict well both speech intelligibility and quality would be highly desirable in voice communication and hearing-aid applications. The objective quality measures are primarily based on the idea that speech quality can be modeled in terms of differences in loudness between the original and processed signals (e.g., Bladon and Lindblom, 1981) or simply in terms of differences in the spectral envelopes [e.g., as computed using a linear predictive coding (LPC) model] between the original and processed signals. The perceptual evaluation of speech quality (PESQ) objective measure (ITU-T, 2000; Rix *et al.*, 2001), for instance, assesses speech quality by estimating the overall loudness difference between the noise-free and processed signals. This measure has been found to predict very reliably ($r > 0.9$) the quality of telephone networks and speech codecs (Rix *et al.*, 2001) as well as the quality of noise-suppressed speech (Hu and Loizou, 2008). Only a few studies (Beerends *et al.*, 2004, 2005) have tested the PESQ measure in the context of predicting speech intelligibility. High correlation ($r > 0.9$) was reported, but it was for a relatively small number of noisy conditions which included speech processed via low-rate vocoders (Beerends *et al.*, 2005) and speech processed binaurally via beamforming algorithms (Beerends *et al.*, 2004). The speech distortions introduced by noise-suppression algorithms (based on single-microphone recordings) differ, however, from those introduced by low-rate vocoders. Hence, it is not known whether the PESQ measure can predict reliably the intelligibility of noise-suppressed speech containing various forms of non-linear distortions, such as musical noise.

The aim of the present work is two-fold: (1) to evaluate the performance of conventional objective measures originally designed to predict speech quality and (2) to evaluate the performance of new speech-based STI measures, modified coherence-based measures (CSII), as well as AI-based measures that were designed to operate on short-term (20–30 ms) intervals in realistic noisy conditions. A number of modifications to the speech-based STI, coherence-based, and AI measures are proposed and evaluated in this study. Much focus is placed on the development of band-importance weighting functions which can be used in situations wherein speech is corrupted by fluctuating maskers. This is pursued with the understanding that a single BIF,

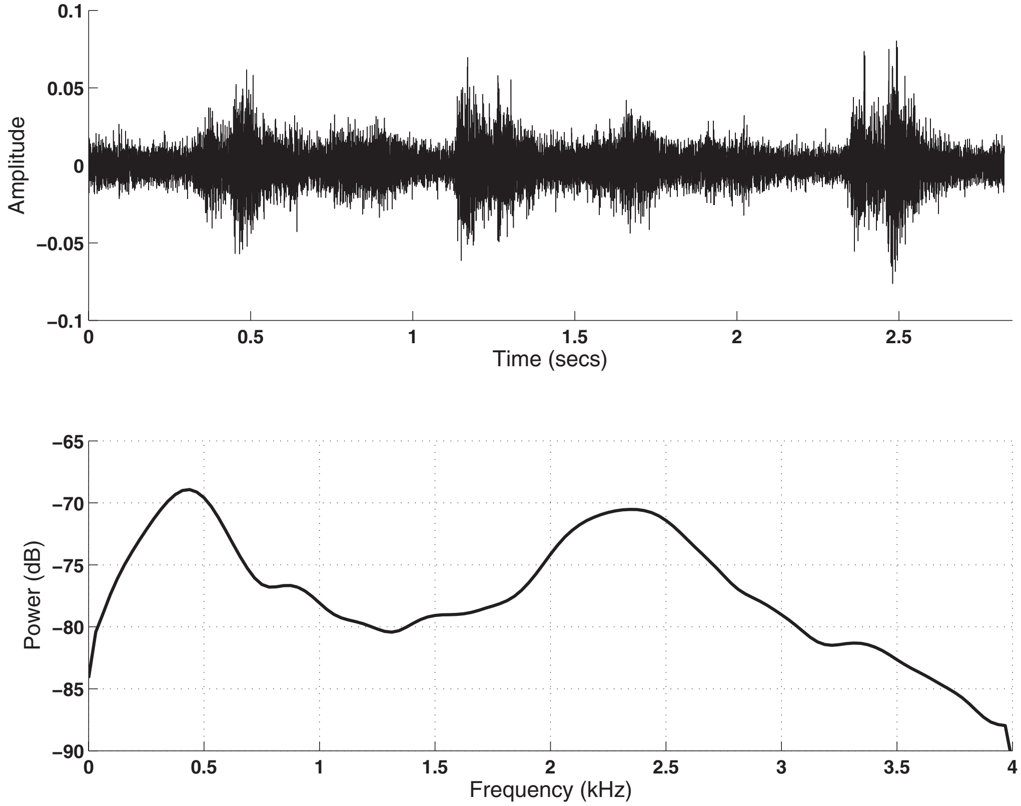


FIG. 1. Waveform (top panel) and long-term averaged spectrum (bottom panel) of the train noise used in the present study.

such as those used in STI and SII indices (ANSI, 1997), might not suitable for evaluating the intelligibility of speech embedded in fluctuating maskers.

II. METHODS

The intelligibility evaluation of noise-corrupted speech processed through eight different noise-suppression algorithms was reported in Hu and Loizou (2007) and is summarized briefly below.

A. Materials and subjects

IEEE sentences (IEEE, 1969) and consonants in/a C a/ format were used as test material. The consonant test included 16 consonants recorded in /a C a/ context, where C =/p,t,k,b,d,g,m,n,dh,l,f,v,s,z,sh,dj/. All consonants were produced by a female speaker, and all sentences were produced by a male talker. The sentences and consonants were originally sampled at 25 kHz and downsampled to 8 kHz. These recordings are available in Loizou (2007). The maskers were artificially added to the speech material. The masker signals were taken from the AURORA database (Hirsch and Pearce, 2000) and included the following real-world recordings from different places: babble, car, street, and train. Figure 1 shows the time-domain waveform and long-term average spectrum of the train noise, illustrating the modulating nature of this masker. The maskers were added to the speech signals at SNRs of 0 and 5 dB.

A total of 40 native speakers of American English were recruited for the sentence intelligibility tests, and 10 additional listeners were recruited for the consonant tests. All subjects were paid for their participation.

B. Noise reduction algorithms

The noise-corrupted sentences were processed by eight different noise-reduction algorithms which included the generalized subspace approach (Hu and Loizou, 2003), the perceptually-based subspace approach (Jabloun and Champagne, 2003), the log minimum mean square error (logMMSE) algorithm (Ephraim and Malah, 1985), the logMMSE algorithm with speech-presence uncertainty (Cohen and Berdugo, 2002), the spectral subtraction algorithm based on reduced-delay convolution (Gustafsson *et al.*, 2001), the multiband spectral-subtractive algorithm (Kamath and Loizou, 2002), the Wiener filtering algorithm based on wavelet-thresholded multitaper spectra (Hu and Loizou, 2004), and the traditional Wiener algorithm (Scalart and Filho, 1996). With the exception of the logMMSE-SPU algorithm which was provided by the authors (Cohen and Berdugo, 2002), all other algorithms were based on our own implementation. The parameters used in the implementation of these algorithms were the same as those published. MATLAB implementations of all noise reduction algorithms tested in the present study are available in Loizou (2007).

C. Procedure

A total of 40 native speakers of American English were recruited for the sentence intelligibility tests. The 40 listeners were divided into four panels (one per type of noise), with each panel consisting of 10 listeners. Each subject participated in a total of 19 listening conditions (=2 SNR levels × 8 algorithms + 2 noisy references + 1 quiet). Two IEEE sentence lists (ten sentences per list) were used for each condition, and none of the sentence lists were repeated. Additional ten listeners were recruited for the consonant recognition task. Subjects were presented with six repetitions of each consonant in random order. The processed speech files (sentences/consonants), along with the clean and noisy speech files, were presented monaurally to the listeners in a double-walled sound-proof booth (Acoustic Systems, Inc.) via Sennheiser's (HD 250 Linear II) circumaural headphones at a comfortable level.

The intelligibility study by Hu and Loizou (2007) produced a total of 72 noisy conditions including the noise-corrupted (unprocessed) conditions. The 72 conditions included distortions introduced by 8 different noise-suppression algorithms operating at two SNR levels (0 and 5 dB) in four types of real-world environments (babble, car, street, and train). The intelligibility scores obtained in the 72 conditions were used in the present study to evaluate the predictive power of a number of old and newly proposed objective measures.

III. OBJECTIVE MEASURES

A number of objective measures are examined in the present study for predicting the intelligibility of speech in noisy conditions. Some of the objective measures (e.g., PESQ) have been used successfully for the evaluation of speech quality (e.g., Quackenbush *et al.*, 1988; Rix et al, 2001), while others are more appropriate for intelligibility assessment. A description of these measures along with the proposed modifications to speech-based STI and AI-based measures is given next.

A. PESQ

Among all objective measures considered, the PESQ measure is the most complex to compute and is the one recommended by ITU-T (2000) for speech quality assessment of 3.2 kHz (narrow-band) handset telephony and narrow-band speech codecs (Rix *et al.*, 2001; ITU-T, 2000). The PESQ measure is computed as follows. The original (clean) and degraded signals are first level equalized to a standard listening level and filtered by a filter with response similar to that of a standard telephone handset. The signals are time aligned to correct for time delays, and then processed through an auditory transform to obtain the loudness spectra. The difference in loudness between the original and degraded signals is computed and averaged over time and frequency to produce the prediction of subjective quality rating. The PESQ produces a score between 1.0 and 4.5, with high values indicating better quality. High correlations ($r > 0.92$) with subjective listening tests were reported by Rix *et al.* (2001) using the above PESQ measure for a large number of testing conditions taken from voice-over-internet protocol applications. High correlation ($r \approx 0.9$) was also reported in Hu and Loizou (2008) with the subjective quality judgments of noise-corrupted speech processed via noise-suppression algorithms.

B. LPC-based objective measures

The LPC-based measures assess, for the most part, the spectral envelope difference between the input (clean) signal and the processed (or corrupted) signal. Three different LPC-based objective measures were considered: the log likelihood ratio (LLR), the Itakura–Saito (IS), and the cepstrum (CEP) distance measures. All three measures assess the difference between the spectral envelopes, as computed by the LPC model, of the noise-free and processed signals. The LLR measure is defined as (Quackenbush *et al.*, 1988)

$$d_{\mathrm{LLR}}(\vec{a}_p,\vec{a}_c) = \log\left(\frac{\vec{a}_p \mathbf{R}_c \vec{a}_p^T}{\vec{a}_c \mathbf{R}_c \vec{a}_c^T}\right), \tag{1}$$

where $\vec{a}_c$ is the LPC vector of the clean speech signal, $\vec{a}_p$ is the LPC vector of the processed (enhanced) speech signal, and $\mathbf{R}_c$ is the autocorrelation matrix of the noise-free speech signal. Only the smallest 95% of the frame LLR values were used to compute the average LLR value (Hu and Loizou, 2008). The segmental LLR values were limited in the range of [0, 2] to further reduce the number of outliers (Hu and Loizou, 2008).

The IS measure is defined as (Quackenbush *et al.*, 1988)

$$d_{\mathrm{IS}}(\vec{a}_p,\vec{a}_c) = \frac{\sigma_c^2}{\sigma_p^2}\left(\frac{\vec{a}_p \mathbf{R}_c \vec{a}_p^T}{\vec{a}_c \mathbf{R}_c \vec{a}_c^T}\right) + \log\left(\frac{\sigma_c^2}{\sigma_p^2}\right) - 1, \tag{2}$$

where σ_c^2 and σ_p^2 are the LPC gains of the clean and processed signals, respectively. The IS values were limited in the range of [0, 100] to minimize the number of outliers.

The CEP distance provides an estimate of the log spectral distance between two spectra and is computed as follows (Kitawaki *et al.*, 1988):

$$d_{\mathrm{CEP}}(\vec{c}_c,\vec{c}_p) = \frac{10}{\log 10}\sqrt{2\sum_{k=1}^{p}[c_c(k) - c_p(k)]^2}, \tag{3}$$

where $\vec{c}_c$ and $\vec{c}_p$ are the CEP coefficient vectors of the noise-free and processed signals, respectively. The CEP distance was limited in the range of [0, 10] to minimize the number of outliers (Hu and Loizou, 2008).

C. Time-domain and frequency-weighted SNR measures

The time-domain segmental SNR (SNRseg) measure was computed as per Hansen and Pellom (1998) as follows:

$$\mathrm{SNRseg} = \frac{10}{M}\sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1}(x(n) - \hat{x}(n))^2}, \tag{4}$$

where $x(n)$ is the input (clean) signal, $\hat{x}(n)$ is the processed (enhanced) signal, N is the frame length (chosen to be 30 ms), and M is the number of frames in the signal. Only frames with SNRseg in the range of $[-10, 35]$ dB were considered in the computation of the average (Hansen and Pellom, 1998).

The frequency-weighted segmental SNR (fwSNRseg) was computed using the following equation (Hu and Loizou, 2008):

$$\text{fwSNRseg} = \frac{10}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)\log_{10}\frac{X(j,m)^2}{(X(j,m)-\hat{X}(j,m))^2}}{\sum_{j=1}^{K} W(j,m)}, \quad (5)$$

where $W(j,m)$ is the weight placed on the jth frequency band, K is the number of bands, M is the total number of frames in the signal, $X(j,m)$ is the critical-band magnitude (excitation spectrum) of the clean signal in the jth frequency band at the mth frame, and $\hat{X}(j,m)$ is the corresponding spectral magnitude of the enhanced signal in the same band. The critical-band spectra $X(j,m)$ in Eq. (5) were obtained by multiplying the FFT magnitude spectra by 25 overlapping Gaussian-shaped windows (Loizou, 2007, Chap. 11) spaced in proportion to the ear's critical bands and summing up the power within each band. Similar to the implementation in Hu and Loizou (2008), the excitation spectra were normalized to have an area of 1. The SNR term in the numerator of Eq. (5) was limited within the range of $[-15,15]$ dB. To assess the influence of the dynamic range on performance, we also considered limiting the SNR range to $[-15,20]$, $[-15,25]$, $[-15,30]$, $[-15,35]$, and $[-10,35]$ dB. The latter range $[-10,35]$ dB) was chosen for two reasons. First, to facilitate comparisons with the SNRseg measure [Eq. (4)], which was also limited to the same range. Second, it was chosen to be consistent with several studies (Boothroyd *et al.*, 1994; Studebaker and Sherbecoe, 2002) that showed that the speech dynamic range often exceeds 30 dB.

For the weighting function $W(j,m)$, we considered the AI weights (given in Table I) as well as the critical-band spectrum of the noise-free signal raised to a power, i.e.,

$$W(j,m) = X(j,m)^p, \quad (6)$$

where p is the power exponent, which can be varied for maximum correlation and can be optimized for different speech materials. In our experiments, we varied p from 0.5 to 4. The AI weights were taken from Table B.1 of the ANSI (1997) standard. For the consonant materials, we used the nonsense syllable weights and for the sentence materials we used the short-passage weights given in Table B.1 (ANSI, 1997). The weights were linearly interpolated to reflect the range of band center-frequencies adopted in the present study.

The value of p in Eq. (6) can control the emphasis or weight placed on spectral peaks and/or spectral valleys. Values of $p<1$, for instance, compress the spectrum, while values of $p>1$ expand the spectrum. Compressive values of $p(p<1)$ equalize the spectrum by boosting the low-intensity components (e.g., spectral valleys). Consequently, the effective dynamic range of the spectrum is reduced, and relatively uniform weights are applied to all spectral components. Figure 2 shows as an example the spectrum of a segment taken from the vowel /ɛ/ (as in "head"), along with the same spectrum raised to powers of 0.25 and 1.25. Note that prior to the compression, the F2 amplitude is very weak compared to the F1 amplitude (compare the top two panels). After the compression, the F2 peak gets stronger and closer in amplitude to F1's. Expansion $(p>1)$, on the other hand, has the opposite effect in that it enhances the dominant spectral peak(s), while suppressing further the weak spectral components (see bottom panel in Fig. 2). In this example, the F2 amplitude was further weakened following the spectrum expansion. In brief, the value of p in Eq. (6) controls the steepness of the compression/expansion function, and in practice, it can be optimized for different speech materials.

TABLE I. AI weights (ANSI, 1997) used in the implementation of the fwSNRseg and AI-ST measures for consonants and sentence materials.

Band	Center frequencies (Hz)	Consonants	Sentences
1	50.0000	0.0000	0.0064
2	120.000	0.0000	0.0154
3	190.000	0.0092	0.0240
4	260.000	0.0245	0.0373
5	330.000	0.0354	0.0803
6	400.000	0.0398	0.0978
7	470.000	0.0414	0.0982
8	540.000	0.0427	0.0809
9	617.372	0.0447	0.0690
10	703.378	0.0472	0.0608
11	798.717	0.0473	0.0529
12	904.128	0.0472	0.0473
13	1020.38	0.0476	0.0440
14	1148.30	0.0511	0.0440
15	1288.72	0.0529	0.0470
16	1442.54	0.0551	0.0489
17	1610.70	0.0586	0.0486
18	1794.16	0.0657	0.0491
19	1993.93	0.0711	0.0492
20	2211.08	0.0746	0.0500
21	2446.71	0.0749	0.0538
22	2701.97	0.0717	0.0551
23	2978.04	0.0681	0.0545
24	3276.17	0.0668	0.0508
25	3597.63	0.0653	0.0449

The last conventional measure tested was the weighted spectral slope (WSS) measure (Klatt, 1982). The WSS distance measure computes the weighted difference between the spectral slopes in each frequency band. The spectral slope is obtained as the difference between adjacent spectral magnitudes in decibels. The WSS measure evaluated in this paper is defined as

$$d_{\text{WSS}} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W_{\text{WSS}}(j,m)(S_c(j,m)-S_p(j,m))^2}{\sum_{j=1}^{K} W_{\text{WSS}}(j,m)}, \quad (7)$$

where $W_{\text{WSS}}(j,m)$ are the weights computed as per Klatt (1982), $K=25$, M is the number of data segments, and $S_c(j,m)$ and $S_p(j,m)$ are the spectral slopes for the jth frequency band of the noise-free and processed speech signals, respectively.

Aside from the PESQ measure, all other measures were computed by segmenting the sentences using 30-ms duration

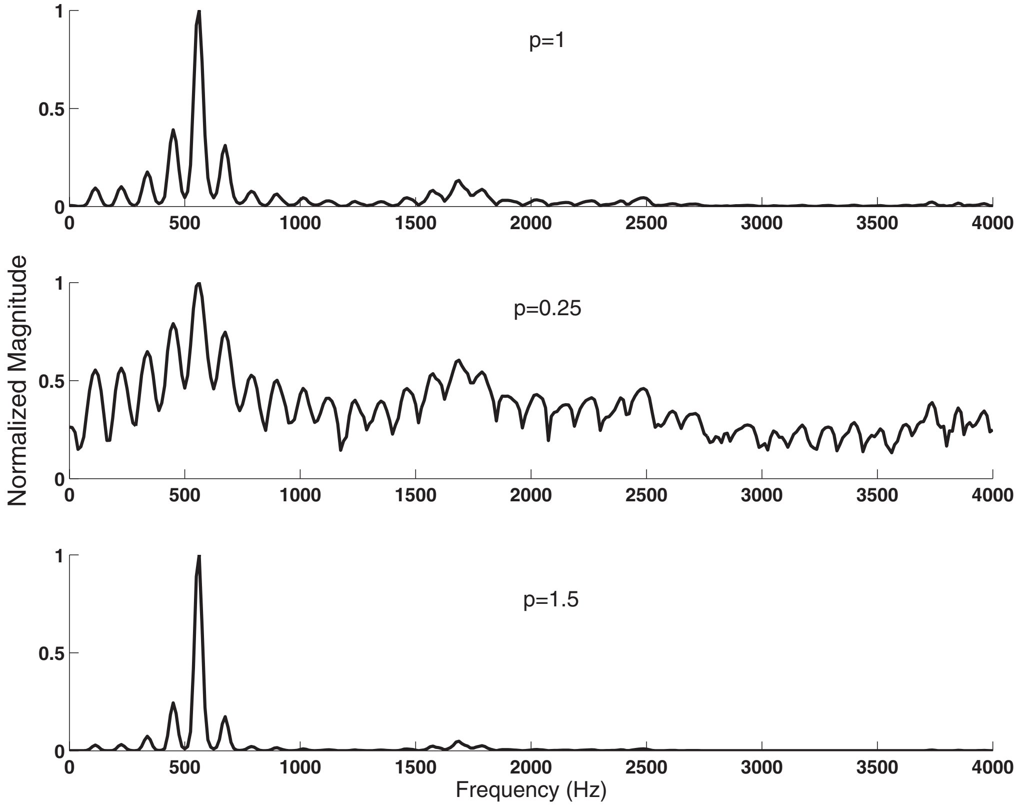


FIG. 2. (Top panel) FFT magnitude spectrum of a segment taken from the vowel /ɛ/ (excised from the word "head" and produced by a male talker). (Middle panel) Same spectrum raised to the power of 0.25. (Bottom panel) Same spectrum raised to the power of 1.5. All spectra are shown in linear units and have been normalized by their maximum for better visual clarity.

Hamming windows with 75% overlap between adjacent frames. This frame duration was chosen to be consistent with that used in our previous study (Hu and Loizou, 2008) which focused on evaluation of objective measures for predicting quality ratings. A tenth-order LPC analysis was used in the computation of the LPC-based objective measures (CEP, IS, and LLR).

D. Normalized covariance metric measures

From the various speech-based STI measures proposed (see review in Goldsworthy and Greenberg, 2004), we chose the normalized covariance metric (NCM) (Hollube and Kollmeier, 1996). This measure is similar to the STI (Steeneken and Houtgast, 1980) in that it computes the STI as a weighted sum of transmission index (TI) values determined from the envelopes of the probe and response signals in each frequency band (Goldsworthy and Greenberg, 2004). Unlike the traditional STI measure, however, which quantifies the change in modulation depth between the probe and response envelopes using the modulation transfer function (MTF), the NCM measure is based on the covariance between the probe (input) and response (output) envelope signals.

The NCM measure is computed as follows. The stimuli were first bandpass filtered into K bands spanning the signal bandwidth. The envelope of each band was computed using the Hilbert transform and then downsampled to 25 Hz, thereby limiting the envelope modulation frequencies to 0–12.5 Hz. Let $x_i(t)$ be the downsampled envelope in the ith band of the clean (probe) signal and let $y_i(t)$ be the downsampled envelope of the processed (response) signal. The normalized covariance in the ith frequency band is computed as

$$r_i = \frac{\Sigma_t(x_i(t)-\mu_i)(y_i(t)-\nu_i)}{\sqrt{\Sigma_t(x_i(t)-\mu_i)^2}\sqrt{\Sigma_t(y_i(t)-\nu_i)^2}}, \qquad (8)$$

where μ_i and ν_i are the mean values of the $x_i(t)$ and $y_i(t)$ envelopes, respectively. Note that the r_i values are limited to $|r_i| \leqslant 1$. A value of r_i close to 1 would suggest that the input [i.e., $x_i(t)$] and processed [i.e., $y_i(t)$] signals are linearly related, while a value of r_i close to 0 would indicate that the input and processed signals are uncorrelated. The SNR in each band is computed as

$$\mathrm{SNR}_i = 10\log_{10}\left(\frac{r_i^2}{1-r_i^2}\right). \qquad (9)$$

and subsequently limited to the range of $[-15,15]$ dB (as done in the computation of the SII measure, ANSI, 1997). The TI in each band is computed by linearly mapping the SNR values between 0 and 1 using the following equation:

$$\mathrm{TI}_i = \frac{\mathrm{SNR}_i + 15}{30}. \qquad (10)$$

Finally, the transmission indices are averaged across all frequency bands to produce the NCM index:

TABLE II. AI weights (ANSI, 1997) used in the implementation of the NCM measure for consonants and sentence materials.

Band	Center freq. (kHz)	Consonants	Sentences
1	0.3249	0.0346	0.0772
2	0.3775	0.0392	0.0955
3	0.4356	0.0406	0.1016
4	0.5000	0.0420	0.0908
5	0.5713	0.0433	0.0734
6	0.6502	0.0457	0.0659
7	0.7376	0.0472	0.0580
8	0.8344	0.0473	0.0500
9	0.9416	0.0471	0.0460
10	1.0602	0.0487	0.0440
11	1.1915	0.0519	0.0445
12	1.3370	0.0534	0.0482
13	1.4980	0.0562	0.0488
14	1.6763	0.0612	0.0488
15	1.8737	0.0684	0.0493
16	2.0922	0.0732	0.0491
17	2.3342	0.0748	0.0520
18	2.6022	0.0733	0.0549
19	2.8989	0.0685	0.0555
20	3.2274	0.0670	0.0514

$$\mathrm{NCM} = \frac{\Sigma_{i=1}^{K} W_i \times \mathrm{TI}_i}{\Sigma_{i=1}^{K} W_i}, \tag{11}$$

where W_i are the weights applied to each of the K bands. The denominator term is included for normalization purposes. The weights W_i are often called BIF in the computation of the SII measure (ANSI, 1997). Fixed weights (given in Table II), such as those used in AI studies, are often used in the computation of the STI measure (Steeneken and Houtgast, 1980). In our study, we consider making those weights signal and frequency (i.e., band) dependent. More precisely, we considered the following two weighting functions:

$$W_i^{(1)} = \left(\sum_t x_i^2(t) \right)^p, \tag{12}$$

$$W_i^{(2)} = \left(\sum_t (\max[x_i(t) - d_i(t), 0])^2 \right)^p, \tag{13}$$

where $d_i(t)$ denotes the (downsampled) scaled masker signal in the time domain. The power exponent p was varied from 0.12 to 1.5. The motivation behind the use of Eq. (12) is to place weight to each TI value in proportion to the signal energy in each band. The motivation behind the use of Eq. (13) is to place weight to each TI value in proportion to the excess masked signal.

To assess the influence of the SNR range used in the computation of the STI measure, we also considered limiting the SNR to the range of $[-15,20]$, $[-15,25]$, $[-15,30]$, $[-15,35]$, and $[-10,35]$ dB. To accommodate for the new range in SNR values, the TI values in Eq. (10) were modified accordingly. So, for instance, to accommodate the $[-10,35]$ dB range, the TI values in Eq. (10) were computed as follows:

$$\mathrm{TI}_i = \frac{\mathrm{SNR}_i + 10}{45}. \tag{14}$$

The above equation ensures that the SNR is linear mapped to values between 0 and 1.

The STI measure is typically evaluated for modulation frequencies spanning 0.63–12.5 Hz. To assess the influence of including higher modulation frequencies ($>$12.5 Hz), we also varied the modulation frequency range to 0–20 and 0–31 Hz. This was motivated by the study of Van Wijngaarden and Houtgast (2004) that showed that extending the modulation bandwidth to 31.5 Hz improved the correlation of the STI index for conversational-style speech.

The NCM computation in Eq. (11) takes into account a total of K bands spanning the signal bandwidth, which was 4 kHz in our study. To assess the contribution of low-frequency envelope information, spanning the range of 100–1000 Hz, we considered a variant of the above NCM measure in which we included only the low-frequency ($<$1000 Hz) bands in the computation. We refer to this measure as the low-frequency NCM measure and denote it as $\mathrm{NCM}_{\mathrm{LF}}$:

$$\mathrm{NCM}_{\mathrm{LF}} = \frac{\Sigma_{i=1}^{8} W_i \times \mathrm{TI}_i}{\Sigma_{i=1}^{8} W_i}. \tag{15}$$

Note that only the first eight low-frequency envelopes, spanning the frequency range of 100–1000 Hz, are used in the computation of the $\mathrm{NCM}_{\mathrm{LF}}$ measure. We considered using uniform weights for all frequency envelopes (i.e., $W_i=1$ for all bands) as well as the weights given in Eq. (12). The $\mathrm{NCM}_{\mathrm{LF}}$ measure can be considered to be a simplified version of the NCM measure, much like the rapid STI (RASTI) measure is a simplified version of the STI measure. The RASTI measure is calculated using only the 500- and 2000-Hz octave bands (IEC 60268, 2003). In terms of prediction accuracy, the RASTI measure was found to produce comparable results to that obtained by the STI measure (Mapp, 2002; Larm and Hongisto, 2006).

E. AI-based measures

A simplified version of the SII measure is considered in this study that operates on a frame-by-frame basis. The proposed measure differs from the traditional SII measure (ANSI, 1997) in many ways: (a) it does not require as input the listener's threshold of hearing, (b) does not account for spread of upward masking, and (c) does not require as input the long-term average spectrum (sound-pressure) levels of the speech and masker signals. The proposed AI-ST measure divides the signal into short (30 ms) data segments, computes the AI value for each segment, and averages the segmental AI values over all frames. More precisely, it is computed as follows:

$$\text{AI-ST} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\Sigma_{j=1}^{K} W(j,m) T(j,m)}{\Sigma_{j=1}^{K} W(j,m)}, \tag{16}$$

where M is the total number of data segments in the signal, $W(j,m)$ is the weight (i.e., band importance function, ANSI, 1997) placed on the jth frequency band, and

$$T(j,m) = \frac{\mathrm{SNR}(j,m) + 15}{30}, \quad (17)$$

$$\mathrm{SNR}(j,m) = 10\log_{10}\frac{\hat{X}(j,m)^2}{D(j,m)^2}, \quad (18)$$

where $D(j,m)$ denotes the critical-band spectrum of the scaled masker signal (obtained before mixing) and $\hat{X}(j,m)$ denotes the enhanced signal's critical-band spectral magnitude in the jth band. Unlike the normalization used in the computation of the fwSNRseg measure [Eq. (5)], the excitation spectra were not normalized to have an area of unity. The SNR term in Eq. (18) was limited within the range of $[-15,15]$ dB and mapped linearly in each band to values between 0 and 1 using Eq. (17). For comparative purposes, we also considered limiting the SNR in Eq. (18) to $[-15,20]$, $[-15,25]$, $[-15,30]$, $[-15,35]$, and $[-10,35]$ dB.

Aside from using the AI weights for $W(j,m)$ (see Table I), the following four band-importance weighting functions were also considered for $W(j,m)$ in Eq. (16):

$$W_1(j,m) = \begin{cases} 1 & \text{if } X(j,m) > D(j,m) \\ 0 & \text{else,} \end{cases} \quad (19)$$

$$W_2(j,m) = \begin{cases} (X(j,m) - D(j,m))^p & \text{if } X(j,m) > D(j,m) \\ 0 & \text{else,} \end{cases} \quad (20)$$

$$W_3(j,m) = \begin{cases} X(j,m)^p & \text{if } X(j,m) > D(j,m), \\ 0 & \text{else} \end{cases} \quad (21)$$

$$W_4(j,m) = X(j,m)^p. \quad (22)$$

The motivation behind the use of the above BIFs [Eqs. (19)–(21)] was to include in the computation of the AI-ST measure only bands with positive SNR, i.e., only bands in which the target is stronger than the masker. The rather simplistic assumption made here is that bands with negative SNR contribute little, if anything, to intelligibility. As such, those bands should not be included in the computation of the AI-ST measure. The power exponent p in Eqs. (20)–(22) was varied from 0.5 to 4. As mentioned earlier, the value of p controls the emphasis or weight placed on spectral peaks and/or spectral valleys. Use of $p > 1$, for instance, places more emphasis on the dominant spectral peaks (see example in Fig. 2).

Unlike the BIFs used in the traditional AI measure (ANSI, 1997) and in the extended (short-term) versions of the AI measure (Kates, 1987; Kates and Arehart, 2005; Rhebergen and Versfeld, 2005), the BIFs proposed in Eqs. (19)–(22) are signal and segment dependent. This was done to account for the fact that the AI-ST values are computed at a (short-duration) segmental level rather than on a global (long-term average spectrum) level. The speech-and masker-spectra vary markedly over time, and this variation is captured to some degree with the use of signal-dependent band-importance (weighting) functions.

F. Coherence-based measures

The magnitude-squared coherence (MSC) function is the normalized cross-spectral density of two signals and has been used to assess distortion in hearing aids (Kates, 1992). It is computed by dividing the input (clean) and output (processed) signals in a number (M) of overlapping windowed segments, computing the cross power spectrum for each segment using the FFT, and then averaging across all segments. For M data segments (frames), the MSC at frequency bin ω is given by

$$\mathrm{MSC}(\omega) = \frac{|\sum_{m=1}^{M} X_m(\omega) Y_m^*(\omega)|^2}{\sum_{m=1}^{M}|X_m(\omega)|^2 \sum_{m=1}^{M}|Y_m(\omega)|^2}, \quad (23)$$

where the asterisk denotes the complex conjugate and $X_m(\omega)$ and $Y_m(\omega)$ denote the FFT spectra of the $x(t)$ and $y(t)$ signals, respectively, computed in the mth data segment. In our case, $x(t)$ corresponds to the clean signal and $y(t)$ corresponds to the enhanced signal. The MSC measure takes values in the range of 0–1. The averaged, across all frequency bins, MSC was used in our study as the objective measure. The MSC was computed by segmenting the sentences using 30-ms duration Hamming windows with 75% overlap between adjacent frames. The use of a large frame overlap ($>50\%$) was found by Carter *et al.* (1973) to reduce bias and variance in the estimate of the MSC.

It should be noted that the above MSC function can be expressed as a weighted MTF (see Appendix), which is used in the implementation of the STI measure (Houtgast and Steeneken, 1985). The main difference between the MTF (Houtgast and Steeneken, 1985) used in the computation of the STI measure and the MSC function is that the latter function is evaluated for all frequencies spanning the signal bandwidth, while the MTF is evaluated only for low modulation frequencies (0.5–16 Hz).

Extensions of the MSC measure were proposed by Kates and Arehart (2005) for assessing the effects of hearing-aid distortions (e.g., peak clipping) on speech intelligibility by normal-hearing and hearing-impaired subjects. More precisely, the new measure, called coherence SII (CSII), was proposed that used the SII index as the base measure and replaced the SNR term with the signal-to-distortion ratio term, which was computed using the coherence between the input and output signals. That is, the $\mathrm{SNR}(j,m)$ term in Eq. (18) was replaced with the following expression:

$$\mathrm{SNR}_{\mathrm{CSII}}(j,m) = 10\log_{10}\frac{\sum_{k=1}^{N} G_j(\omega_k) \times \mathrm{MSC}(\omega_k)|Y_m(\omega_k)|^2}{\sum_{k=1}^{N} G_j(\omega_k)[1-\mathrm{MSC}(\omega_k)]|Y_m(\omega_k)|^2}, \quad (24)$$

where $G_j(\omega)$ denotes the ro-ex filter (Moore and Glasberg, 1993) centered around the jth critical band, $\mathrm{MSC}(\omega)$ is given by Eq. (23), $Y(\omega_k)$ is the FFT spectrum of the enhanced signal, and N is the FFT size. The above SNR term is limited to $[-15,15]$ dB and mapped linearly between 0 and 1 using Eq. (17) to produce a new $T_{\mathrm{CSII}}(j,m)$ term. Finally, the latter term is substituted in Eq. (16) to compute the CSII value as follows:

$$\mathrm{CSII} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)T_{\mathrm{CSII}}(j,m)}{\sum_{j=1}^{K} W(j,m)}. \qquad (25)$$

The above CSII measure is computed using all M speech segments of the utterance. Kates and Arehart (2005) found that a three-level version of the CSII measure yielded higher correlation with speech intelligibility than the above CSII measure. The three measures were computed by first dividing the M speech segments into three level regions and computing separately the CSII measure for each region. The high-level region consisted of segments at or above the overall root-mean-square (rms) level of the whole utterance. The mid-level region consisted of segments ranging from the overall rms level to 10 dB below, and the low-level region consisted of segments ranging from rms–10 dB to rms–30 dB. The three-level CSII measures obtained for the low-, mid-, and high-level segments were denoted as $\mathrm{CSII}_{\mathrm{low}}$, $\mathrm{CSII}_{\mathrm{mid}}$, and $\mathrm{CSII}_{\mathrm{high}}$, respectively. A linear combination of the three CSII values followed by a logistic function transformation was subsequently used to model the intelligibility scores. The resulting intelligibility measure, termed I3 (Kates and Arehart, 2005), will be evaluated and compared against other measures in the present study. The I3 measure was later extended by Arehart *et al.* (2007) to model judgments of quality ratings of noise and hearing-aid type of distortions. A new measure, termed Q3, was developed based on a different linear combination of the three-level CSII measures (Arehart *et al.*, 2007).

The critical-band spacing was used in the implementation of the above CSII measures (Kates and Arehart, 2005). A total of 16 critical bands spanning the bandwidth of 100–3700 Hz were used in our implementation. The BIF given in Table B.1 (ANSI, 1997) were used in Eq. (25) for $W(j,m)$. In addition, the four band-importance weighting functions proposed in Eqs. (19)–(22) were tested.

IV. RESULTS

Two figures of merit were used to assess the performance of the above objective measures in terms of predicting speech intelligibility. The first figure of merit was Pearson's correlation coefficient, r, and the second figure of merit was an estimate of the standard deviation of the error computed as $\sigma_e = \sigma_d\sqrt{1-r^2}$, where σ_d is the standard deviation of the speech recognition scores in a given condition, and σ_e is the computed standard deviation of the error. A smaller value of σ_e indicates that the objective measure is better at predicting speech intelligibility.

The average intelligibility scores obtained by normal-hearing listeners in the 72 different noisy conditions (see Sec. II) were subjected to correlation analysis with the corresponding mean values obtained with the objective measures. As mentioned earlier, these conditions involved noise-suppressed speech (consonants and sentences) originally corrupted by four different maskers (car, babble, train, and street interferences) at two different SNR levels. The computed correlation coefficients (and prediction error) are tabulated separately for the consonants and sentence materials and are given in Tables III and IV, respectively.

TABLE III. Correlation coefficients, r, and standard deviations of the error, σ_e, between consonant recognition scores and the various objective measures examined. The BIFs used in some measures are indicated in the second column. In the implementation of the fwSNRseg, NCM, CSII, and AI-ST measures the SNR was restricted in the range of $[-15,15]$ dB.

Objective measure	Band-importance function	r	σ_e
PESQ		0.77	0.08
LLR		−0.51	0.10
SNRseg		0.40	0.12
WSS		−0.33	0.11
Itakura–Saito (IS)		−0.35	0.12
Cepstrum (CEP)		−0.48	0.11
Coherence (MSC)		0.76	0.08
CSII	ANSI (1997)	0.76	0.08
$\mathrm{CSII}_{\mathrm{high}}$	ANSI (1997)	0.80	0.07
$\mathrm{CSII}_{\mathrm{mid}}$	ANSI (1997)	0.80	0.07
$\mathrm{CSII}_{\mathrm{low}}$	ANSI (1997)	0.36	0.12
I3		0.80	0.07
Q3		0.79	0.07
mI3		0.82	0.07
CSII	W_4, $p=0.5$, Eq. (22)	0.77	0.08
$\mathrm{CSII}_{\mathrm{high}}$	W_4, $p=0.5$, Eq. (22)	0.80	0.07
$\mathrm{CSII}_{\mathrm{mid}}$	W_4, $p=0.5$, Eq. (22)	0.78	0.08
$\mathrm{CSII}_{\mathrm{low}}$	W_4, $p=4$, Eq. (22)	0.68	0.09
fwSNRseg	ANSI (Table I)	0.59	0.10
fwSNRseg	Eq. (6), $p=4$	0.68	0.09
$\mathrm{NCM}_{\mathrm{LF}}$	$W_i=1$	0.65	0.09
$\mathrm{NCM}_{\mathrm{LF}}$	$W_i^{(1)}$, $p=1$, Eq. (12)	0.72	0.09
NCM	ANSI (Table II)	0.66	0.09
NCM	$W_i^{(1)}$, $p=0.5$, Eq. (12)	0.77	0.08
NCM	$W_i^{(2)}$, $p=1$, Eq. (13)	0.72	0.09
AI-ST	ANSI (Table I)	0.39	0.11
AI-ST	W_1, Eq. (19)	0.56	0.10
AI-ST	W_2, $p=4$, Eq. (20)	0.68	0.09
AI-ST	W_3, $p=4$, Eq. (21)	0.67	0.09
AI-ST	W_4, $p=4$, Eq. (22)	0.52	0.11

A. Subjective quality measures

Of the seven measures designed for subjective quality assessment, the PESQ and fwSNRseg measures performed the best. When applied to the sentence materials, the fwSNRseg measure, based on the weighting function given in Eq. (6), performed better than the PESQ measure and yielded a correlation of $r=0.81$, compared to $r=0.79$ obtained with the PESQ measure. When applied to the consonant materials, the PESQ measure performed better than the fwSNRseg measure. The LLR measure, which was found in Hu and Loizou (2008) to yield a correlation coefficient that was nearly as good as that of the PESQ measure, performed comparatively worse than the PESQ measure. The MSC, which has been used to assess hearing-aid distortion, performed modestly well ($r=0.71-0.77$) for both sentence and consonant materials. We believe that the modest performance of the MSC measure can be attributed to the fact that the MSC function can be expressed as a weighted MTF (see Appendix), which is used in the implementation of the STI measure. Higher correlation ($r=0.79-0.88$) was obtained with the coherence-based Q3 measure, which was used by Arehart *et al.* (2007) for modeling subjective quality judgments of

TABLE IV. Correlation coefficients, r, and standard deviations of the error, σ_e, between sentence recognition scores and the various objective measures examined. The BIFs used in some measures are indicated in the second column. In the implementation of the fwSNRseg, NCM, CSII, and AI-ST measures the SNR was restricted in the range of [−15,15] dB.

Objective measure	Band-importance function	r	σ_e
PESQ		0.79	0.11
LLR		−0.56	0.15
SNRseg		−0.46	0.15
WSS		−0.27	0.17
Itakura–Saito (IS)		−0.22	0.17
Cepstrum (CEP)		−0.49	0.15
Coherence (MSC)		0.71	0.12
CSII		0.82	0.10
$\text{CSII}_{\text{high}}$		0.85	0.09
CSII_{mid}		0.91	0.07
CSII_{low}		0.86	0.09
I3		0.92	0.07
Q3		0.88	0.08
mI3		0.92	0.07
CSII	W_4, p=4, Eq. (22)	0.86	0.09
$\text{CSII}_{\text{high}}$	W_4, p=2, Eq. (22)	0.88	0.08
CSII_{mid}	W_4, p=1, Eq. (22)	0.94	0.06
CSII_{low}	W_4, p=0.5, Eq. (22)	0.86	0.09
fwSNRseg	ANSI (Table I)	0.78	0.11
fwSNRseg	Eq. (6), p=1	0.81	0.10
NCM_{LF}	W_i=1	0.81	0.10
NCM_{LF}	$W_i^{(1)}$, p=2, Eq. (12)	0.87	0.09
NCM	ANSI (Table II)	0.82	0.10
NCM	$W_i^{(1)}$, p=1.5, Eq. (12)	0.89	0.07
NCM	$W_i^{(2)}$, p=1.5, Eq. (13)	0.89	0.08
AI-ST	ANSI (Table I)	0.33	0.16
AI-ST	W_1, Eq. (19)	0.66	0.13
AI-ST	W_2, p=3, Eq. (20)	0.80	0.11
AI-ST	W_3, p=3, Eq. (21)	0.80	0.11
AI-ST	W_4, p=4, Eq. (22)	0.62	0.14

hearing-aid distortion. In summary, of all the measures tested previously (Hu and Loizou, 2008) for subjective quality predictions, the fwSNRseg and PESQ measures seem to predict modestly well both speech quality and speech intelligibility.

B. Intelligibility measures

Of all the intelligibility measures considered, the coherence-based (CSII) and NCM measures performed the best. The highest correlations were obtained with the CSII measures for both consonants and sentence materials. The I3 measure (Kates and Arehart, 2005), in particular, produced the highest correlation for consonants (r=0.80) and sentence (r=0.92) materials. Figure 3 shows the scatter plot of the predicted I3 scores against the listeners' recognition scores for consonants and sentences. Figures 4 and 5 show the individual scatter plots broken down by noise type for consonant and sentence recognition, respectively. As can be seen, a high correlation was maintained for all noise types, including modulated (e.g., train) and non-modulated (e.g., car) maskers. The correlations with consonant recognition scores ranged from r=0.82 with street noise to r=0.85 with car noise. The correlations with sentence recognition scores ranged from r=0.88 with train noise to r=0.98 with babble.

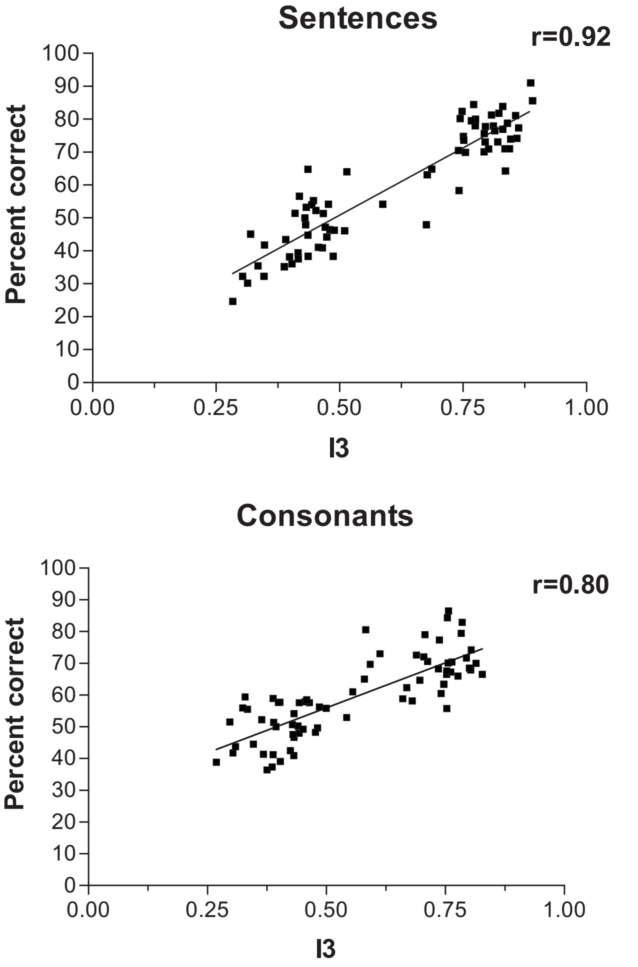


FIG. 3. Scatter plot of sentence recognition scores (top panel) and consonant recognition scores (bottom panel) against the predicted I3 values.

Among the three-level CSII measures, the mid-level CSII (CSII_{mid}) measure yielded the highest correlation for both consonant and sentence materials, consistent with the outcome reported by Kates and Arehart (2005). The CSII_{mid} measure captures information about envelope transients and spectral transitions, critical for the transmission of information regarding place of articulation. Similar to the approach taken in Kates and Arehart (2005), a multiple-regression analysis was run on the three CSII measures, yielding the following predictive models for consonant and sentence intelligibility. For consonants, the modified I3 measure, indicated as mI3, is given by

$$m\text{I3} = 0.026 - 1.033 \times \text{CSII}_{\text{low}} + 0.822 \times \text{CSII}_{\text{mid}} + 0.506 \times \text{CSII}_{\text{high}}, \tag{26}$$

and for sentences, it is given by

$$m\text{I3} = -0.029 - 0.055 \times \text{CSII}_{\text{low}} + 2.206 \times \text{CSII}_{\text{mid}} - 0.349 \times \text{CSII}_{\text{high}}. \tag{27}$$

Subsequent logistic transformations of the mI3 measure did not improve the correlations. The correlations of the above mI3 measures with consonant and sentence recognition

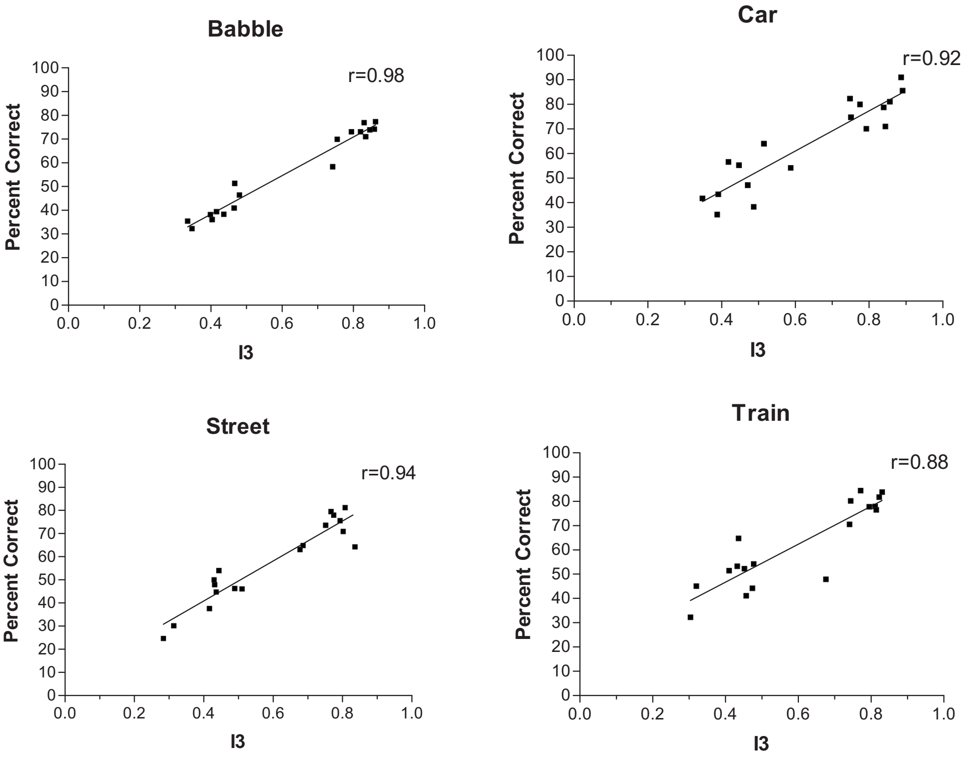


FIG. 4. Individual scatter plots of predicted I3 values against sentence recognition scores for the four types of maskers used.

scores are given in Tables III and IV, respectively. The mI3 measure, via the use of Eq. (26), improved the I3 correlation from 0.80 to 0.82, making it the highest correlation attained for consonants. For sentences, the improvement in performance, over that attained by the I3 measure, was marginal and not evident in Table IV due to the rounding of the correlation values to two decimal places. Further improvements in correlation were obtained with the three-level CSII measures for the sentence materials after applying the proposed signal- and phonetic-segment dependent band-importance functions given in Eq. (22). The correlation of the modified CSII$_{\text{mid}}$ measure improved from $r=0.92$ (7% prediction error) with ANSI (1997) weights to $r=0.94$ (6% prediction error) with the proposed BIF given in Eq. (22). The resulting correlation was higher than that attained with the I3 measure proposed by Kates and Arehart (2005), and it was the highest correlation obtained in the present study.

The next highest correlations were obtained with the modified NCM measure that used the BIF in Eq. (12). The resulting correlation coefficient for sentences was $r=0.89$

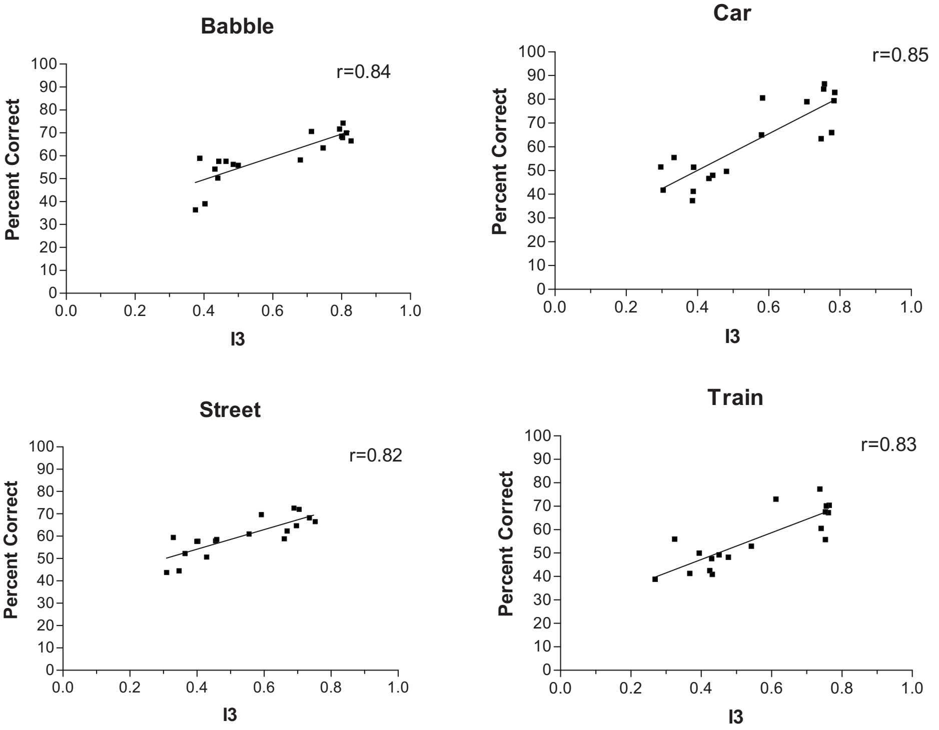


FIG. 5. Individual scatter plots of predicted I3 values against consonant recognition scores for the four types of maskers used.

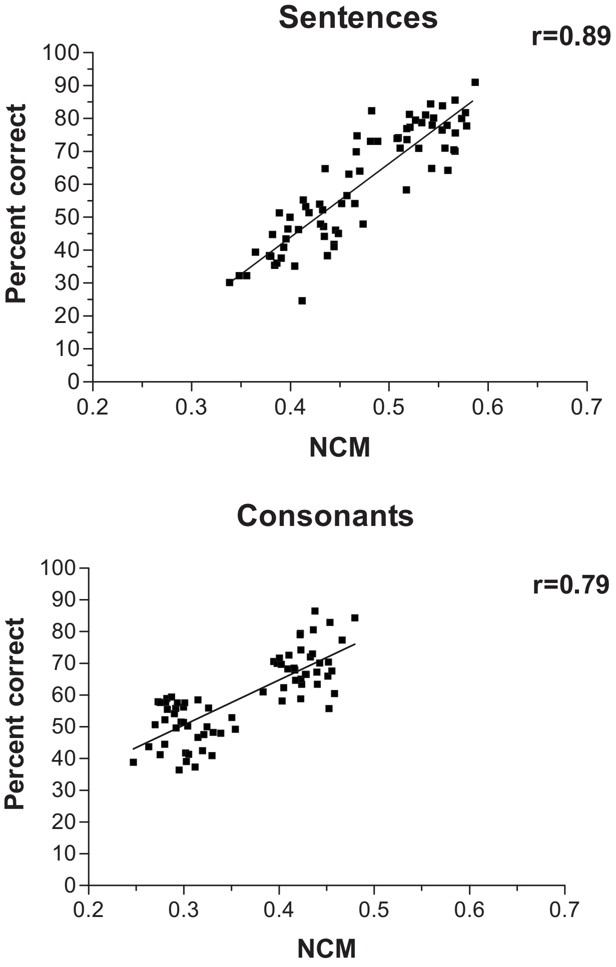


FIG. 6. Scatter plot of sentence recognition scores (top panel) and consonant recognition scores (bottom panel) against the predicted NCM values. In the implementation of the NCM metric, the SNR range was restricted to $[-10,35]$ dB and the BIF was set to that given in Eq. (12) with $p=1.5$ for the sentence materials and $p=0.25$ for the consonant materials.

(7% prediction error) and for consonants it was $r=0.79$ (8% error) when the $[-10,35]$ dB SNR range was used. Figure 6 shows the scatter plot of the predicted NCM scores against the listeners' speech recognition scores. Figures 7 and 8 show the individual scatter plots broken down by noise type for consonant and sentence recognition, respectively. A high correlation was maintained for all noise types, including modulated (e.g., train) and non-modulated (e.g., car) maskers. The correlations obtained with consonant recognition scores ranged from $r=0.75$ with babble to $r=0.89$ with train noise. The correlations obtained with sentence recognition scores ranged from $r=0.85$ with car noise to $r=0.94$ with babble.

As shown in Tables III and IV, performance was clearly influenced by the choice of the band-importance function. In all cases, the lowest correlation was obtained when the AI weights, taken from the ANSI (1997) standard, were used. This clearly demonstrates that the BIFs are material dependent, something that is already accounted for in the ANSI (1997) standard. Different sets of weights are provided for different speech materials (see Table B.1, ANSI, 1997). Complex procedures followed by lengthy experiments are needed to obtain the BIFs tabulated in the ANSI (1997) standard. In contrast, the proposed weighting functions, given in Eqs. (19)–(22), suggest an alternative and easier way for deriving the BIFs.

In the implementation of the NCM measure, we fixed the number of bands to 20, the speech dynamic range to $[-15,15]$ dB, and the range of modulation frequencies to $0-12.5$ Hz. Additional experiments were run to assess the influence of the number of bands, range of modulation frequencies, and speech dynamic range on the prediction of speech intelligibility in noise. Note that the conventional STI measure uses seven $1/3$-octave bands (Houtgast and

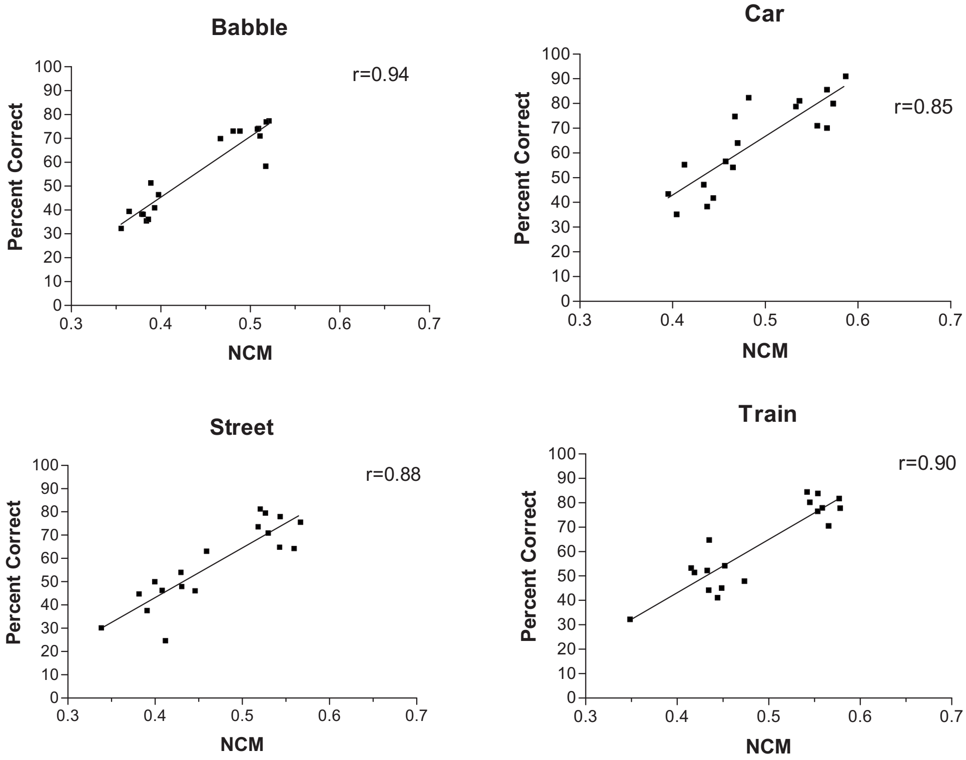


FIG. 7. Individual scatter plots of predicted NCM values against sentence recognition scores for the four types of maskers used.

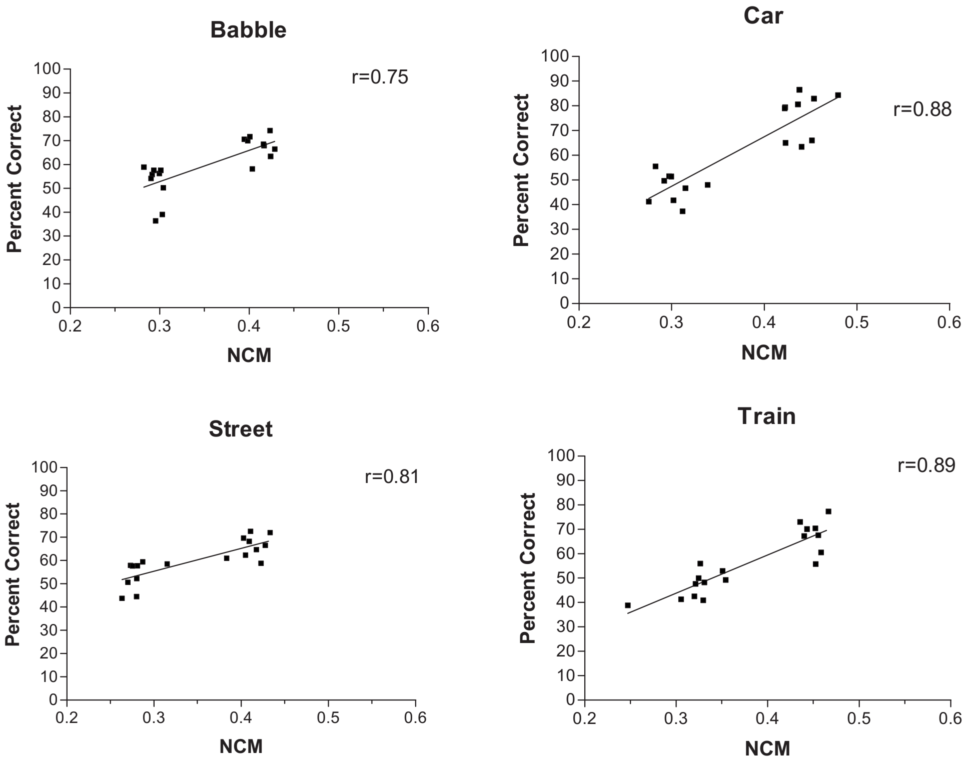


FIG. 8. Individual scatter plots of predicted NCM values against consonant recognition scores for the four types of maskers used.

Steeneken, 1985). To assess the influence of the number of bands on the computation of the NCM measure, we varied the number of bands from 7 to 20. The band center frequencies were logarithmically spaced in the 300–3400 Hz bandwidth. The weighting function given in Eq. (12) with $p=1.5$ was used in all conditions. The resulting correlation coefficients are given in Table V. As can be seen, there is a small, but non-significant, improvement in the correlation as the number of bands increases. Hence, the number of bands used in the computation of the NCM measure does not influence significantly its prediction power.

The implementation of the STI measure typically uses a set of 14 modulation frequencies ranging from 0.63 to 12.5 Hz (Houtgast and Steeneken, 1985). To further assess whether including higher (>12.5 Hz) modulation frequencies would improve the correlation of the NCM measure, we tested two additional implementations that included modulation frequencies up to 20 Hz and up to 31 Hz. The results obtained for different SNR ranges and different ranges of modulation frequencies are tabulated in Table VI. As can be seen there is no improvement for sentences, but a small improvement for consonants. The small improvement obtained with consonants might reflect a difference in the speaking style between the production of consonants vs. sentences (van Wijngaarden and Houtgast, 2004) used in the present study. The sentences used in the present study (taken from Loizou, 2007) were produced with a clear, rather than conversational, speaking style.

TABLE V. Correlation coefficients, r, and standard deviations of the error, σ_e, between sentence recognition scores and the NCM measure as a function of the number of bands used.

No. of bands	r	σ_e
7	0.88	0.08
12	0.88	0.08
16	0.89	0.08
20	0.89	0.08

The correlations obtained with the NCM measure after varying the SNR dynamic range from $[-15,15]$ to $[-15,35]$ dB are shown in Table VII. Performance improved on the consonant recognition task. The correlation coefficient, for instance, obtained with the NCM measure improved from 0.77 to 0.79 when the speech dynamic range increased from 30 to 45 dB. No improvement was noted for the sentence recognition task, at least for the indicated band-importance function. Table VIII shows in more detail the correlations obtained with other band-importance functions and with the SNR dynamic range set to $[-10,35]$ dB. Overall, correlations improved for both consonants and sentences when a wider dynamic range was used.

The performance obtained with the AI-ST measure was quite poor ($r=0.39$ for consonants and $r=0.33$ for sentences) when the AI AI weights were used, but improved considerably when the proposed BIFs were used ($r=0.68$ for consonants and $r=0.80$ for sentences). Compared to the SII implementation (ANSI, 1997) which incorporates upward-spread of masking effects, the AI-ST implementation is rather simplistic. In addition, the averaging of the individual frame AI-ST values in Eq. (16) implicitly assumes that all short (phonetic) segments should be weighted uniformly, i.e., that equal emphasis should be placed on consonant segments, steady-state vowels, and/or vowel-consonant transitions. Furthermore, it is assumed that the same weighting function should be applied to vowels and consonants. Further work is thus needed to develop weighting functions specific to consonants and vowels.

TABLE VI. Correlation coefficients obtained with the NCM measure for different modulation bandwidths and different SNR dynamic ranges. For the sentence materials, the W_1 band-importance function was used with $p=1.5$ and for the consonant materials the W_1 function was used with $p=0.25$.

Material	Modulation bandwidth (Hz)	SNR dynamic range (dB)			
		[−15,20]	[−15,25]	[−15,30]	[−15,35]
Sentences	20.0	0.89	0.89	0.89	0.89
	31.5	0.88	0.88	0.88	0.88
Consonants	20.0	0.74	0.74	0.74	0.74
	31.5	0.77	0.77	0.77	0.77

In the computation of the SII index, the time interval over which the noise and signal are integrated is 125 ms (ANSI, 1997). Within this integration time, the distribution of the speech rms values is approximately linear within a 30 dB dynamic range (Dunn and White, 1940), which is the range adopted for the computation of the SII and STI measures. Several studies have argued, however, that this estimate of speech dynamic range is conservative (e.g., Boothroyd *et al.*, 1994; Studebaker and Sherbecoe, 2002). Studebaker and Sherbecoe (2002), for instance, reported that the dynamic range of BIFs (derived for monosyllabic words) ranged from 36 to 44 dB, with an average value of about 40 dB. Hence, we considered varying the speech dynamic range for both the AI-based and fwSNRseg measures. The resulting correlation coefficients obtained with the wider dynamic range are given in Table VII. As can be seen, the larger dynamic range seemed to influence the performance of the AI-ST measure, but not the fwSNRseg and NCM measures.

Unlike the SII standard (ANSI, 1997) which uses a 125-ms integration window, a 30-ms integration window was used in our present study for the implementation of the AI-ST measure. To assess the influence of window duration, we varied the window duration from 30 to 125 ms. The resulting correlation coefficients are tabulated in Tables IX and X for consonants and sentences, respectively. As can be seen from these tables, performance was influenced by both the weighting function and window duration used. Small improvements were obtained in the prediction of consonant recognition when the window duration increased (Table IX), and considerably larger improvements were obtained in the prediction of sentence recognition (Table X).

V. DISCUSSION

The PESQ measure, which was originally designed to predict quality of speech transmitted over IP networks (ITU-T, 2000), performed modestly well ($r=0.77-0.79$) on predicting intelligibility of consonants and sentences in noise. This was surprising at first, given that this measure assesses overall loudness differences between the input (clean) and processed speech signals, and as such it is more appropriate for predicting subjective quality ratings (Bladon and Lindblom, 1981) than intelligibility. The PESQ measure has been shown in Hu and Loizou (2007) to correlate well ($r=0.81$) with subjective ratings of speech distortion introduced by noise-suppression algorithms. Hence, on this regard it is reasonable to expect that a measure that assesses accurately speech distortion (and overall quality) should also be suitable for assessing speech intelligibility. This is based on the premise (and expectation) that the distortion often introduced by noise-suppression algorithms (e.g., spectral attenuation near formant regions) and imparted on the speech signal, should degrade speech intelligibility. Indeed, the intelligibility study by Hu and Loizou (2007) showed that some noise-suppression algorithms may degrade speech intelligibility in noisy conditions.

Among all objective measures examined in the present study, the modified CSII and NCM measures incorporating

TABLE VII. Correlation coefficients obtained for different SNR dynamic ranges. The band-importance function (BIF) used is given in the third column.

Material	Objective measure	BIF	SNR dynamic range (dB)					
			[−15,15]	[−15,20]	[−15,25]	[−15,30]	[−15,35]	[−15,35]
Sentences	fwSNRseg	$p=2$, Eq. (6)	0.81	0.79	0.78	0.77	0.77	0.80
	NCM	W_1, $p=1.5$, Eq. (12)	0.89	0.89	0.89	0.89	0.89	0.89
	AI-ST	W_2, $p=3$, Eq. (20)	0.80	0.81	0.82	0.83	0.83	0.83
Consonants	fwSNRseg	$p=2.5$, Eq. (6)	0.68	0.65	0.65	0.64	0.64	0.64
	NCM	W_1	0.77	0.78	0.78	0.78	0.78	0.79
	AI-ST	W_3, $p=4$, Eq. (21)	0.67	0.68	0.69	0.69	0.69	0.69

TABLE VIII. Correlation coefficients, r, and standard deviations of the error, σ_e, between sentence/consonant recognition scores and the various objective measures examined. The BIF are given in the third column. The SNR was restricted in the range of $[-10,35]$ dB.

Material	Objective measure	Band-importance function	r	σ_e
Consonants	fwSNRseg	ANSI (Table I)	0.60	0.10
	fwSNRseg	Eq. (6), $p=2.5$	0.64	0.09
	NCM_{LF}	$W_i=1$	0.69	0.09
	NCM_{LF}	$W_i^{(1)}$, $p=2$, Eq. (12)	0.74	0.08
	NCM	ANSI (Table II)	0.73	0.08
	NCM	$W_i^{(1)}$, $p=0.25$, Eq. (12)	0.79	0.08
	NCM	$W_i^{(2)}$, $p=0.25$, Eq. (13)	0.76	0.08
	AI-ST	ANSI (Table I)	0.42	0.11
	AI-ST	W_1, Eq. (19)	0.57	0.10
	AI-ST	W_2, $p=4$, Eq. (20)	0.69	0.09
	AI-ST	W_3, $p=4$, Eq. (21)	0.68	0.09
	AI-ST	W_4, $p=4$, Eq. (22)	0.62	0.10
Sentences	fwSNRseg	ANSI (Table I)	0.78	0.11
	fwSNRseg	Eq. (6), $p=2$	0.80	0.10
	NCM_{LF}	$W_i=1$	0.81	0.10
	NCM_{LF}	$W_i^{(1)}$, $p=1.5$, Eq. (12)	0.87	0.09
	NCM	ANSI (Table II)	0.84	0.09
	NCM	$W_i^{(1)}$, $p=1.5$, Eq. (12)	0.89	0.08
	NCM	$W_i^{(2)}$, $p=0.25$, Eq. (13)	0.86	0.08
	AI-ST	ANSI (Table I)	0.43	0.16
	AI-ST	W_1, Eq. (19)	0.66	0.13
	AI-ST	W_2, $p=3$, Eq. (20)	0.83	0.10
	AI-ST	W_3, $p=3$, Eq. (21)	0.83	0.10
	AI-ST	W_4, $p=4$, Eq. (22)	0.73	0.11

signal-specific weighting information have been found to perform the best in terms of predicting speech intelligibility in noise. The CSII measures have been found previously to correlate highly with both speech intelligibility (Kates and Arehart, 2005) and speech quality (Arehart *et al.*, 2007), at least for sentence materials subjected to hearing-aid type of distortions (e.g., clipping). On this regard, the present study extends the utility of the CSII measures for the prediction of the intelligibility of noise-suppressed speech. The proposed band-importance functions [Eq. (22)] had a big influence on the performance of the modified CSII measures, particularly for the prediction of sentence intelligibility scores. The correlation coefficient of the CSII_{mid} measure with sentence recognition scores, in particular, improved from $r=0.92$ to $r=0.94$ after using the proposed BIF given in Eq. (22). Similar improvement was noted for consonants, but only for the CSII_{low} measure. The lack of improvement for the CSII_{mid} measure can be attributed to the non-uniform, and perhaps skewed, distribution of segments falling in the three regions, at least for the consonant materials used in this study (note that for sentences, a roughly equal number of segments fall in the three regions). Only a small percentage ($<16\%$) of segments were found to be classified as mid-level, suggesting that perhaps different regions need to be considered for consonants. Further work is thus warranted to optimize the selection of regions for isolated vowel-consonant–vowel syllables.

High performance was expected of the NCM measure as it belongs to the speech-based STI measures, which have been shown in many studies to correlate highly with the intelligibility of nonsense syllables (e.g., Steeneken and Houtgast, 1982; Houtgast and Steeneken, 1985). The speech-based STI measures (Goldsworthy and Greenberg, 2004) generally assess the amount of reduction in temporal-envelope modulations incurred when the input signal goes through a sound transmission system. In our case, the NCM measure [Eq. (8)] assesses the fraction of the processed envelope signal that is linearly dependent on the input (clean) envelope signal at each frequency band. This measure accounts for the average envelope power in each band as well as for the low-frequency (<12.5 Hz) envelope modulations, which are known to carry critically important information about speech (e.g., Drullman *et al.*, 1994a, 1994b; Arai *et al.*, 1996). Compared to the conventional NCM measure (Hollube and Kollmeier, 1996) which uses fixed (for all speech stimuli) weights, the modified NCM measure uses signal-dependent weighting functions and performed substantially better. Overall, the proposed BIFs [Eqs. (12) and (13)] had a big influence on the performance of the modified NCM measure. The correlation coefficient obtained with the consonant materials improved from 0.66 when fixed ANSI (1997) weights were used to 0.77 when the signal-dependent weighting function given in Eq. (12) was used (Table III). Similar improvements were also noted on the sentence recognition task (Table IV). Aside from the use of the proposed BIFs, the use of a wider speech dynamic range (45 dB) improved slightly the performance of the NCM measure (see

TABLE IX. Correlation coefficients between consonant recognition scores and the AI-ST measure as a function of window duration (in milliseconds), SNR range, and BIF.

SNR range	Band-importance function	Window duration			
		30 ms	60 ms	100 ms	125 ms
$[-15,15]$ dB	W_1, Eq. (19)	0.56	0.56	0.59	0.59
	W_2, $p=1$, Eq. (20)	064	0.64	0.66	0.68
	W_3, $p=2$, Eq. (21)	0.65	0.64	0.66	0.66
	W_4, $p=2$, Eq. (22)	0.51	0.53	0.56	0.59
$[-10,35]$ dB	W_1, Eq. (19)	0.57	0.55	0.57	0.58
	W_2, $p=1$, Eq. (20)	0.66	0.65	0.66	0.67
	W_3, $p=2$, Eq. (21)	0.67	0.65	0.67	0.67
	W_4, $p=2$, Eq. (22)	0.60	0.61	0.63	0.64

TABLE X. Correlation coefficients between sentence recognition scores and the AI-ST measure as a function of window duration (in milliseconds), SNR dynamic range, and BIF.

SNR range	Band-importance function	Window duration			
		30 ms	60 ms	100 ms	125 ms
$[-15,15]$ dB	W_1, Eq. (19)	0.66	0.71	0.75	0.76
	W_2, $p=1$, Eq. (20)	0.77	0.81	0.84	0.85
	W_3, $p=2$, Eq. (21)	0.79	0.82	0.85	0.86
	W_4, $p=2$, Eq. (22)	0.67	0.68	0.71	0.73
$[-10,35]$ dB	W_1, Eq. (19)	0.66	0.71	0.74	0.75
	W_2, $p=1$, Eq. (20)	0.80	0.83	0.85	0.86
	W_3, $p=2$, Eq. (21)	0.82	0.84	0.86	0.86
	W_4, $p=2$, Eq. (22)	0.71	0.73	0.75	0.77

Table VII). However, neither the use of a wider range of modulation frequencies (see Table VI) nor the use of smaller number of channels (see Table V) influenced significantly the performance of the NCM measure. The power exponent (p) used in the BIFs can be clearly optimized for different speech materials, but only a slight dependence on the specific value of the power exponent was observed (see Table XI), at least for the NCM measure.

The performance of the proposed low-frequency (100–1000 Hz) version of the NCM measure [see Eq. (15)] was comparable to that of the NCM measure. This suggests that the low-frequency region of the spectrum carries critically important information about speech. The low-frequency region of the spectrum is known to carry F1 and voicing information, which in turn provides listeners with access to low-frequency acoustic landmarks of the signal (Li and Loizou, 2008). These landmarks, often blurred in noisy conditions, are critically important for understanding speech in noise as it aids listeners to better determine syllable structure and word boundaries (Stevens, 2002; Li and Loizou, 2008).

The performance of the AI-ST measure was modest and comparable to that obtained with the PESQ measure. Higher performance was expected with the AI-ST measure, at least for predicting consonant recognition in noise, given the success of the AI index in predicting the intelligibility of nonsense syllables (e.g., Kryter, 1962b). Our implementation, however, was rather simplistic as it did not incorporate upward spread of masking or any other non-linear auditory effects modeled in the ANSI (1997) standard. Furthermore, the AI-ST measure operates at a short, segmental (phonetic) level, while the SII measure operates on the average long-term spectra of the target and masker signals. Operating at a short-term (segmental) level was found necessary in the present study in order to capture the changing temporal/spectral characteristics of fluctuating maskers (e.g., train), but it imposes some limitations on the AI-ST measure that are difficult to overcome. For one, the segmental AI-ST values were averaged over all segments to produce one value. In doing so, it is implicitly assumed that all short (phonetic) segments should be weighted uniformly, i.e., that equal emphasis should be placed on consonant segments, steady-state vowels, and/or vowel-consonant transitions. Since our knowledge is limited as to how normal-hearing listeners integrate over time vowel and consonant information for sentence recognition, one can consider devising separate BIFs that are more appropriate for vowels and consonants. A better temporal weighting function, perhaps one derived psychoacoustically and incorporating forward/backward masking effects (e.g., Rhebergen *et al.*, 2006), might be needed to improve further the performance of the AI-ST measure.

The performance of the AI-ST measure on the prediction of sentence intelligibility in noise was higher than that on consonant intelligibility. This was surprising since the AI-ST measure as well as the other measures examined in this study do not model contextual or any other high-level (involving central processes) effects, which are known to play a significant role on sentence recognition. We speculate that this was accomplished, or perhaps compensated, by the use of signal-dependent BIFs. In the absence of those functions, the performance of the AI-ST measure on the sentence recognition task was found to be poor ($r<0.4$).

The data shown in Tables III and IV clearly demonstrate that the performance of the AI-ST measure depends largely on the choice of the BIF. The BIF given in Eq. (20), in particular, was found to work the best on both consonant and sentence recognition tasks. The performance, for instance, of the AI-ST measure when applied to sentence recognition improved from $r=0.33$ with ANSI (1997) weights to $r=0.80$ with the proposed BIF given in Eq. (20). The results from the present study clearly suggest that the traditional SII index (ANSI, 1997), as well as the STI index, could benefit from

TABLE XI. Correlation coefficients, r, and standard deviations of the error, σ_e, between sentence recognition scores and the NCM measure as a function of the power exponent, p, used in the BIF in Eq. (12).

Power exponent, p	r	σ_e
0.12	0.85	0.09
0.25	0.87	0.08
0.50	0.89	0.08
0.62	0.89	0.08
0.75	0.89	0.08
1.00	0.89	0.08
1.50	0.89	0.07

the use of signal-dependent BIFs, such as those given in Eqs. (19)–(22).

VI. CONCLUSIONS

The present study evaluated the performance of traditional (e.g., SNRseg) as well as new objective measures in terms of predicting speech intelligibility in realistic noisy conditions. The objective measures were tested in a total of 72 noisy conditions which included processed sentences and nonsense syllables corrupted by four real-world types of noise (car, babble, train, and street). The distinct contributions of the present work include the following:

(1) An AI-ST measure was proposed operating on short-term (30 ms) segments. This measure was found to predict modestly ($r=0.68-0.83$) well the intelligibility of speech embedded in fluctuating maskers when the proposed BIFs were used. The performance of the AI-based measure was quite poor ($r=0.33$) when the ANSI (1997) AI weights were used, but improved to $r=0.83$ when the proposed (segment-dependent) BIFs were used.

(2) A low-frequency version of the NCM measure was proposed that incorporates only low-frequency (100–1000 Hz) envelope information in its computation. The correlation obtained with this measure for predicting sentence recognition scores was high ($r=0.87$) and nearly as good as that obtained with the full-bandwidth (300–3400 Hz) NCM measure ($r=0.89$). This outcome provides additional support for the importance of low-frequency (<1000 Hz) acoustic landmarks on speech recognition (Li and Loizou, 2008).

(3) The conventional SNRseg measure, which is widely used for assessing performance of noise-suppression and speaker-separation algorithms, predicted poorly ($r=0.40-0.46$) the intelligibility of consonants and sentences.

(4) The PESQ measure, which was originally designed to predict speech quality, performed modestly well ($r=0.77-0.79$) on predicting speech intelligibility in noise. Of all the conventional subjective quality measures tested, the fwSNRseg and PESQ measures performed modestly well in terms of predicting *both* quality and intelligibility.

(5) The influence of speech dynamic range (varying from 30 to 50 dB), integration window (varying from 30 to 125 ms), number of bands (varying from 7 to 20 bands), and range of modulation frequencies (varying from 12.5 to 30 Hz) was assessed on the performance of the AI-based and STI-based (i.e., NCM) measures. Of all these parameters, only the use of a wider dynamic range (45–50 dB) improved somewhat the correlation of the NCM and AI-ST measures. Increasing the window duration also improved the correlation of the AI-ST measure in predicting sentence recognition (Table X).

(6) Of all parameters examined in this study, the BIFs influenced the performance of the AI-based, STI-based (NCM), and coherence-based (CSII) measures the most. The proposed signal and phonetic-segment dependent BIFs [Eqs. (19)–(22)] were found to be suitable for predicting the intelligibility of speech in fluctuating maskers. Additional flexibility is built in the proposed band-importance functions for emphasizing spectral peaks and/or spectral valleys. The proposed BIFs improved consistently the performance of all three sets of measures. This outcome clearly suggests that the traditional SII index (ANSI, 1997) as well as the STI index could benefit from the use of signal-dependent band-importance functions, such as those proposed in Eqs. (19)–(22).

(7) Among all objective measures examined in the present study, the modified CSII and NCM measures incorporating signal-specific weighting information have been found to perform the best in terms of predicting speech intelligibility in noise. The modified CSII_{mid} measure, in particular, which only includes vowel/consonant transitions and weak consonants in its computation, yielded the highest correlation ($r=0.94$) with sentence recognition scores. This outcome further corroborates the large contribution of weak consonants on speech recognition in noise (Li and Loizou, 2008).

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC007527 (P.C.L.) and R03 DC008887 (Y.H.) from the National Institute of Deafness and other Communication Disorders, NIH.

APPENDIX

The MSC function is given by

$$\text{MSC}(\omega) = \frac{|S_{XY}(\omega)|^2}{S_{XX}(\omega)S_{YY}(\omega)}. \tag{A1}$$

Let the MTF at frequency ω be given by (Drullman *et al.*, 1994b)

$$\text{MTF}(\omega) = \alpha\sqrt{\frac{S_{YY}(\omega)}{S_{XX}(\omega)}}, \tag{A2}$$

where α is a normalization factor, and let $W(\omega)$ be the following weighting function at frequency ω:

$$W(\omega) = \frac{1}{\alpha}\frac{|S_{XY}(\omega)|^2}{\sqrt{S_{XX}(\omega)}(S_{YY}(\omega))^{3/2}}. \tag{A3}$$

Then, the MSC function can be written as a weighted MTF, i.e.,

$$\text{MSC}(\omega) = W(\omega)\cdot\text{MTF}(\omega). \tag{A4}$$

Allen, J. B. (**1994**). "How do humans process and recognize speech," IEEE Trans. Speech Audio Process. **2**, 567–577.

Anderson, W. B., and Kalb, J. T. (**1987**). "English verification of STI method for estimating speech intelligibility of a communications channel," J. Acoust. Soc. Am. **81**, 1982–1985.

ANSI (**1997**). "Methods for calculation of the speech intelligibility index," S3.5–1997 (American National Standards Institute, New York).

Arai, T., Pavel, M., Hermansky, H., and Avendano, C. (**1996**). "Intelligibility of speech with filtered time trajectories of spectral envelopes," in *Proceedings of the ICSLP*, pp. 2490–2493.

Arehart, K., Kates, J., Anderson, M., and Harvey, L. (**2007**). "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 1150–1164.

Beerends, J., Larsen, E., Lyer, N., and van Vugt, J. (**2004**). "Measurement of speech intelligibility based on the PESQ approach," in Proceedings of the Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN), Prague, Czech Republic.

Beerends, J., van Wijngaarden, S., and van Buuren, R. (**2005**). "Extension of ITU-T recommendation P.862 PESQ towards measuring speech intelligibility with vocoders," in *New Directions for Improving Audio Effectiveness*, Proceedings of the RT0-MP-HFM-123, Neuilly-sur-Seine, France, pp. 10-1–10.6.

Bladon, R., and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.

Boothroyd, A., Erickson, F. N., and Medwetsky, L. (**1994**). "The hearing aid input: A phonemic approach to assessing the spectral distribution of speech," Ear Hear. **6**, 432–442.

Brachmanski, S. (**2004**). "Estimation of logatom intelligibility with STI method for polish speech transmitted via communication channels," Arch. Acoust. **29**, 555–562.

Carter, C., Knapp, C., and Nuttall, A. (**1973**). "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," IEEE Trans. Audio Electroacoust. **AU-21**, 337–344.

Cohen, I., and Berdugo, B. (**2002**). "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Process. Lett. **9**, 12–15.

Drullman, R., Festen, J., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Drullman, R., Festen, J., and Plomp, R., (**1994b**). "Effect of reducing slow temporal modulations on speech reception" J. Acoust. Soc. Am. **95**, 2670–2680.

Dunn, H., and White, S. (**1940**). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. **11**, 278–288.

Ephraim, Y., and Malah, D. (**1985**). "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-33**, 443–445.

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

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

Goldsworthy, R., and Greenberg, J. (**2004**). "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," J. Acoust. Soc. Am. **116**, 3679–3689.

Gustafsson, H., Nordholm, S., and Claesson, I. (**2001**). "Spectral subtraction using reduced delay convolution and adaptive averaging," IEEE Trans. Speech Audio Process. **9**, 799–807.

Hansen, J., and Pellom, B. (**1998**). "An effective quality evaluation protocol for speech enhancement algorithms," in Proceedings of the International Conference on Spoken Language Processing, Vol. **7**, pp. 2819–2822.

Hirsch, H., and Pearce, D. (**2000**). "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *ISCA Tutorial and Research Workshop ASR2000*, Paris, France.

Hohmann, V., and Kollmeier, B. (**1995**). "The effect of multichannel dynamic compression on speech intelligibility," J. Acoust. Soc. Am. **97**, 1191–1195.

Hollube, I., and Kollmeier, K. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am. **100**, 1703–1715.

Houtgast, T., and Steeneken, H. J. M. (**1971**). "Evaluation of speech transmission channels by using artificial signals," Acustica **25**, 355–367.

Houtgast, T., and Steeneken, H., (**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.

Hu, Y., and Loizou, P. C. (**2003**). "A generalized subspace approach for enhancing speech corrupted by colored noise," IEEE Trans. Speech Audio Process. **11**, 334–341.

Hu, Y., and Loizou, P. C. (**2004**). "Speech enhancement based on wavelet thresholding the multitaper spectrum," IEEE Trans. Speech Audio Process. **12**, 59–67.

Hu, Y., and Loizou, P. C., (**2007**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**, 1777–1786.

Hu, Y., and Loizou, P. C. (**2008**). "Evaluation of objective quality measures for speech enhancement," IEEE Trans. Audio, Speech, Lang. Process. **16**, 229–238.

IEC 60268-16 (**2003**). "Sound system equipment—Part 16: Objective rating of speech intelligibility by speech transmission index," Ed. 3 (International Electrotechnical Commission, Geneva, Switzerland).

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

ITU-T (**2000**). "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU-T Recommendation P. 862.

Jabloun, F., and Champagne, B. (**2003**). "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Process. **11**, 700–708.

Kamath, S., and Loizou, P. C. (**2002**). "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL.

Kates, J. (**1987**). "The short-time articulation index," J. Rehabil. Res. Dev. **24**, 271–276.

Kates, J. (**1992**). "On using coherence to measure distortion in hearing aids," J. Acoust. Soc. Am. **91**, 2236–2244.

Kates, J., and Arehart, K. (**2005**). "Coherence and the speech intelligibility index," J. Acoust. Soc. Am. **117**, 2224–2237.

Kitawaki, N., Nagabuchi, H., and Itoh, K. (**1988**). "Objective quality evaluation for low bit-rate speech coding systems," IEEE J. Sel. Areas Commun. **6**, pp. 262–273.

Klatt, D. H. (**1982**). "Prediction of perceived phonetic distance from critical-band spectra: A first step," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. **2**, pp. 1278–1281.

Kryter, K. D. (**1962a**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Kryter, K. D. (**1962b**). "Validation of the articulation index," J. Acoust. Soc. Am. **34**, 1698–1706.

Larm, P., and Hongisto, V. (**2006**). "Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index," J. Acoust. Soc. Am. **119**, 1106–1117.

Li, N., and Loizou, P. (**2008**). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise," J. Acoust. Soc. Am. **124**, 498–509.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC, Boca Raton, FL).

Ludvigsen, C., Elberling, C., and Keidser, G. (**1993**). "Evaluation of a noise reduction method—Comparison of observed scores and scores predicted from STI," Scand. Audiol. Suppl. **38**, 50–55.

Ludvigsen, C., Elberling, C., Keidser, G., and Poulsen, T. (**1990**). "Prediction of intelligibility of non-linearly processed speech," Acta Oto-Laryngol., Suppl. **469**, 190–195.

Mapp, P. (**2002**). "A comparison between STI and RASTI speech intelligibility measurement systems," in The 111th AES Convention, Los Angeles, CA, Preprint No. 5668.

Moore, B., and Glasberg, B. (**1993**). "Suggested formulas for calculation auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Quackenbush, S. R., Barnwell, T. P., and Clements, M. A. (**1988**). *Objective Measures of Speech Quality*, (Prentice-Hall, Englewood Cliffs, NJ).

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.

Rix, A., Beerends, J., Hollier, M., and Hekstra, A. (**2001**). "Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. **2**, pp. 749–752.

Scalart, P., and Filho, J. (**1996**). "Speech enhancement based on a priori signal to noise estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 629–632.

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

Steeneken, H., and Houtgast, T. (**1982**). "Some applications of the speech transmission index (STI) in auditoria," Acustica **51**, 229–234.

Stevens, K. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Studebaker, G., and Sherbecoe, R. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–1436.

van Buuren, R., Festen, J., and Houtgast, T. (**1999**). "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," J. Acoust. Soc. Am. **105**, 2903–2913.

Van Wijngaarden, S., and Houtgast, T. (**2004**). "Effect of talker and speaking style on the speech transmission index," J. Acoust. Soc. Am. **115**, 38L–41L.

Relative-pitch tracking of multiple arbitrary sounds

Paris Smaragdis[a]
Adobe Systems Inc., 275 Grove Street, Newton, Massachusetts 02466

(Received 29 July 2008; revised 22 February 2009; accepted 1 March 2009)

Perceived-pitch tracking of potentially aperiodic sounds, as well as pitch tracking of multiple simultaneous sources, is shown to be feasible using a probabilistic methodology. The use of a shift-invariant representation in the constant-*Q* domain allows the modeling of perceived pitch changes as vertical shifts of spectra. This enables the tracking of these changes in sounds with an arbitrary spectral profile, even those where pitch would be an ill-defined quantity. It is further possible to extend this approach to a mixture model, which allows simultaneous tracking of varying mixed sounds. Demonstrations on real recordings highlight the robustness of such a model under various adverse conditions, and also show some of its unique conceptual differences when compared to traditional pitch tracking approaches.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3106529]

PACS number(s): 43.75.Zz, 43.66.Hg [NHF] Pages: 3406–3413

I. INTRODUCTION

Pitch tracking has long been a fascinating subject in musical acoustics. This is a problem, which has been tackled using a rich variety of approaches and continues to inspire a considerable amount of research. Approaches to pitch track extraction have ranged from straightforward period estimation, to sophisticated statistical methods, some employing time domain techniques and others sophisticated front-ends that reveal more of the pitch structure.[1–8] The applications of pitch tracking cover a wide range of applications ranging from musical transcription, to emotion recognition in speech, to animal acoustics. To facilitate such a wide variety of applications various biases are often imposed to facilitate a reasonable answer for the domain at hand. In this paper we present a general approach to tracking a pitch-like measure, which makes minimal assumptions about the nature of the input sound, or the kind of pitch content at hand.

We present an additive shift-invariant decomposition, which when coupled with a constant-*Q* analysis front-end can be used to track movements of spectral structures along the log-frequency space. Such movements correlate very strongly to how we perceive pitch, and can be used to infer relative-pitch changes. Using this model we set forth to address a number of issues. The primary goal is to present a formulation, which allows *soft* decisions that do not result in deterministic estimates, but rather a probability distribution describing the relative likelihood of these shifts along the frequency axis. This probabilistic approach, which is most valuable when designing systems with input of high uncertainty, also provides an easy way to extend such a system by using statistical methods that take advantage of domain knowledge that can further help achieve robust performance. An additional point we wish to address is that of tracking in the case of mixtures. The assumption of clean input sounds is rarely valid in real recordings, and often we need to compute pitch tracks of either noisy or multiple sources. The model we present is additive by design, so that multiple overlapping frequency shifts can be tracked simultaneously. This allows us to process inputs with multiple sources without serious complications. Finally, using this particular representation allows us to deal with unpitched or inharmonic sounds whose absolute pitch value is hard to pinpoint, yet they can be used to a melodic effect. Examples of such cases are chords, or certain percussive instrument sounds (e.g., cymbals) that individually do not have a strong pitch characteristic, but once used in a certain succession they can invoke the percept of a melody or tonality. In these cases, the tracked shifts along the frequency axis provide an indication of a likely perceived melody, something that methods based on harmonicity assumptions would not be able to provide. Through various experiments we show that this approach can deal with sound sources, which have challenging spectral profiles, as well as sources that exhibit a dynamic spectral character,

The remainder of this paper is structured as follows. We will begin by describing a frequency shifting approach to modeling pitch changes, we briefly discuss the constant-*Q* transform and its utility for our purposes, we will then introduce the computational details of our approach, and then demonstrate it with a variety of pitch tracking situations that highlight its abilities to overcome difficult situations.

II. A SPECTRAL SHIFTING APPROACH TO MODELING PITCH CHANGES

Pitch, especially as we perceive it, is an elusive concept. Most trivially we can link it to the fundamental vibrating frequency of a sound-generating object. However, it is very easy to find examples of aperiodic, or otherwise harmonically incoherent sounds where this assumption can break. Because pitch is hard to estimate, and in some cases nonexistent, attempting to construct pitch tracks in terms of a series of instantaneous pitches is an inherently risky endeavor. Instead, in this paper, we use a different approach, which we argue is more akin to how we perceive pitch.

We will approach the pitch tracking problem as a frequency-shift problem. Instead of trying to estimate abso-

[a] Electronic mail: paris@adobe.com

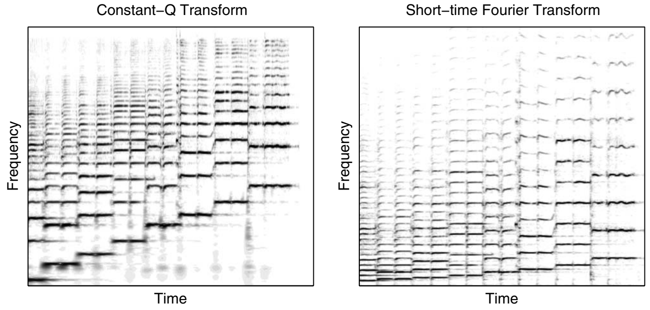


FIG. 1. A comparison between the constant-Q and the short-time Fourier transform. The input is a recording of an arpeggio performed by a violin. Note how in the constant-Q transform shown on the left the individual harmonics of the violin sound maintain their relative distance regardless of the note being played, whereas in the short-time Fourier transform they get spread apart as the notes move to higher frequencies.

lute pitch quantities at every point in time, we will instead track relative changes of pitch across time. This is very similar to how most of us perceive pitch where we can note relative changes but not necessarily actual values. Aside from this connection, more importantly we sidestep the issue of defining and estimating the actual pitch. Instead of pitch measurements, we track the movement of spectral structures along the log-frequency axis. These shifts correlate very much to our perception of a pitch change and can be directly used to infer pitch movements. This can also allow us to deal with inharmonic or aperiodic sounds, which in isolation do not have a clearly defined pitch, but when modulated create that percept. To accommodate this broader notion of pitch we will be using the term *spectral pitch* to indicate this particular movement across the frequency axis.

In Secs. III and IV we will describe the representation that can reveal this modulation, and the machinery involved in detecting it.

III. CONSTANT-*Q* REPRESENTATIONS OF SOUNDS

The constant-Q transform is a time/frequency decomposition that exhibits a logarithmic frequency scale.[9] It is defined so that each octave of available frequencies spans the same number of frequency bins. Examining the magnitude of the constant-Q transforms results in visualizing the amount of acoustic energy at any point in the time/frequency plane. For the remainder of this paper we will be referring to the magnitude of the constant-Q transform and discard the phase information.

A very important property of this type of transformation is that changes in spectral pitch can be clearly visualized as shifts along the frequency axis. An example of this, as contrasted with the short-time Fourier transform, is shown in Fig. 1. On the left we show the constant-Q transform of an arpeggio performed on a real violin, and on the right its equivalent through a short-time Fourier transform. Upon closer examination it is easy to see that the note changes in the constant-Q plot are represented as vertical shifts of approximately the same spectral shape. For the short-time Fourier transform the spacing between the individual harmonics becomes wider for higher notes. This observation will be our starting point in defining the tracking model in this paper. Noting that in the constant-Q transform, the major variation that distinguishes different notes of the same instrument is a simple shift along the frequency axis, we will endeavor to track it and interpret it as a pitch movement.

An underlying assumption in this model is that the spectral shape of an individual sound is relatively constant as it changes pitch, so that the measurement of the shift is feasible. Theoretical arguments on that point are difficult to make since they rely on the expected statistics on the inputs, but as we will demonstrate later on this assumption holds well for sounds with widely varying spectral character.

Another point we need to make here is that of the approximate additivity of the magnitude constant-Q transform. The actual transform results in a complex valued output and is a linear operation, which maintains that the transform of the sum of two signals equals the sum of the transforms of the two signals. When we compute the magnitude of the transform, however, there is no guarantee of linearity since for pair of any complex numbers $\{z_1,z_2\}\in\mathbb{C}$ we have $\|z_1\| + \|z_2\| \neq \|z_1+z_2\|$. However, when observing mixtures of multiple sounds, there is often a high degree of disjointedness in their spectra and the likelihood of both sounds being significantly active at the same time/frequency cell is often very low. In addition to that we seldom observe complete phase cancellations so even in the cases where significant energy overlaps we still have an effect approximate to addition. This assumption has been very commonly used for multiple audio processing systems and is generally understood to be valid for practical purposes. Under this assumption, when we observe the mixture of multiple notes, we will expect the observed constant-Q transform to be composed out of the addition of constant-Q transforms that are appropriately composed out of shifted spectra, denoting each instrument or each note being played. This complicates the operation we wish to resolve, by requiring that we track potentially multiple spectra, that shift independently. If the input is composed of the same sound exhibiting multiple simultaneous pitches (such a polyphonic piano passage), then we would observe the same spectrum being shifted and overlaid accordingly for each note. If we have multiple instruments we would expect each instrument to have its own spectral shape, which shifts and overlays according to its melodies. In Sec. IV we will present an algorithm that allows us to track these simultaneous shifting movements, and help us interpret them as a relative spectral pitch change.

IV. A MODEL FOR TRACKING SPECTRAL PITCH SHIFTS

The computational model we will use in this section is the one developed in Ref. 10. For reasons that will become clearer later, we will be interpreting the magnitude constant-Q transform as a probability distribution. The contents at the time and frequency coordinates t,ω will be interpreted as an arbitrary scaling of the probability of existence of energy at that point. Due to this we will notate constant-Q transforms as $P(\omega,t)$ and assume the proper scaling so that they integrate to unity.

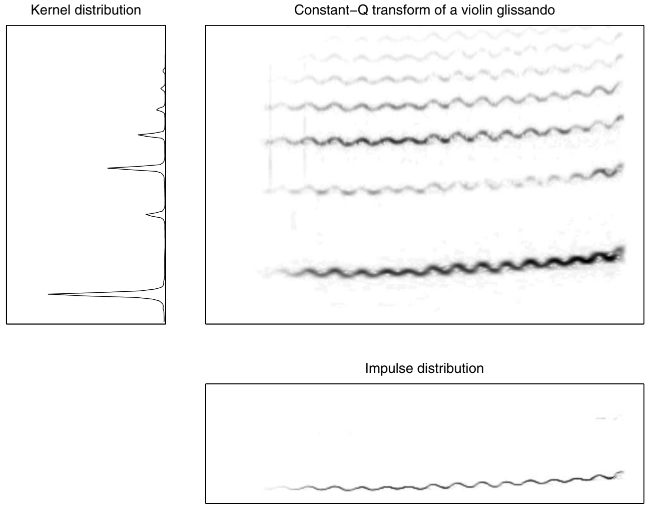


FIG. 2. An illustrative analysis of the constant-Q transform in the top right plot. The top left plot shows the extracted kernel distribution and the bottom plot shows the corresponding impulse distribution. The impulse distribution can be used to represent the spectral pitch changes in the input.

A. The single source formulation

Starting with this model we wish to discover shift-invariant components across the ω dimension. In the simple case where we assume one shifting spectrum we can notate this model as

$$P(\omega,t) = P_K(t) * P_I(f_0,t), \qquad (1)$$

where $P(\omega,t)$ is the input magnitude constant-Q transform, $P_K(\omega)$ is a frequency distribution, $P_I(f_0,t)$ is a time/frequency distribution, and the star operator denotes convolution (note that the convolution is two dimensional since the second operant is of that rank). We will refer to $P_K(\omega)$ as the *kernel distribution*, and $P_I(f_0,t)$ as the *impulse distribution*. Their interpretation is rather straightforward. The kernel distribution $P_K(\omega)$ is a frequency distribution, i.e., a constant-Q spectrum, or rather a prototypical vertical slice of a constant-Q transform. The impulse distribution $P_I(f_0,t)$ is a time/frequency distribution, which gets convolved with the kernel distribution. Due to this relationship we can interpret the impulse distribution as expressing the likelihood that the kernel distribution will take place at any given frequency shift or time. To illustrate this concept consider the case in Fig. 2. In the top right panel we show the constant-Q transform of a recording of a real violin performing a glissando with vibrato. The ideal decomposition given our model is also shown. To the left we see the kernel distribution, which denotes the spectral character of the violin, and in the bottom plot we see the corresponding impulse distribution. Convolving these two distributions we would approximate the input. It is easy to see that the impulse distribution graphically represents the frequency-shift variations in a very convenient manner. In fact, if we assume that the present instrument is well defined by the kernel distribution we can interpret the impulse distribution as a probability distribution of frequency shift, and by extension spectral pitch, across time. Because we also learn the actual spectral character of the input sound, we do not impose any requirements that the source has to be harmonic or otherwise structured, as long as the spectral pitch change is characterized by a shift in the frequency axis. As we will demonstrate later on this allows us to deal with arbitrary sounds very easily.

At this point, this model is quite closely related to the one in Ref. 4, and any such similar approach that employs shift-tracking on a log-frequency scale. The main points of difference between what we present and these approaches can be summarized in two points. First, we do not make the assumption that the signal we track is harmonic. Unlike past work we do not assume a known harmonic template whose movements are being tracked. In our framework, we allow for the flexibility to learn the particular spectral structure that characterizes the sound we track, which, as we show later, can allow us to use this approach even in cases where what we track is not clearly defined as pitch. Secondly, such techniques traditionally use cross-correlation to find the most likely placement of a harmonic series along the log-frequency axis. The most likely placement is taken to be the peak of the cross-correlation. In our approach we actually produce a distribution that describes the likelihood of shift, which is a much more informative measure, especially when it needs to be incorporated into a larger reasoning system. The implied non-negativity in this operation also means that we will not be obtaining negative cross-correlation values, which can obscure the interpretation of this operation, and make use of cross-cancellations, which can impede finding the true function peak. This last argument will become increasingly more important as we move into multi-source formulations in Sec. IV B.

B. The multi-source formulation

In the case where we expect to encounter multiple sounds with different spectra, we can generalize the above model to

$$P(\omega,t) = \sum_{z=1}^{R} P(z) P_K(\omega|z) * P_I(f_0,t|z). \qquad (2)$$

The difference with Eq. (1) is that now we use a latent variable z as an index to allow us having R distinct kernel distributions $P_K(\omega|z)$, each with its own corresponding shifting pattern denoted by R impulse distributions $P_I(f_0,t|z)$. We also introduce a prior distribution $P(z)$, which allows us to arbitrarily weight these pairs of convolution operands to approximate the input. This is essentially an additive generalization of the previous model, which allows the simultaneous tracking of spectral pitches by multiple sources with distinct spectral characters. Each kernel distribution conditional on z will describe the spectral shape of each source, and each impulse distribution conditional on z will describe its corresponding spectral pitch track. The priors' distribution will effectively denote the mixing proportions of each spectral template, or more simply how much of it we will observe in relation to the others. The choice of R is up to the user. For $R=1$ this model collapses to the model in Sec. IV A, and tracks the movement of a single spectral template across the log-frequency space. If we know that the input we are analyzing is constructed by multiple and distinctly different spectra then we can set R to their count so that we can simultaneously track the shifts of multiple spectral templates at the same time.

C. Learning the model

In order to estimate the unknown distributions $P_K(\omega|z)$, $P_I(f_0,t|z)$, and $P(z)$ in the above model we can use the expectation-maximization (EM) algorithm.[11] We first rewrite the model in order to express the convolution in a more explicit manner as

$$P(\omega,t) = \sum_{z=1}^{R} P(z) \sum_{f_0} P_K(\omega - f_0|z) P_I(f_0,t|z). \quad (3)$$

EM estimation will break down the learning process in an iterative succession of two steps. The first step, the E-step, computes the contribution of each spectral template to the overall model reconstruction by

$$Q(\omega,t,f_0,z) = \frac{P(z)P_K(\omega - f_0|z)P_I(f_0,t|z)}{\sum_{z'} P(z')\sum_{f_0'} P_K(\omega - f_0'|z')P_I(f_0',t|z')}. \quad (4)$$

This results in a "weighting" factor, which tells us how important each kernel distribution is in reconstructing the input at any possible shift in the time/frequency space. During the second step, the M-step, we estimate the wanted model distributions by essentially performing matched filtering between each operant and the input weighted by the appropriate $Q(\omega,t,f_0,z)$. The equations for the M-step are

$$P(z)^* = \sum_{\omega}\sum_{t}\sum_{f_0} P(\omega,t)Q(\omega,t,f_0,z),$$

$$P_K(\omega|z^*) = \frac{\sum_t \sum_{f_0} P(\omega + f_0,t)Q(\omega + f_0,t,f_0,z)}{\sum_{\omega'}\sum_t \sum_{f_0} P(\omega' + f_0,t)Q(\omega' + f_0,t,f_0,z)},$$

$$P_I(f_0,t|z)^* = \frac{\sum_{\omega} P(\omega,t)Q(\omega,t,f_0,z)}{P(z)^*}, \quad (5)$$

where the $(\cdot)^*$ notation indicates the new estimate. Iterating over the above steps we converge to a solution after about 30–50 iterations. Although there is no guarantee that this process will find the global optimum, it predominantly converges to qualitatively the same solutions over repeated runs. The dominant variation in these solutions is an arbitrary shift in the spectral distribution, which is counteracted by an opposing shift in the impulse distribution in order to ensure the correct reconstruction. This introduces a variation in the resulting outputs, but not one that interferes with the quality of fit, or (as we examine in Sec. IV E) with the interpretation of the decomposition.

D. Sparsity constraints

Upon closer consideration one can see that the above model is overcomplete. This means that we can potentially have more information in the model itself than we have in the input. This is of course a major problem because it can result in outputs, which have overfitted to the input, or models which are hard to interpret. A particular instance of this problem can be explained using the commutativity property of convolution. Referring back to Fig. 2 we note that an output in which the impulse distribution was identical to the input and the kernel distribution was a delta function would be also an acceptable answer. That particular decomposition would not offer any information at all since the spectral pitch track would have to be identical to the input. Likewise any arbitrary shift of information from one distribution to another that would lie between what we have plotted above and the outcome just described would result in an infinite set of correct solutions.

In order to regulate the potential increase in information from input to output we will make use of an *entropic prior*.[12] This prior takes the form of $P(\theta) \propto e^{-\beta\mathcal{H}(\theta)}$, where $\mathcal{H}(\cdot)$ is entropy and θ can be any distribution from the ones estimated in our model. The parameter β is simply adjusting the amount of bias toward a high or a low entropy preference. A β value, which is less than zero, will bias the estimation toward a high entropy (i.e., flatter) distribution, and a positive value will bias it toward a low entropy (i.e., spikier) distribution. The magnitude of β determines how important this entropic manipulation is so that larger values will put more stress in it, whereas values closer to 0 will not. In the extreme case where $\beta=0$ the prior does not come in effect. Imposing this prior can be done by inserting an additional procedure in the M-step, which enforces the use of this prior. The additional step involves re-estimating the distribution in hand by

$$\theta^{**} = \frac{-(\hat{\theta}^*)/\beta}{\mathcal{W}(-(\hat{\theta}^*)e^{1+\lambda/\beta}/\beta)}, \quad (6)$$

where $\mathcal{W}(\cdot)$ is Lambert's function,[13] θ^{**} is the estimate of θ with the entropic prior, and $\hat{\theta}^*$ is the estimate according to Eq. (5) but without the division by $P(z)^*$. The quantity λ comes from a Lagrangian due to the constraint that $\Sigma\theta_i=1$, which results in the expression

$$\lambda = -\left[\frac{\hat{\theta}^*}{\theta_i^{**}} + \beta + \beta \log \theta_i^{**}\right]. \quad (7)$$

These two last equations can be repeatedly evaluated in succession and they converge to a stable estimate of θ after a small number of iterations. This prior and the computational details as they relate to this model involved are described in more detail in Ref. 14.

With the problem at hand we would want to have a high entropy kernel distribution and a low entropy impulse distribution. This will result in a frequency-shift track, which will be as fine as possible, and a spectrum estimate, which will account for most of the energy of the source's spectrum. To illustrate the effect of this prior consider the different outcomes shown in Fig. 3. The input was the same as in Fig. 2. The top plots show the results we obtain using the entropic prior, whereas the bottom plots show the results we get if we do not use it. When using the entropic prior we bias the

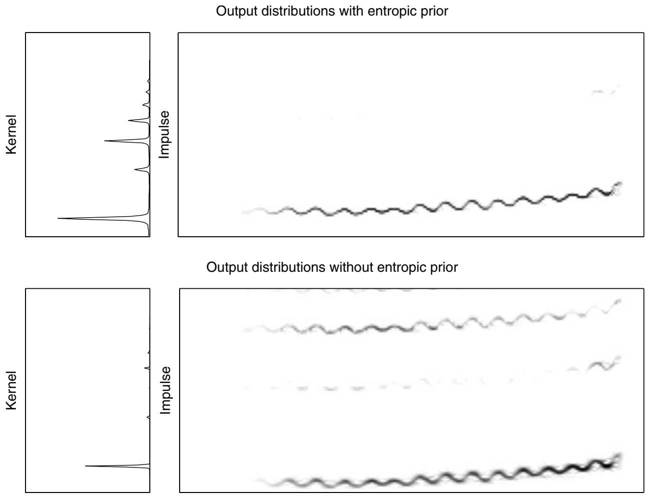


FIG. 3. Illustrating the effect of the entropic prior. The top figures show the output we obtain when analyzing the example in Fig. 2 and employ the entropic prior with a high entropy constraint on the kernel distribution and a low entropy constraint on the impulse distribution. The pair of the bottom plots shows the kernel and impulse distributions learned without using the prior.

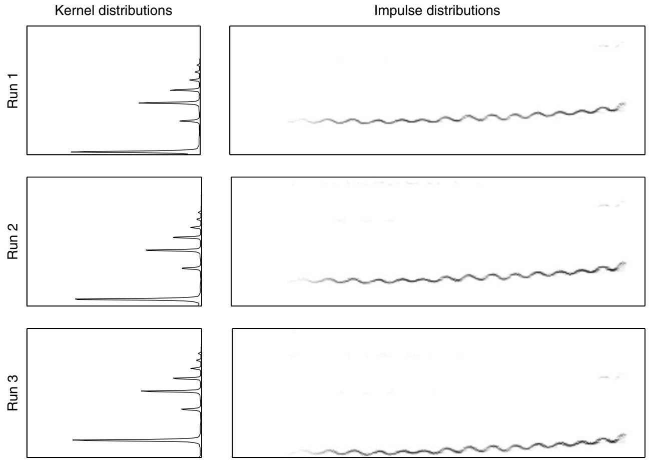


FIG. 4. Results from multiple runs on the input shown in Fig. 2. Notice that although in a qualitative sense the results are identical, there are vertical shifts between them that differentiate them. When the kernel distribution is positioned higher in the frequency axis, the impulse distribution is positioned lower and vice-versa.

learning toward a high entropy kernel and a low entropy impulse. This resulted in a clean and sparse track, as opposed to the non-interpretable one when not using the prior.

For practical reasons we can strengthen the low entropy bias by requiring the height of the impulse distribution to be only as big as the expected pitch range (therefore implying that values outside that range will be zero thus lowering the entropy). This results in faster training requiring convolutions of smaller quantities, but also restricts the possible solutions to the range we are interested in, thus aiding in a speedier convergence. The convolutions involved can also be evaluated efficiently using the fast Fourier transform. Training for the examples in this paper took less than 1 min on a current mid-range laptop computer using a MATLAB implementation of this algorithm.

E. Interpreting the model

Let us now examine exactly how the results of this analysis could be interpreted. As we have stressed before, this approach tracks frequency shifts, which correspond to relative-pitch movements, and does not compute an absolute spectral pitch. This means that at any point in time we do not know what the spectral pitch is, but rather how much it has changed in relation to other parts. To illustrate this idea consider the plots in Fig. 4. Each row of plots displays the results from a different simulation on the input in Fig. 2. Note that although the results appear somewhat identical they still differ by an arbitrary vertical shift. This shift is counterbalanced between the kernel and the impulse distribution such that when they convolve they result in the same output. However, we cannot expect the impulse distribution between multiple runs to exhibit the same shift since the model is shift-invariant. This means that we can recover the relative-pitch changes in the input, but we cannot infer the actual spectral pitch values. If one is inclined to mark the fundamental in the kernel distribution then we can easily obtain the spectral pitch values. However, detecting the fundamental will not always be a trivial pursuit, especially when dealing with noisy or aperiodic kinds of sounds. Regardless, for the scope of this paper, our objective is to detect relative changes and not the actual spectral pitch so we refer to this estimation as something that can be pursued in future work. Another issue that this analysis presents is that of the relationship of the output with the intensity of the input. As the intensity of the input fluctuates across time, we will see a similar fluctuation in the intensity of the impulse distribution across time. This is an effect that can be seen in all the previous plots, especially during the attack portion, where the impulse distribution transitions from faint to strong as it follows the performed crescendo. Additionally, at times where the violin is not performing, we do not observe any intensity in the impulse distribution. This is because the impulse distribution is a distribution over both frequency and time, measuring the presence of the relative amount of presence of the kernel distribution as compared to all time points and frequency shifts. If we are only interested in the frequency offset of the kernel distribution then we need to examine each time slice of the impulse distribution. This is the distribution $P(f_0|t)=P(f_0,t)P(t)$, where $P(t)$ is overall input energy at time t, i.e., $P(t)=\int P(f_0,t)df_0$. To estimate the most likely frequency shift at time t we only need to find the mode of $P(f_0,t)$. However, using $P(f_0,t)$ instead of $P(f_0|t)$ is a more intuitive choice since the estimate we will be normalized by the likelihood that the signal is active at that point in time, and thus also provide us with an amplitude estimate, which we can use for note onset. If we decide to use $P(f_0|t)$ instead the silent sections will be quite uniform indicating that there is no dominant candidate for a frequency shift. This can be interpreted either as an unpitched section, or a silence. In order to avoid this ambiguity we perform the estimation on $P(f_0,t)$, which offers a more user friendly format.

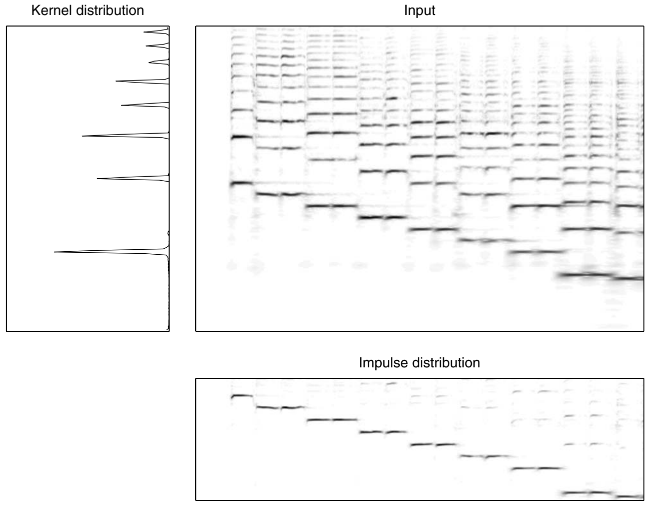


FIG. 5. Analysis of a violin arpeggio. The top right plot is the input constant-Q transform. The left plot is the extracted spectrum of the sound and the bottom plot is the estimated frequency-shift track.

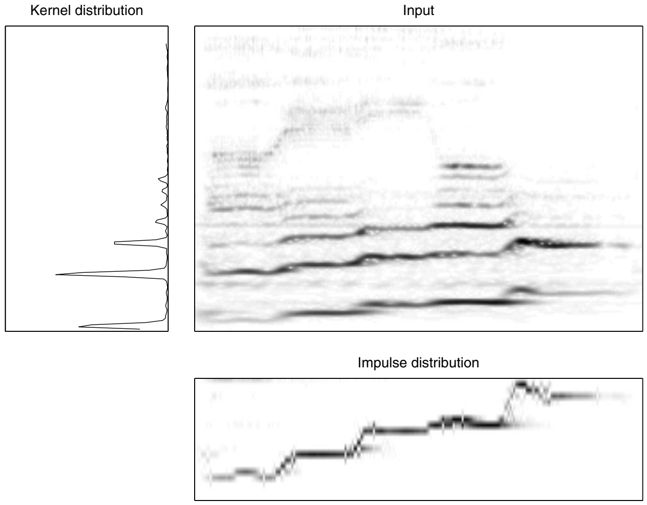


FIG. 6. Example of spectral pitch tracking singing voice with a changing spectral character. The left plot is the extracted spectrum of the input and the bottom plot is the implied spectral pitch track.

V. EXAMPLES

Let us now show how this approach works with some more complex sounds, especially with challenging situations where conventional pitch tracking techniques can result in unreliable, or hard to interpret, estimates.

A. Single source, monophonic examples

In this case we will show results from analyzing two real sound samples containing only one note instance at a time. The first example is a violin recording performing a descending arpeggio with each note (except the first and the last) played twice in succession. The signal was analyzed from a frequency range of 300–8000 Hz. The results of the analysis are shown in Fig. 5. The kernel distribution is clearly a harmonic series (on a constant-Q frequency scale), and the impulse distribution clearly shows the notes that were played. Subtle nuances such as the pitch correction at the beginning of the first note as well as the double note repeats can be easily seen. There are some artifacts as well, mostly in the form of octave echoes, which are more present during the low notes. Since this representation displays the likelihood of spectral pitch these are not necessarily erroneous estimates since they are clearly of lower likelihood than the actual note, and represent the periodicity along the frequency axis of the constant-Q transform. Picking the maximal values of each column of the impulse distribution will easily result in the correct relative spectral pitch estimate at that time. Another more challenging example is shown in Fig. 6. In this case the analyzed source is a vocal recording of the first five notes of a major scale, each note being sung with a different vowel. This experiment is used to test the assumption that the spectrum of the input has to be constant. As is clearly seen in the figure the spectral character of each note is substantially different from the others. Examining the results we observe an averaged spectrum as the converged kernel distribution, and an appropriate frequency-shift track from the impulse distribution. It is important to stress this robustness when dealing with spectrally dynamic sounds since our original assumption of a constant spectrum is unrealistic for real recordings. It is well known that musical instruments exhibit a varying formant character at different registers and that not all notes can be modeled as a simple shift of others. As shown by all the experiments in this paper (and more so by the current one) the constant spectrum assumption in this approach is not very strict at all and does not pose any serious problems with dynamically changing sources. For the final example of the monophonic cases we show how this approach performs when dealing with highly inharmonic sounds. The input in this case was the first four bars of the deep purple recording of the song “Smoke on the Water.” The recording features a well known guitar pattern of an interval of a fifth being appropriately transposed to form a characteristic melody. The guitar sound is highly distorted, which, in addition to the fact that the melody involves multiple notes, creates a highly inharmonic sound that technically does not exhibit pitch (although perceptually it sounds tonal). However, since the same sound is being transposed to form a melody, it is clearly perceived by a human listener as a melodic sequence. Figure 7 shows the results of this analysis. The melodic line is clearly displayed in the impulse distribution, and the spectral profile of the source as represented by the kernel distribution is as expected a highly inharmonic and busy spectral pattern. Similar results can be obtained when using inharmonic or aperiodic sounds, such as cymbals, tom-toms, or bells, without any complications due to their non-harmonic spectral character.

B. Single source, multiple notes

Since the model is additive it is also able to deal with multiple notes. An example of this case is shown in Fig. 8. The input in this case was a country violin recording, which included the simultaneous playing of two notes during most of the time. As expected the analysis of this sound results in an impulse distribution, which has multiple peaks at each time frame that represent the two notes sounding at that

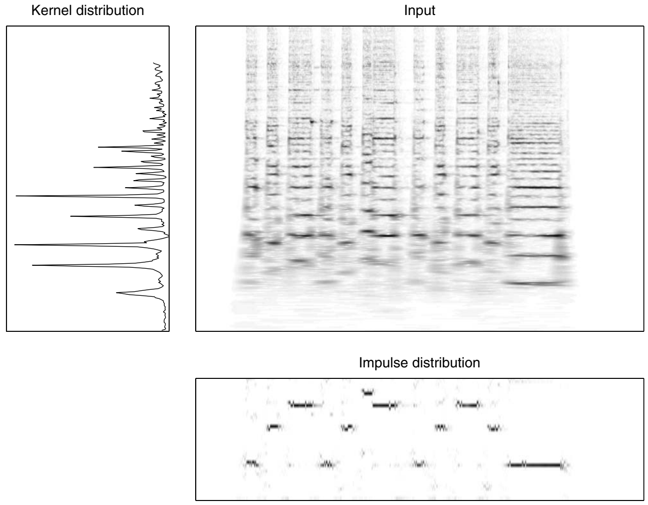


FIG. 7. Analysis of the first four bars of Deep Purple's Smoke on the Water. Despite the absence of a strict pitch at any point in time the transposed chord sequence forms a melody, which we can clearly represent in the impulse distribution.

point. In the impulse distribution it is easy to see the long sustained notes and the simultaneous ornamental fiddle playing.

C. Multiple sources, multiple notes

Finally we demonstrate the ability of this approach to deal with multiple different sources playing different melodies. For this experiment we use as an input a recording of a singing voice accompanied by tubular bells playing a few bars from the round “Frére Jacques.” In this case because the spectral characteristics of the two sources are distinct (the harmonic voice vs the inharmonic tubular bells), we need to perform an analysis in which the latent variable assumes two values. This means that we will be estimating two kernel and impulse distributions, each fitting the pattern of each source. The results of the analysis are shown in Fig. 9. As is evident from the figure the input is a very dense distribution where the included melodies are very hard to spot visually. However, the distinct difference between the two spectra representing the two sounds forces the two kernel distributions to converge to their shape, and helps segment the input in the two instrument parts. Upon examining the impulse distributions we extract we can easily see the structure of the two melodies. Likewise examining the recovered kernel distributions we can see the two different spectra, which represent the characteristics of the two sources in the mixture. The same approach can be applied on mixtures of an arbitrary number of sounds, although as the number of sources increases the disjointedness assumption in the input will gradually be weakened and result in poorer estimates. Having instruments with very similar spectra (e.g., a flute and a clarinet) can also be a problematic case since there will not be sufficient difference between the tracked spectra to easily distinguish them. However, using temporal continuity priors or prior knowledge on the structure of the mixture instruments can allow us to offset that problem. This idea is developed in Ref. 15. If the spectra of the mix instruments are sufficiently different it is also possible to use rank estimation methods, such as the Akaike or Bayesian information criterion, to estimate the number of sources. This is, however, an intrinsically unreliable approach and is only expected to produce good results in cases with a strong contrast in the character of the instruments in the mix.

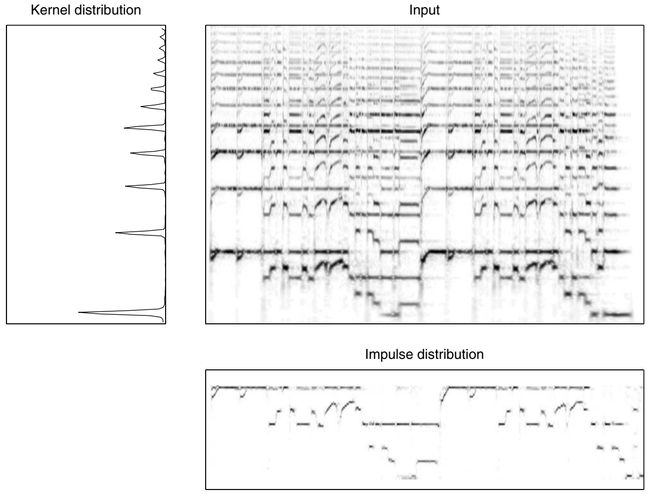


FIG. 8. Analysis of a country violin recording, which involves the continuous simultaneous sounding of two notes. The impulse distribution clearly represents the two notes at any point in time and provides an accurate description of the melodic content.

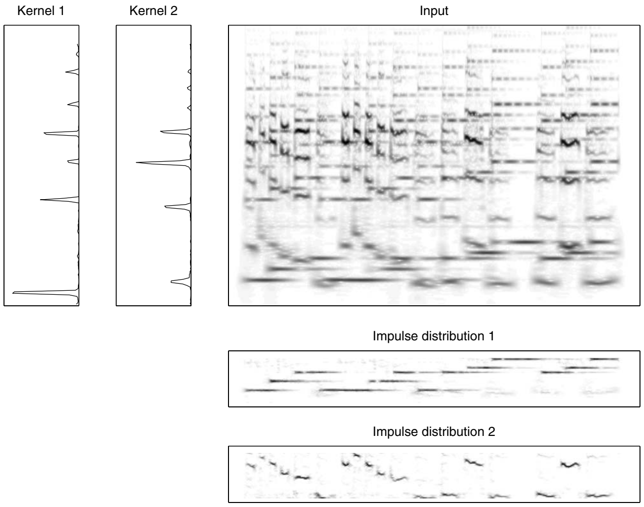


FIG. 9. Analysis of a mixture of two different sources performing simultaneous notes. The analysis results in two kernel and two impulse distributions, each pair describing one of the two sources. The top right plot displays the input. The two left plots show the two extracted kernel distributions, and the two bottom right plots show the impulse distributions that contain the recovered spectral pitch tracks of the two sources. Note that the second and fifth to last notes were performed an octave higher, and the displayed results do not exhibit an octave error.

The approach of using multiple spectral templates can also be very beneficial when attempting to recover the frequency shifts of a source, which is contaminated by additive noise. In this situation we can expect one kernel distribution to latch on to the spectrum of the source we wish to track and another one latching on to the background noise source. The impulse distribution corresponding to the tracked source will again be the spectral pitch track whereas the other impulse distribution will converge to some less structured form that is adequate to describe the presence of noise but will not carry any information about pitch.

If we are provided a way to invert the constant-Q transform, or a similar transform that is also invertible, we can even use this information to selectively reconstruct the input and thus isolate the melody of each source. However, recovering a time waveform from such a decomposition is not straightforward process and we postpone its discussion in future publications.

VI. DISCUSSION

The model we presented is able to overcome some of the challenges we set forth in the introduction of this paper. Although it might seem cumbersome at first, the probabilistic interpretation we have chosen provides a very flexible framework, which is easy to extend in multiple ways. In the presence of musical signals one can impose a prior on the structure of the impulse distribution so that it follows the frequency-shift expectations of the musical style at hand. Or if one is interested in temporally smoother spectral pitch tracks, modeling the temporal behavior of a specific kind of source, the application of a dynamical system such as a Kalman filter or a Markov model can incorporate prior temporal knowledge in order to provide more appropriate results.[15] Likewise rules on harmony and counterpoint can enhance this approach to allow polyphonic transcription.

This model is also useful when one knows the spectral characteristics of the sounds that need to be tracked. These characteristics can be applied as a prior on the kernel distribution.[15] Or in the case where these are exactly known, the kernel distributions can be preset and fixed as we only update the impulse distribution. This is essentially a straightforward non-negative deconvolution process. The only complication is that we need to maintain that the output has to be positive and a probability distribution. This results in a more powerful and interpretable representation compared to a cross-correlation output that a straightforward deconvolution would produce.

In conclusion, we presented a frequency-shift tracking model, which is flexible enough to deal with situations that can be challenging. This creates a robust front-end for performing pitch tracking, which makes soft decisions that can be highly desirable in complex music transcription systems. We presented results, which demonstrate the ability of this model to deal with mixtures, inharmonic sounds, and complex tracking situations. We also presented this model in a probabilistic framework, which allows clean statistical reasoning and makes it a good candidate for extensions that incorporate statistical priors depending on the input signal.

[1] B. Kedem, "Spectral analysis and discrimination by zero-crossings," Proc. IEEE **74**, 1477–1493 (1986).

[2] D. Terez, "Fundamental frequency estimation using signal embedding in state space," J. Acoust. Soc. Am. **112**, 2279 (2002).

[3] J. A. Moorer, "On the transcription of musical sound by computer," Comput. Music J. **1**, 32–38 (1977).

[4] J. C. Brown, "Musical fundamental frequency tracking using a pattern recognition method," J. Acoust. Soc. Am. **92**, 1394–1402 (1992).

[5] A. Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930 (2002).

[6] B. Doval and X. Rodet, "Estimation of fundamental frequency of musical sound signals," in International Conference on Acoustics, Speech and Signal Processing (1991), pp. 3657–3660.

[7] M. Goto, "A robust predominant-F0 estimation method for real-time detection of melody and bass lines in CD recordings," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey (2000).

[8] A. P. Klapuri, "Wide-band pitch estimation for natural sound sources with inharmonicities," in Proceedings of the 106th Audio Engineering Society Convention, Munich, Germany (1999).

[9] J. C. Brown, "Calculation of a constant Q spectral transform," J. Acoust. Soc. Am. **89**, 425–434 (1991).

[10] P. Smaragdis, B. Raj, and M. V. Shashanka, "Sparse and shift-invariant feature extraction from non-negative data," in Proceedings of the IEEE International Conference on Acoustics and Speech Signal Processing, Las Vegas, NV (2008).

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. Ser. B (Methodol.) **39**, 1–38 (1977).

[12] M. E. Brand, "Structure learning in conditional probability models via an entropic prior and parameter extinction," Neural Comput. **11**, 115–1182 (1999).

[13] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," Adv. Comput. Math. **5**, 329–359 (1996).

[14] M. V. Shashanka, B. Raj, and P. Smaragdis, "Advances in neural information processing systems 20," *Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, Dec. 3–6, 2007 (MIT Press, Cambridge).

[15] G. Mysore and P. Smaragdis, "Relative pitch estimation of multiple instruments," in Proceedings of the IEEE International Conference on Acoustics and Speech Signal Processing, Taipei, Taiwan (2009).

Velocity dispersion of guided waves propagating in a free gradient elastic plate: Application to cortical bone

Maria G. Vavva
Department of Materials Science and Engineering, Department of Computer Science, and Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110 Ioannina, Greece

Vasilios C. Protopappas
Department of Computer Science and Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110 Ioannina, Greece

Leonidas N. Gergidis and Antonios Charalambopoulos
Department of Materials Science and Engineering, University of Ioannina, GR 45110 Ioannina, Greece

Dimitrios I. Fotiadis[a)]
Department of Computer Science and Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110 Ioannina, Greece

Demosthenes Polyzos
Department of Mechanical Engineering and Aeronautics, University of Patras, GR 26500 Patras, Greece

(Received 6 October 2008; revised 9 March 2009; accepted 10 March 2009)

The classical linear theory of elasticity has been largely used for the ultrasonic characterization of bone. However, linear elasticity cannot adequately describe the mechanical behavior of materials with microstructure in which the stress state has to be defined in a non-local manner. In this study, the simplest form of gradient theory (Mindlin Form-II) is used to theoretically determine the velocity dispersion curves of guided modes propagating in isotropic bone-mimicking plates. Two additional terms are included in the constitutive equations representing the characteristic length in bone: (a) the gradient coefficient g, introduced in the strain energy, and (b) the micro-inertia term h, in the kinetic energy. The plate was assumed free of stresses and of double stresses. Two cases were studied for the characteristic length: $h=10^{-4}$ m and $h=10^{-5}$ m. For each case, three subcases for g were assumed, namely, $g>h$, $g<h$, and $g=h$. The values of g and h were of the order of the osteons size. The velocity dispersion curves of guided waves were numerically obtained and compared with the Lamb modes. The results indicate that when g was not equal to h (i.e., $g\neq h$), microstructure affects mode dispersion by inducing both material and geometrical dispersion. In conclusion, gradient elasticity can provide supplementary information to better understand guided waves in bones. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3110203]

PACS number(s): 43.80.Ev, 43.80.Jz, 43.20.Bi [CCC] Pages: 3414–3427

I. INTRODUCTION

In the quite rich literature of the quantitative ultrasound assessment of human bone status, some interesting techniques, such as the axial-transmission, have been proved effective in providing ultrasonic parameters that are related to long bone's mechanical and geometrical properties. Axial transmission methods and devices have determined the ultrasound propagation velocity by measuring the transit time of the first arriving signal (FAS) aiming at the assessment of osteoporosis (Bossy *et al.*, 2002, 2004; Camus *et al.*, 2000; Nicholson *et al.*, 2002; Njeh *et al.* 1999; Tatarinov *et al.*, 2005; Wear, 2007) and the evaluation of bone fracture healing (Dodd *et al.*, 2007; Protopappas *et al.*, 2005, 2008). More recently, axial transmission approaches have also been proposed to investigate the use of surface or guided waves for osteoporosis (Bossy *et al.*, 2004; Moilanen, 2008; Nicholson *et al.*, 2002; Tatarinov *et al.*, 2005) and bone healing evaluation (Protopappas *et al.*, 2006, 2007; Vavva *et al.*, 2008).

Guided waves are particularly attractive for characterization of bone status because they propagate throughout the cortical thickness and are thus sensitive to both mechanical and geometrical properties. Since they interact continuously with the boundaries of the bone, they propagate in different modes with velocities, which depend on the frequency. Although this multimodal and dispersive nature of guided waves makes their handling, control, and measurement much more difficult and complicated than bulk waves, guided ultrasound can provide various useful non-destructive testing parameters (Chimenti, 1997). Thus, it is apparent that understanding of how ultrasonic guided waves propagate through a bone is of paramount importance for the qualitative and quantitative inspection process.

Modeling of guided ultrasound in long human bones is a difficult task due to the complex microstructure of bones. In all the works dealing with guided waves in bones and appearing so far in the literature, the bone is mimicked as a

[a)] Author to whom correspondence should be addressed. Electronic mail: fotiadis@cs.uoi.gr linear elastic and homogenized medium. Bone is a strongly heterogeneous natural composite with a complex structure. Its architecture is organized at different hierarchical levels (Rho, 1998) which correspond to different structural elements, i.e., the macrostructure (cortical bone), the microstructure (harvesian systems, osteons, and single trabeculae), the sub-microstructure (lamellae), the nanostructure (fibrillar collagen and embedded mineral), and the sub-nanostructure (mineral, collagen, and non-collagenous organic proteins). At the microstructural level (i.e., 10–500 μm), bone consists of the osteons or harvesian systems with cylindrically-resembling shape of 200–250 μm diameter. Each harvesian system is formed from arrays (lamellae) of mineralized collagen fibers wrapped in concentric layers around a central canal (Rho, 1998). Previous *in vitro* measurements on parallelepipedic cortical bone samples, classified into different microstructures, have made clear that the bone's microstructure has a significant impact on the characteristics of ultrasonic propagation (Sasso *et al.*, 2008; Yamato *et al.*, 2006).

Experimental observations have also indicated that in heterogeneous materials with dimensions comparable to the length scale of the microstructure, microstructural effects become important and the state of stress has to be theoretically defined in a non-local manner. Since the classical theory of linear elasticity is associated with concepts of homogeneity and locality of stresses, it becomes obvious that it cannot provide a complete description of bone's dynamic mechanical behavior. Bone's microstructural effects can be successfully modeled in a macroscopic framework by employing enhanced elastic theories such as the couple stresses theory proposed by Cosserat brothers (Cosserat and Cosserat, 1909) and generalized later by Eringen as the micropolar elastic theory (Eringen, 1966), the general higher-order gradient elastic theory proposed by Mindlin (Mindlin, 1964), and the non-local theory of elasticity of Eringen (Eringen, 2002). For a literature review on the subject of these theories, one can consult the review articles (Tiersten and Bleustein, 1974; Exadaktylos and Vardoulakis, 2001), the literature review in the recent paper (Tsepoura *et al.*, 2002), the book (Vardoulakis and Sulem, 1995), as well as the paper of (Chakraborty, 2008) where a non-local poroelastic theory has been used to predict the reduction in the velocity with frequency in cancellous bone, the so-called negative dispersion which has been experimentally observed in various types of bone.

According to couple stresses theories, as proposed by Cosserat and Cosserat (1909) and Eringen (1966), the deformation of the medium is described not only by the displacement vector but also by an independent rotation vector. Displacements and rotations are associated with stresses and couple stresses through constitutive relations, which contrary to the classical theory of elasticity define non-symmetric stress and couple stress tensors. Both Cosserat and micropolar elastic theories have been successfully exploited to explain microstructural size effects in bones (Fatemi *et al.*, 2002; Hsia *et al.*, 2006; Lakes, 1981; Park and Lakes, 1986; Yang and Lakes, 1981, 1982; Yoon and Katz, 1983). Due to the rotations, couple stress theories are able to capture wave dispersion phenomena, which are not observed in the classical theory of elasticity. In the context of wave propagation in couple stress continuum, many papers have appeared in the past 20 years in the literature. Some representative papers are those of Chen *et al.* (2003) and Suiker *et al.* (2001) dealing with wave dispersion in free spaces, Tomar and Gogna (1995) solving wave reflection problems in flat interfaces, Kumar and Partap (2006) and Ottosen *et al.* (2000), investigating Rayleigh waves in micropolar half spaces, and those of Kulesh *et al.* (2007) and Midya (2004), treating propagation of dispersive waves in waveguides. However, no previous work has been reported to investigate guided wave propagation in bones in the context of higher-order gradient theories of elasticity.

Higher order gradient theories can be considered as generalizations of linear theory of elasticity, utilizing the displacement vector to describe the deformation of the continuum and introducing in both potential and kinetic energy higher-order terms associated with internal length scale parameters, which correlate microstructural effects with the macrostructural behavior of the considered material. In the regime of isotropic linear elastic behavior, the most general and comprehensive gradient elastic theory is the one due to Mindlin (1964, 1965). However, in order to introduce higher-order gradients of strains in the expression of potential energy, as well as to correlate the micro-strains with macro-strains, Mindlin (1964, 1965) utilized 18 new constants rendering; thus, his initial general theory is very complicated from physical and mathematical points of view. In the sequel considering long wavelengths, in order to have an elastic continuum with potential energy-density dependent on strain and strain gradient, kinetic energy-density dependent on velocity and velocity gradient and assuming the same deformation for macro- and micro-structure, Mindlin (1964, 1965) proposed three new simplified versions of his theory, known as Form-I, Form-II, and Form-III, where beyond the 2 Lamé constants 5 other ones are introduced instead of 16 employed in his initial model. In Form-I, the strain energy-density function is assumed to be a quadratic form of the classical strains and the second gradient of displacement; in Form-II, the second displacement gradient is replaced by the gradient of strains; in Form-III, the strain energy function is written in terms of the strain, the gradient of rotation, and the fully symmetric part of the gradient of strain. The most important difference among the aforementioned three simplified versions of general Mindlin's theory is the fact that Form-II leads to a total stress tensor, which is symmetric as in the case of classical elasticity. This symmetry avoids the problems introduced by the non-symmetric stress tensors in Cosserat and Cosserat (1909) and couple stress theories.

Ru and Aifantis (1993) and Altan *et al.* (1996) proposed very simple static and dynamic, respectively, versions of Mindlin's theory requiring only one new gradient elastic constant plus the standard Lamé ones. Although elegant, the problem with Aifantis and co-workers models is first the use of boundary conditions which are not compatible with the corresponding correct ones provided by Mindlin (1964, 1965) and second the fact that they ignore of the contribution of the inertia of the microstructure to the dynamic behavior of the gradient elastic material. Both drawbacks were corrected later by Georgiadis *et al.* (2004) and Vardoulakis and

Georgiadis (1997), proposing two very simple and elegant gradient elastodynamic theories called gradient elasticity with surface energy and dipolar gradient elastic theory, respectively. The latter, exploited in the present work, can be considered as the simplest possible special case of Mindlin's Form-II gradient elastic theory. In the framework of wave propagation, in infinite and semi-infinite gradient elastic spaces, one can mention the works of Aggelis *et al.* (2004), Bennet *et al.* (2007), Chang *et al.* (1998), Erofeyev (2003), Georgiadis and Velgaki (2003), Georgiadis *et al.* (2004), Papargyri-Beskou and Beskos (2008), Sluys *et al.* (1993), Vardoulakis and Georgiadis (1997), and Yerofeyev and Sheshenina (2005).

Finally, in the non-local theory of elasticity of Eringen (2002), stresses at any point of the considered continuum are assumed to be a function not only of the strains defined at the point itself but also of the strain states defined at all other points of the elastic body. This consideration leads to an integro-differential stress-strain constitutive relation, which contains integrals defined over the entire region of interest and kernels comprising weighted averages of the contributions of the strains of all points of the elastic body. The integro-differential form of the constitutive equations renders non-local elastic theory complex for practical applications. Some special cases where the integro-differential constitutive equations can be converted to differential ones are convenient or similar to the aforementioned micropolar and higher-order gradient elastic theories (Artan and Altan, 2002; Chakraborty, 2007).

In the present work, the simplified Mindlin (1964, 1965) Form-II or dipolar gradient elastic theory is exploited for the determination of symmetric and antisymmetric modes propagating in a two-dimensional (2D) and free of stresses gradient elastic plate. The material properties of the plate as well as the internal length scale parameters introduced by the considered enhanced elastic theory are compatible with the physiological properties of human bones. To the authors' best knowledge, no such theory has been proposed up to now for the simulation of propagating guided waves in long bones. The main advantages of the utilized gradient elastic theory as compared to other couple stresses, micropolar, gradient elastic, and non-local elastic theories, are its simplicity and the symmetry of all classical and non-classical stress tensors involved. This paper is organized as follows: Sec. II is devoted to the basics of the aforementioned gradient elastic theory. In the same section, the dispersive nature of the waves propagating in an infinitely extended gradient elastic continuum is illustrated. Next, in Sec. III, the determination of the internal length scale parameters employed in the present study is explained, while in Sec. IV the modes corresponding to guided waves traveling in a free gradient elastic plate are explicitly derived. The obtained results are discussed in detail in Sec. V. Finally, the main conclusions of the present study are drawn.

II. MINDLIN'S FORM II SIMPLIFIED GRADIENT ELASTIC THEORY AND WAVE DISPERSION

Mindlin (1964) in the Form-II version of his gradient elastic theory considered that the potential energy density $\hat{W}$ is a quadratic form of the strains ε_{ij} and the gradient of strains, $\hat{\kappa}_{ijk}$, i.e.,

$$\hat{W} = \tfrac{1}{2}\lambda\varepsilon_{ii}\varepsilon_{jj} + \mu\varepsilon_{ij}\varepsilon_{ij} + \hat{\alpha}_1\hat{\kappa}_{iik}\hat{\kappa}_{kjj} + \hat{\alpha}_2\hat{\kappa}_{ijj}\hat{\kappa}_{ikk} + \hat{\alpha}_3\hat{\kappa}_{iik}\hat{\kappa}_{jjk} + \hat{\alpha}_4\hat{\kappa}_{ijk}\hat{\kappa}_{ijk} + \hat{\alpha}_5\hat{\kappa}_{ijk}\hat{\kappa}_{kji}, \tag{1}$$

where

$$\varepsilon_{ij} = \tfrac{1}{2}(\partial_i u_j + \partial_j u_i), \quad \hat{\kappa}_{ijk} = \partial_i\varepsilon_{jk} = \tfrac{1}{2}(\partial_i\partial_j u_k + \partial_i\partial_k u_j) = \hat{\kappa}_{ikj}, \tag{2}$$

with ∂_i denoting space differentiation, u_i the displacement, and λ and μ representing the Lamé constants. The constants $\hat{\alpha}_1, \ldots, \hat{\alpha}_5$ have units of m^2 explicitly defined in Mindlin, 1964.

Extending the idea of non-locality to the inertia of the continuum with microstructure, Mindlin (1964) proposed for the isotropic case an enhanced expression for the kinetic energy-density function $\hat{T}$, which beyond velocities includes the gradients of the velocities, i.e.,

$$\hat{T} = \tfrac{1}{2}\rho\dot{u}_i\dot{u}_i + \tfrac{1}{6}\rho\tilde{d}^2_{pkmn}\partial_m\dot{u}_n\partial_p\dot{u}_k, \tag{3}$$

where ρ is the mass density, overdots indicate differentiation with respect to time t, and $\tilde{d}^2_{pkmn}$ is a fourth order tensor explained in Mindlin, 1964.

Strains ε_{ij} and gradient of strains $\hat{\kappa}_{ijk}$ are dual in energy with the Cauchy and double stresses, respectively, defined as

$$\hat{\tau}_{ij} = \frac{\partial\hat{W}}{\partial\varepsilon_{ij}} = \hat{\tau}_{ji}, \tag{4}$$

$$\hat{\mu}_{ijk} = \frac{\partial\hat{W}}{\partial\hat{\kappa}_{ijk}} = \hat{\mu}_{ikj}, \tag{5}$$

where Einstein's summation convention is employed. Equations (4) and (5) imply that

$$\hat{\tau}_{pq} = 2\mu\varepsilon_{pq} + \lambda\varepsilon_{ii}\delta_{pq}, \tag{6}$$

and

$$\hat{\mu}_{pqr} = \tfrac{1}{2}\hat{\alpha}_1[\hat{\kappa}_{rii}\delta_{pq} + 2\hat{\kappa}_{iip}\delta_{qr} + \hat{\kappa}_{qii}\delta_{rp}] + 2\hat{\alpha}_2\hat{\kappa}_{pii}\delta_{qr} + \hat{\alpha}_3(\hat{\kappa}_{iir}\delta_{pq} + \hat{\kappa}_{iiq}\delta_{pr}) + + 2\hat{\alpha}_4\hat{\kappa}_{pqr} + \hat{\alpha}_5(\hat{\kappa}_{rpq} + \hat{\kappa}_{qrp}). \tag{7}$$

The total stress tensor $\hat{\sigma}_{pq}$ is then defined as

$$\hat{\sigma}_{pq} = \hat{\tau}_{pq} - \partial_r\hat{\mu}_{rpq}, \tag{8}$$

which is symmetric, since both Cauchy stresses $\hat{\tau}_{pq}$ and relative stresses $\partial_r\hat{\mu}_{rpq}$ are symmetric according to Eqs. (4) and (5).

Considering an isotropic continuum with microstructural effects confined by a smooth boundary and taking the variation in strain and kinetic energy, according to Hamilton's principle, the following equation of motion is obtained for a continuum with microstructure (Mindlin, 1964):

$$\partial_j(\hat{\tau}_{jk} - \partial_i \hat{\mu}_{ijk}) + F_k = \rho \ddot{u}_k - \tfrac{1}{3}\partial_p(\rho d^2_{pkmn}\partial_m \ddot{u}_n), \tag{9}$$

accompanied by the classical and non-classical boundary conditions, respectively:

$$\hat{p}_k = p_k^{\text{prescribed}}, \tag{10}$$

$$\hat{R}_k = R_k^{\text{prescribed}}. \tag{11}$$

The traction vector $\hat{p}_k$ and the double traction vector $\hat{R}_k$ are defined as

$$\hat{p}_k = n_j \tau_{jk} - n_i n_j D \hat{\mu}_{ijk} - (n_j D_i + n_i D_j)\hat{\mu}_{ijk} + (n_i n_j D_l n_l - D_j n_i)\hat{\mu}_{ijk} + \tfrac{1}{3}\rho n_p \widetilde{d}^2_{pkmn}(D_m \ddot{u}_n + n_m D \ddot{u}_n), \tag{12}$$

$$\hat{R}_k = n_i n_j \hat{\mu}_{ijk}, \tag{13}$$

where $D_i=(\delta_{ij}-n_i n_j)\partial_j$ is the surface gradient operator, δ_{ij} is the Kronecker delta, $D=n_l\partial_l$ is the normal gradient operator, and n_l is the outward unit normal vector to the boundary (Mindlin, 1964).

In terms of displacements, the equation of motion (9) becomes

$$(\lambda + 2\mu)(1 - g_1^2\nabla^2)\nabla\nabla\cdot\mathbf{u} - \mu(1 - g_2^2\nabla^2)\nabla\times\nabla\times\mathbf{u} = \rho(\ddot{\mathbf{u}} - h_1^2\nabla\nabla\cdot\ddot{\mathbf{u}} + h_2^2\nabla\times\nabla\times\ddot{\mathbf{u}}), \tag{14}$$

where $\mathbf{u}$ stands for displacement vector and

$$g_1^2 = 2(\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 + \hat{\alpha}_4 + \hat{\alpha}_5)/(\lambda + 2\mu),$$

$$g_2^2 = (\hat{\alpha}_3 + 2\hat{\alpha}_4 + \hat{\alpha}_5)/2\mu,$$

$$h_1^2 = d^2[2\alpha^2 + (\alpha + \beta)^2]/3,$$

$$h_2^2 = d^2(1 + \beta^2)/6, \tag{15}$$

with g_1, h_1 being the internal length scale parameters which affect longitudinal waves, g_2, h_2 the corresponding ones affecting shear waves, and d, a, β being constants illustrated in (Mindlin, 1964).

Positive definiteness of $\hat{W}$ (for reasons of uniqueness and stability) requires that (Mindlin, 1964) $\mu>0$, $\lambda+2\mu>0$, $g_i^2>0$, and $h_i^2>0$.

In the simplest possible case of the gradient theory of elasticity (Mindlin, 1964) where the potential energy density $\hat{W}$ and the kinetic energy $\hat{T}$ are defined as

$$\hat{W} = \varepsilon_{ij}\tau_{ij} + g^2\partial_i\varepsilon_{jk}\partial_i\tau_{jk}, \quad \hat{T} = \tfrac{1}{2}\rho\dot{u}_i\dot{u}_i + \tfrac{1}{6}\rho h^2(\partial_i\dot{u}_i)(\partial_i\dot{u}_i), \tag{16}$$

the constants $\hat{\alpha}_1, \ldots, \hat{\alpha}_5$, d, α, and β become $\hat{\alpha}_1=\hat{\alpha}_3=\hat{\alpha}_5=0$, $\hat{\alpha}_2=\lambda g^2/2$, $\hat{\alpha}_4=\mu g^2$, $d^2/6=h^2$, $\alpha=1$, and $\beta=0$ Then, the constants $g_1^2, g_2^2, h_1^2, h_2^2$ in Eq. (14) obtain the form $g_1^2\equiv g_2^2 = g^2$ and $h_1^2\equiv h_2^2=h^2$. Under the above simplifications, the stresses in Eqs. (6)–(8) become

$$\tau_{ij} = 2\mu\varepsilon_{ij} + \lambda\varepsilon_{ii}\delta_{ij}, \tag{17}$$

$$\mu_{ijk} = g^2\partial_i\tau_{jk}, \tag{18}$$

$$\sigma_{ij} = \tau_{ij} - g^2\nabla^2\tau_{ij}, \tag{19}$$

and the equation of motion, i.e., Eq. (14), through the well known identity $\nabla^2=\nabla\nabla\cdot-\nabla\times\nabla\times$ obtains the simple form

$$(1 - g^2\nabla^2)[\mu\nabla^2\mathbf{u} + (\lambda + \mu)\nabla\nabla\cdot\mathbf{u}] = \rho(\ddot{\mathbf{u}} - h^2\nabla^2\ddot{\mathbf{u}}), \tag{20}$$

where $g^2\nabla^2[\mu\nabla^2\mathbf{u}+(\lambda+\mu)\nabla\nabla\cdot\mathbf{u}]$ and $\rho h^2\nabla^2\ddot{\mathbf{u}}$ are the microstructural and the micro-inertia terms, respectively, and the operator ∇^2 is the Laplacian.

Taking the divergence and the curl of Eq. (20), the equations governing the propagation of dilatations and rotations are derived, i.e.,

$$(\lambda + 2\mu)(1 - g^2\nabla^2)\nabla^2\nabla\cdot\mathbf{u} = \rho(1 - h^2\nabla^2)\nabla\cdot\ddot{\mathbf{u}}, \tag{21}$$

$$\mu(1 - g^2\nabla^2)\nabla^2\nabla\times\mathbf{u} = \rho(1 - h^2\nabla^2)\nabla\times\ddot{\mathbf{u}}. \tag{22}$$

Considering plane waves of the form

$$\nabla\cdot\mathbf{u} = Ae^{i(K\hat{k}\cdot r - \omega t)},$$

$$\nabla\times\mathbf{u} = \mathbf{B}e^{i(K\hat{k}\cdot r - \omega t)}, \tag{23}$$

where $A, \mathbf{B}$ represent amplitudes, $\mathbf{r}$ stands for the position vector, $\hat{\mathbf{k}}$ is the direction of incidence, K and ω are the wave number and the frequency of the propagating waves, respectively, and $i=\sqrt{-1}$.

Inserting Eq. (23) into Eqs. (21) and (22) and representing by C_L, C_T the classical phase velocities of longitudinal (L) and shear (T) waves, respectively, we obtain the following dispersion relation:

$$\omega^2 = C_L^2\frac{K_L^2(1 + g^2K_L^2)}{1 + h^2K_L^2}, \quad C_L^2 = \frac{\lambda + 2\mu}{\rho}, \tag{24}$$

for longitudinal waves and the relation

$$\omega^2 = C_T^2\frac{K_T^2(1 + g^2K_T^2)}{1 + h^2K_T^2}, \quad C_T^2 = \frac{\mu}{\rho}, \tag{25}$$

for shear waves.

Using Eqs. (24) and (25), the following expressions for the phase velocities V_L and V_T of the longitudinal and shear waves, respectively, are obtained:

$$V_{L,T} = \frac{\omega}{K_{L,T}} = C_{L,T}\sqrt{\frac{1 + g^2K_{L,T}^2}{1 + h^2K_{L,T}^2}}. \tag{26}$$

Equations (25) and (26) reveal that, unlike the classical elastic case characterized by constant velocities of longitudinal and shear waves and hence non-dispersive wave propagation, the gradient elastic case is characterized by phase velocities for longitudinal and shear waves, which are functions of the wave number, indicating wave dispersion. This dispersion is entirely due to the presence of the two microstructural material constants g^2 and h^2. When g and h are set equal to zero in Eq. (26) it becomes obvious that $V_{L,T}=C_{L,T}$, i.e., the classical elastic case with constant wave speeds and hence no dispersion.

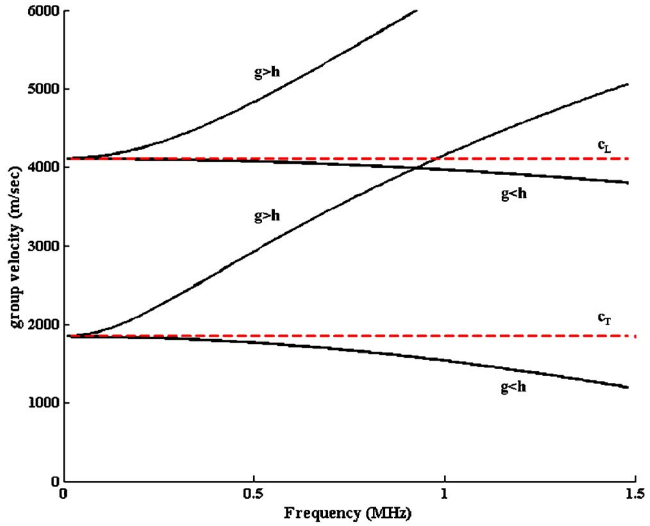


FIG. 1. (Color online) Dispersion curves of the group velocity versus frequency for elastic medium with microstructure.

Solution of Eqs. (24) and (25) regarding the wave numbers K_L and K_T, respectively, leads to the following relation:

$$K_{L,T}=\sqrt{\frac{-(C_{L,T}^2-\omega^2h^2)+\sqrt{(C_{L,T}^2-\omega^2h^2)^2+4\cdot C_{L,T}^2g^2\omega^2}}{2C_{L,T}^2g^2}}. \tag{27}$$

Since in a dispersive medium energy propagates with the group velocity $V_{L,T}^g$ instead of phase velocity $V_{L,T}$ (Rose, 1999), we can find from Eq. (27) that

$$V_{L,T}^g=\frac{d\omega}{dK_{L,T}}=\sqrt{\frac{K_{L,T}^2h^2+1}{K_{L,T}^4C_{L,T}^2g^2+K_{L,T}^2C_{L,T}^2}} \times\frac{2K_{L,T}C_{L,T}^2(K_{L,T}^4g^2h^2+2K_{L,T}^2g^2+1)}{(K_{L,T}^2h^2+1)^2}. \tag{28}$$

Figure 1 provides the dispersion curves governing the group velocity of longitudinal (L) and shear (T) waves propagating in an infinitely extended gradient elastic medium for various combinations of g and h as a function of frequency according to Eq. (28).

III. DETERMINATION OF THE INTERNAL LENGTH SCALE PARAMETERS

One of the most crucial and difficult issues in the enhanced linear elastic theories as they apply to real problems is the determination of the internal length scale parameters g_1^2, g_2^2, g^2 and h_1^2, h_2^2, h^2 appearing in Eqs. (14) and (20). This section tries to shed some light on the subject based on some observations as well as on theoretical and experimental results found in the literature.

The gradient coefficients g_1, g_2, g and the inertia coefficients h_1, h_2, h are internal length scale parameters, which indicate how the elastic properties and the inertia of the microstructure affect the macrostructural behavior of materials and structures. All have units of length. Equation (14) indicates that the internal length scale parameters for longitudinal waves, i.e., g_1, h_1, are different from the corresponding ones, g_2, h_2 for shear waves. The simplification made in Eqs. (16) and (20) is that both longitudinal and transverse waves are affected by the same gradient and inertia internal lengths g and h, respectively. Let us see some examples which reveal the nature of these parameters. Ben-Amoz (1976) considering a particulate composite material found that

$$g_1^2=\frac{l^2}{4}\left[1-\frac{\mu_f-\mu_m}{\mu_v}(v_f-4I_f)\right], \tag{29}$$

$$g_2^2=\frac{l^2}{4}\left[1-\frac{(\lambda_f+2\mu_f)-(\lambda_m+2\mu_m)}{\lambda_v+2\mu_v}(v_f-4I_f)\right], \tag{30}$$

$$h_1^2=\frac{l^2}{8}[1+(\rho_R/\rho_m-\rho_R/\rho_f)(v_f-4I_f)]^{-1}, \tag{31}$$

$$h_2^2=\frac{l^2}{4}[1+(\rho_R/\rho_m-\rho_R/\rho_f)(v_f-4I_f)]^{-1}, \tag{32}$$

where l is the size of the representative volume element of the composite consisting of a particle embedded in a matrix with Lamé constants λ_f, μ_f and λ_m, μ_m, respectively. All the other parameters involved in Eqs. (29)–(32) are explicitly given in Ben-Amoz, 1976. This is an example where all the internal length scale parameters depend on the size as well as on the elastic properties and the inertia of the constituents of the representative volume element of the material with microstructural effects. To the authors' best knowledge, the study of Ben-Amoz (1976) was the first to provide closed form relations for the coefficients g_1, g_2, h_1, h_2.

Tekoglu and Onck (2005), working in cellular materials with regular and irregular hexagonal microstructures and comparing the results obtained from a discrete and a micropolar model, found that the micropolar characteristic length l_m obtains values in the range $0.15d$–$0.28d$ for the regular and $0.47d$–$0.51d$ for the irregular hexagonal microstructure, with d representing the size of the unit cell. Later, Tekoglu (2007), comparing the solutions of the same problem corresponding to micropolar and gradient elastic theories, found that the micropolar and the gradient internal length scale parameters l_m and g, respectively, are correlated with each other via the relation $g=0.707l_m$. Here, the gradient parameter g depends on both the size of the unit cell and the way that these cells form the entire cellular material.

For a dynamic case, Georgiadis *et al.* (2004), considering a material composed wholly of unit cells having the form of cubes with edges of size $2h$ and comparing the dispersion curves of Rayleigh waves obtained by the Toupin–Mindlin approach with those obtained by the atomic-lattice analysis of Gazis *et al.* (1960), found that for polycrystals, alloys, and granular materials g^2 is of the order of $(0.1h)^2$.

As in the case of material and geometrical wave dispersion, the coefficients g and h besides their dependence on the size and the properties of the microstructure can be correlated with some specific geometric characteristics of the structure. For example, Vardoulakis and Giannakopoulos (2006), working on the bending of a classic Bernoulli–Euler beam and taking into account the double shear forces appearing in the cross-section of a T-type beam, found that a Timoshenko beam is nothing more than a gradient elastic Bernoulli–Euler beam (Papargyri-Beskou *et al.*, 2003) with gradient coefficient $g=4(1+\nu)H$, where ν is the Poisson ratio and H is the distance of the horizontal plate of the T-beam from the center of the cross-section. Thus, in this case as the authors pointed out "The interesting point is that this length scale effect has a definite structural origin, other than a material one." Similarly, considering the propagation of one-dimensional axial waves in a beam with lateral inertia correction, known as Rayleigh model (Graff, 1975), it can be shown that the same phenomenon can be explicitly described by the one-dimensional analog of Eq. (20) with $g^2=0$ and $h^2=\nu\gamma$, where ν is Poisson's ratio and γ is the polar radius of gyration of the cross-section of the bar.

Furthermore, there is experimental evidence that g and h are influenced by the type of loading. For example, Lakes (1981) and Yang and Lakes (1982) found that the micropolar internal length scale parameters corresponding to torsion and bending of human compact bones are $l_t=2.2\times10^{-4}$ m and $l_b=4.5\times10^{-4}$ m, respectively, a fact that according to Tekoglu (2007) reflects different values of g in torsion and bending.

Finally, for a given gradient coefficient g, the inertia coefficient h can be determined through comparisons with experimental dispersion curves taken by different type of materials. More specifically, it is apparent from Eq. (26) that for $h=g$ there is no dispersion and $V^g_{L,T}\equiv C_{L,T}$. This occurs because this specific value of h separates two large categories of materials. For $h>g$ negative dispersion is obtained, i.e., $V^g_{L,T}$ decreases as frequency increases, and $V^g_{L,T}<C_{L,T}$. This is a physically acceptable case, which is in agreement with results of crystal lattice theories for the 2D space (Suiker *et al.*, 2001; Yim and Sohn, 2000) and the 2D half-space (Gazis *et al.*, 1960). The relation $h>g$ was first found to lead to results in agreement with lattice theories during the numerical studies of (Georgiadis *et al.*, 2004) for wave dispersion in the half-plane. This case is also in agreement with experimental results on metals and alloys (Erofeyev, 2003; Kondratev, 1990). For $h<g$, there is dispersion, $V^g_{L,T}>C_{L,T}$, and $V^g_{L,T}$ increases with increasing frequency, exhibiting thus positive dispersion in agreement with experimental results on granular type of materials, such as marble, sand, concrete, granular composites, bones, and cellular materials (Aggelis *et al.*, 2004; Chen and Lakes, 1989; Erofeyev, 2003; Lakes, 1982; Stavropoulou *et al.*, 2003).

Despite the intensive work in the literature, there are only few experimental works dealing with the determination of internal parameters in solids with microstructural effects. The majority of them are referred to the determination of the micropolar internal length l_m than to the gradient one g. Regardless to which parameter they measure, their main conclusion is that the internal lengths such as g and h in metals, alloys, composites, cellular and granular materials, foams, bones, ceramics, etc., are smaller than 10^{-3} m. In general, there is strong experimental evidence (Aifantis, 1999; Georgiadis *et al.*, 2004; Exadaktylos and Vardoulakis, 2001; Lakes, 1982; Lakes, 1995; Lam *et al.*, 2003; Yang and Lakes, 1982) that g should be of the same order as the size of the basic building block of microstructure, e.g., the osteons in bones; however, no final conclusion has been drawn in the literature.

TABLE I. Values of g and h for each one of the six subcases.

Cases	Gradient coefficient g (m)	Intrinsic characteristic length h (m)
Case-1a	5×10^{-4}	10^{-4}
Case-1b	10^{-5}	10^{-4}
Case-1c	10^{-4}	10^{-4}
Case-2a	10^{-4}	10^{-5}
Case-2b	5×10^{-6}	10^{-5}
Case-2c	10^{-5}	10^{-5}

In the present work, the values of g and h, presented in Table I, have been considered taking into account the experimental results of Lakes and co-workers for compact bones, the conclusion of Tekoglu (2007) that $g<l_m$ and the two different cases of dispersion ($h>g$ and $h<g$). The relation $h>g$ physically shows that the inertia of bone's microstructure plays the most important role in the dispersion phenomena, as opposed to the relation $h<g$ which shows that the phenomena are mostly due to the elastic behavior of the microstructure.

IV. WAVE PROPAGATION IN A GRADIENT ELASTIC FREE PLATE

Considering a free 2D plate and a Cartesian coordinate system Ox_1x_2 with the axis Ox_1 being the axis of symmetry of the plate. Assuming plane strain conditions, the components of the displacement vector can be written as

$$u_1=u_1(x_1,x_2,t),$$

$$u_2=u_2(x_1,x_2,t),$$

$$u_3=0. \tag{33}$$

Solution to the equation of motion, i.e., Eq. (20), is given using the method of potentials. The displacement vector field is decomposed according to decomposition as a gradient of a scalar and the curl of the zero divergence vector, i.e.,

$$\mathbf{u}=\nabla\varphi+\nabla\times\mathbf{A},\qquad \nabla\cdot\mathbf{A}=0. \tag{34}$$

Substituting Eq. (34) into the equation of motion derived from the gradient elastic theory, i.e., Eq. (20), the following two partial differential equations are obtained:

$$(1-g^2\nabla^2)\nabla^2\phi=\frac{(1-h^2\nabla^2)}{C_L^2}\ddot{\phi}, \tag{35a}$$

$$(1-g^2\nabla^2)\nabla^2\mathbf{A}=\frac{(1-h^2\nabla^2)}{C_T^2}\ddot{\mathbf{A}}. \tag{35b}$$

It is obvious that for the case $g=h$ the expressions $g^2\nabla^2$ and $h^2\nabla^2$ in Eqs. (35a) and (35b) are identical and can be eliminated, resulting thus in the corresponding expressions obtained in the classical elastic theory.

Writing the scalar and vector potentials as $\varphi = \phi(x_1,x_2,t)$ and $\mathbf{A}=(A_1,A_2,A_3)$: $A_1=A_2=0$, $A_3=\psi(x_1,x_2,t)$, the displacement vector in terms of potentials is expressed as

$$\mathbf{u} = \hat{x}_1\left(\frac{\partial\phi}{\partial x_1} + \frac{\partial\psi}{\partial x_2}\right) + \hat{x}_2\left(\frac{\partial\phi}{\partial x_2} + \frac{\partial\psi}{\partial x_1}\right), \qquad (36)$$

where $\hat{x}_1,\hat{x}_2$ are the unit vectors on the directions x_1,x_2, respectively.

Thus, Eqs. (35a) and (35b) are expressed as

$$(1 - g^2\nabla^2)\nabla^2\phi = \frac{1}{C_L^2}(1 - h^2\nabla^2)\ddot{\phi}, \qquad (37a)$$

$$(1 - g^2\nabla^2)\nabla^2\psi = \frac{1}{C_T^2}(1 - h^2\nabla^2)\ddot{\psi}. \qquad (37b)$$

Assuming traveling waves in the x_1 direction and standing waves in the x_2 direction of the forms

$$\phi = \Phi(x_2)\exp\{i(Kx_1 - \omega t)\}, \qquad (38)$$

$$\psi = \Psi(x_2)\exp\{i(Kx_1 - \omega t)\}, \qquad (39)$$

where Φ and Ψ are unknown functions. Substituting Eqs. (38) and (39) into Eqs. (37a) and (37b), respectively, the following differential equations are obtained:

$$-g^2\frac{\partial^4\Phi}{\partial x_2^4} + \left(1 + 2K^2g^2 - \frac{h^2\omega^2}{C_L^2}\right)\frac{\partial^2\Phi}{\partial x_2^2} + (K_L^2 - K^2 - g^2K^4 + h^2K^2K_L^2)\Phi = 0, \qquad (40a)$$

$$-g^2\frac{\partial^4\Psi}{\partial x_2^4} + \left(1 + 2K^2g^2 - \frac{h^2\omega^2}{C_T^2}\right)\frac{\partial^2\Psi}{\partial x_2^2} + (K_T^2 - K^2 - g^2K^4 + h^2K^2K_T^2)\Psi = 0. \qquad (40b)$$

Note here that K represents the wavenumber of the propagating guided disturbance and $K_L^2=\omega^2/C_L^2$ and $K_T^2=\omega^2/C_T^2$.

It can be observed that when the volumetric strain gradient coefficient g^2 and the inertia term h^2 becomes zero, Eqs. (40a) and (40b) become identical to those obtained in the classical elastic case (Rose, 1999). The solutions of Eqs. (40a) and (40b) can be written in the forms

$$\Phi(x_2) = Q\sin px_2 + R\cos px_2 + S\exp\{r_p x_2\} + T\exp\{-r_p x_2\}, \qquad (41)$$

$$\Psi(x_2) = U\sin qx_2 + V\cos qx_2 + W\exp\{r_s x_2\} + Z\exp\{-r_s x_2\}, \qquad (42)$$

where

$$p,q = i\frac{\sqrt{1 + 2K^2g^2 - K_{L,T}^2h^2 - \sqrt{(1 + 2K^2g^2 - K_{L,T}^2h^2)^2 + 4g^2(K_{L,T}^2 - K^2 - g^2K^4 + h^2K_{L,T}^2K^2)}}}{g\sqrt{2}}, \qquad (43)$$

$$r_{p,s} = \frac{\sqrt{1 + 2K^2g^2 - K_{L,T}^2h^2 + \sqrt{(1 + 2K^2g^2 - K_{L,T}^2h^2)^2 + 4g^2(K_{L,T}^2 - K^2 - g^2K^4 + h^2K_{L,T}^2K^2)}}}{g\sqrt{2}}, \qquad (44)$$

and the constants (Q,R,S,T,U,V,W,Z) are unknown amplitudes, which can be determined by satisfying the classical and non-classical boundary conditions of the problem, respectively:

$$\mathbf{P}|_{x_2=d/2} = \mathbf{P}|_{x_2=-d/2} = 0, \qquad (45a)$$

$$\mathbf{R}|_{x_2=d/2} = \mathbf{R}|_{x_2=-d/2} = 0, \qquad (45b)$$

where $\mathbf{P}$ and $\mathbf{R}$ are the traction vector and the double stress traction vector, respectively. Satisfaction of the boundary conditions result in two systems of four equations: the first for the unknowns R,U,S,Z corresponding to the symmetric modes and the second for unknowns Q,T,V,W corresponding to the antisymmetric modes. The components of the determinant of the two systems are given in Appendix. Vanishing of each determinant yields the characteristic dispersion equations for the propagation of the symmetric and antisymmetric modes in a gradient elastic plate.

V. APPLICATION TO CORTICAL BONE PLATES

In this section, a free 2D gradient elastic plate is considered to have mechanical properties typically used for bone, i.e., Young's modulus $E_{\text{bone}}=14$ GPa, Poisson's ratio $\upsilon_{\text{bone}}=0.37$ and density $\rho_{\text{bone}}=1500$ Kg/m^3 (Bossy *et al.*, 2004; Nicholson *et al.*, 2002; Protopappas *et al.*, 2005; Protopappas *et al.*, 2006). The plate thickness is 4 mm, which is a common cortical value found in several types of human long bones (Njeh *et al.*, 1999). The resulting classical bulk longitudinal and shear velocities are 4063 m/s and 1846 m/s, respectively (Bossy *et al.*, 2004; Protopappas *et al.*, 2006). Two different cases for the inertia characteristic length h are investigated. In the first case, denoted herein as Case-1, $h=10^{-4}$ m, whereas in the second, denoted as Case-2, $h=10^{-5}$ m. In both cases, the values of h are comparable to the size of bone's microstructure (Harvesian systems, osteons), i.e., from 10 to 500 μm (Rho, 1998).

As it is analyzed in Sec. III, the value of g should be of the same order as the size of the basic building block of microstructure, e.g., the osteons in bone. Therefore, in the present analysis for each of the previous two cases for h, we consider three subcases for g, resulting totally in six different combinations between g and h. In the first two subcases (Case-1a and Case-1b), g is assumed higher and smaller than h, respectively, whereas in the third subcase (Case-1c), the value of g is assumed to be equal to that of h. For Case-2, three subcases are constructed in a similar way as in Case-1. The values for h and g in each subcase are presented in Table I.

In what follows, the symmetric and antisymmetric modes propagating in the bone-mimicking plate with microstructure, in the form of frequency-group velocity (f,c_g) dispersion curves for the different combinations of g and h, are presented. The symmetric and antisymmetric modes obtained from the simplified version of Mindlin's Form-II gradient elastic theory are denoted herein as g_Sn and g_An, respectively, where $n=0,1,2\ldots$ represents the mode number.

The numerical solution of the problem is achieved using a symbolic algebra software (MATHEMATICA, Wolfram Research, Inc., 2004). Calculations were carried out in which the frequency step was progressively decreasing to investigate the dependence of the derived curves on the frequency step. The obtained dispersion curves are all calculated for a frequency step equal to 10 krad/s since no differences were observed for smaller steps.

Figures 2(a)–2(c) illustrate the group velocity dispersion curves of the symmetric modes for Cases-1a, 1b, and 1c, respectively. The group velocity dispersion curves of the Lamb modes for a classical elastic plate with the same geometrical and mechanical properties are also presented in each figure (dashed lines) for comparison purposes. The Lamb modes are denoted as Sn and An, where $n=0,1,2\ldots$ In Figs. 2(a) and 2(b), it can be observed that the bulk shear wave (denoted as g_c_T.) is dispersive, a result that is in fully agreement with the curves shown in Fig. 1. More specifically, in Fig. 2(a) which represents Case-1a $(g>h)$, the velocity of the bulk shear wave derived from the gradient theory for zero frequency is equal to the bulk shear velocity of the medium in classical elasticity (depicted in the figure by a straight dashed line extending across the whole spectrum, denoted as c_T). As frequency increases, the dispersion curve of the bulk shear wave predicted by the gradient theory significantly deviates from the bulk shear velocity value in the classical case taking higher values. For Case-1b $(g<h)$, g_c_T takes lower values than c_T, as opposed to Case-1a, and its deviation starts after 0.37 MHz. However in this case the deviation from c_T is less pronounced than in Case-1a. Finally, in Case-1c $(g=h)$, the velocity of the bulk shear wave in the gradient elastic case exhibits no dispersion, as expected, i.e., its value remains constant with increasing frequency exactly the same as the bulk shear velocity of bone.

Concerning the guided waves in Fig. 2(a), the dispersion of the modes predicted by the gradient theory is strongly modified from that of the classical elasticity. The velocity of the lowest order g_S0 mode is similar to that of the Lamb S0 mode for very low frequencies (up to 0.06 MHz). However, as frequency increases, the g_S0 mode starts rapidly to diverge from the S0 mode. It is known that S0 as well as A0 Lamb modes approach asymptotically the Rayleigh velocity (Rose, 1999). As it is shown in Fig. 2(a), the g_S0 mode seems to approach the dispersive values of the bulk shear wave. Given that in classical elasticity the Rayleigh velocity is very close to the bulk shear velocity ($c_R=0.92c_T$, where c_R is the Rayleigh velocity) (Rose, 1999), we can say that the g_S0 mode converges actually to the velocity of Rayleigh wave, which according to (Georgiadis *et al.*, 2004) is also dispersive.

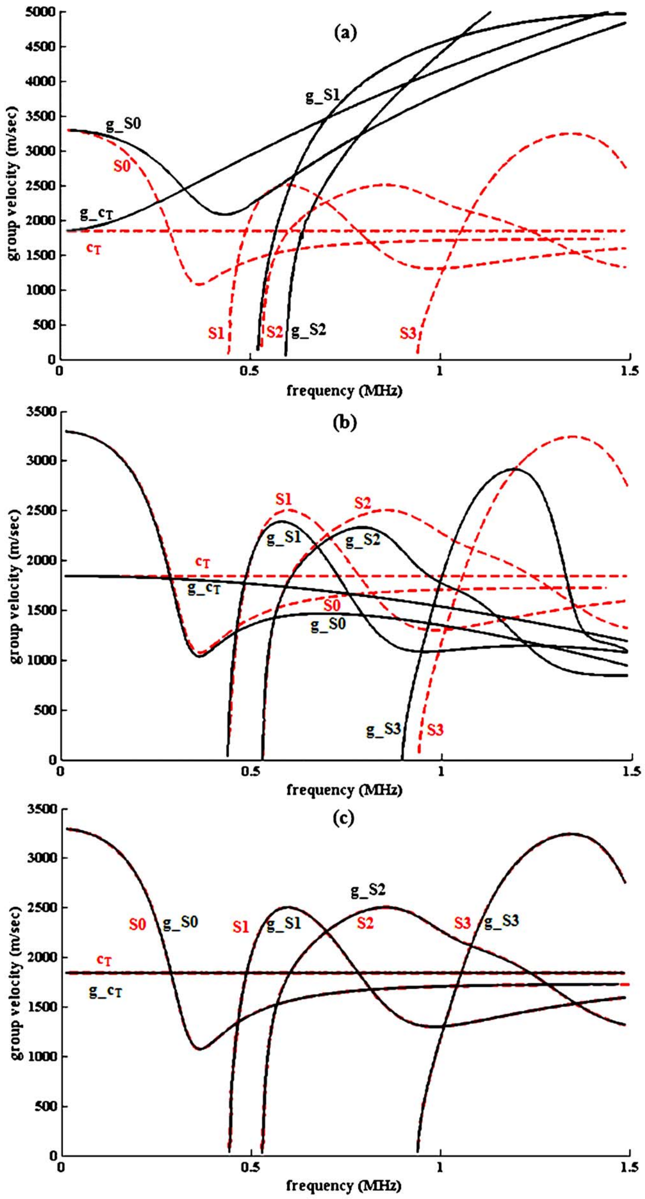


FIG. 2. (Color online) Group velocity dispersion curves of the symmetric modes for a free bone-mimicking plate for the case of the classical (dashed lines) and the gradient theory of elasticity (solid lines) for (a) Case-1a: $g>h$ $(g=5\times10^{-4},\ h=10^{-4})$, (b) Case-1b: $g<h$ $(g=10^{-5},\ h=10^{-4})$, and (c) Case-1c: $g=h$ $(g=10^{-4},\ h=10^{-4})$.

The velocity dispersion of the higher-order modes is considerably different from the Lamb modes, even at low frequencies. More specifically, g_S1 and g_S2 have different cut-off frequencies (0.52 MHz and 0.59 MHz, respectively) and their group velocities rapidly increase with frequency. The group velocity of g_S1 seems to converge to the velocity value of the bulk shear wave in the gradient elasticity (note that in classical elasticity the group velocities of the higher-order Lamb modes converge to the bulk shear velocity of bone). However, for all the other higher-order modes, convergence is expected at higher frequencies than those computed in the present study. Additional modes, such as g_S3, are anticipated but their cut-off frequencies are higher than 1.5 MHz. Similarly, in Case-1b ($g < h$), the dispersion curves of the guided modes also diverge from the classical Lamb modes (the mode g_S3 can now be seen in this case). Similar to what occurred for the bulk shear wave, the modes' group velocities take again lower values than those of the classical Lamb waves; nevertheless their deviation is again less pronounced than in Case-1a. In this case, the convergence of the modes to the bulk shear velocity can be better observed.

As opposed to the aforementioned two subcases, in Case-1c ($g = h$), no differences exist in mode dispersion that is identical between the two theories of elasticity, which is also predicted by the theory.

Figures 3(a)–3(c) represent the group velocity dispersion curves of the antisymmetric modes obtained from the gradient elastic plate in Cases-1a, 1b, and 1c, respectively. Similar to the bulk shear waves and as it is expected (Fig. 1), the bulk longitudinal wave, denoted in the figures as g_c_L, becomes dispersive. The bulk longitudinal velocity, denoted herein as g_c_L, for very low frequencies, is equal to that of the bulk longitudinal wave propagating in a classical elastic medium (depicted by a straight dashed line and is denoted as c_L). Nevertheless, for higher frequencies in Case-1a, the g_c_L rapidly increases from the bulk longitudinal velocity in the classical case, whereas in Case-1b, it decreases exhibiting a similar behavior to the corresponding bulk shear wave. In Case-1c ($g = h$), as expected, the bulk longitudinal wave in the gradient elastic case is non-dispersive.

Regarding the velocity dispersion of the antisymmetric guided modes, similar conclusions can be drawn to those for the symmetric modes. In particular, in Case-1a [Fig. 3(a)], the group velocities of the g_A0 and g_A1 modes are close to those of the Lamb A0 and A1 modes for very low frequencies, but as the frequency increases they become significantly higher than the Lamb modes. Analogous trends are also observed for the g_A2 and g_A3 modes. These modes converge to the bulk shear values. In Case-1b [Fig. 3(b)], the antisymmetric modes are again affected by the microstructure; the group velocities are lower than those of Lamb modes and the effect is less pronounced than Case-1a. Finally, in the Case-1c [Fig. 3(c)], the velocity dispersion curves are again identical to those for the Lamb waves.

Figures 4(a)–4(c) and 5(a)–5(c) illustrate the group velocity dispersion curves for Case-2, i.e., for $h = 10^{-5}$. Similar trends can be observed for the Cases-2a and 2c [Figs. 4(a) and 4(c) and Figs. 5(a) and 5(c)], but the observed microstructural effects are much more mitigated than in Figs. 2(a), 2(c), 3(a), and 3(c). However, for the Case-2b, i.e., for $g = 5 \times 10^{-6}$, the dispersion curves for the symmetric and antisymmetric modes are almost equal to those in the classical elasticity.

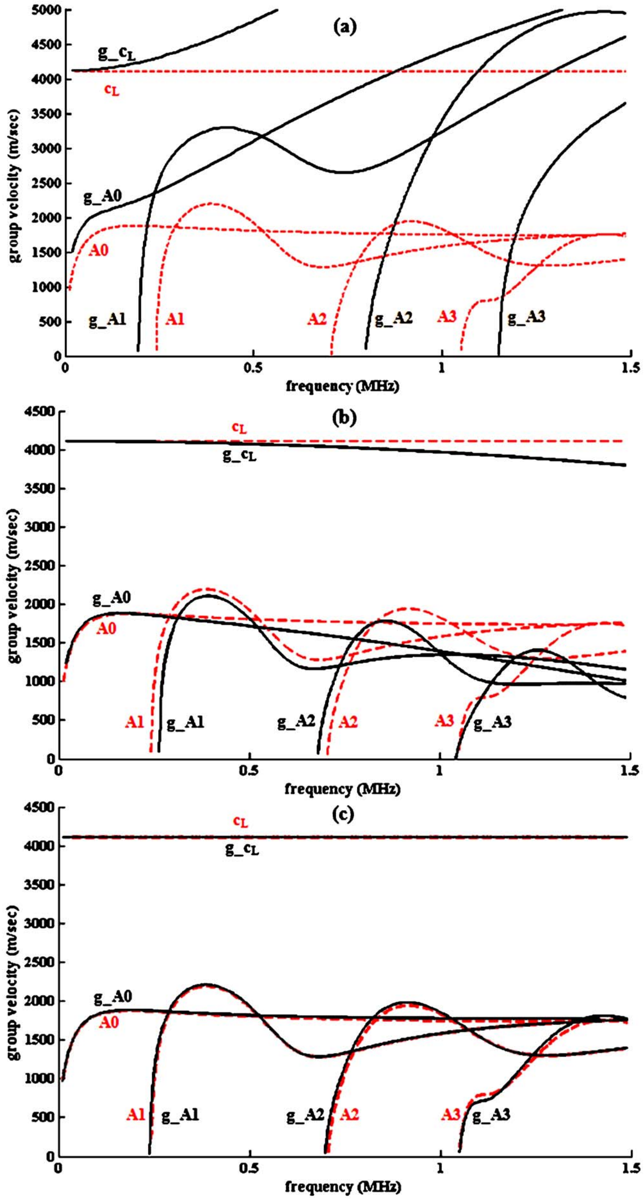


FIG. 3. (Color online) Group velocity dispersion curves of the antisymmetric modes for a free bone-mimicking plate for the case of the classical (dashed lines) and the gradient theory of elasticity (solid lines) for (a) Case-1a: $g > h$ ($g = 5 \times 10^{-4}$, $h = 10^{-4}$), (b) Case-1b: $g < h$ ($g = 10^{-5}$, $h = 10^{-4}$), and (c) Case-1c: $g = h$ ($g = 10^{-4}$, $h = 10^{-4}$).

VI. DISCUSSION

In this paper, a study on the propagation of ultrasound in a free plate with microstructural effects was presented. The dipolar gradient elasticity is the enhanced theory exploited for the dynamic behavior of the considered plate. Comparisons with the solutions derived from the Lamb problem in the classical elasticity were also made to investigate the effect of the microstructure on guided wave propagation in 2D plates. Group velocity dispersion curves were obtained for a testing case in which the medium was assumed to have properties similar to those of cortical bone.

The bone is a material with microstructural effects, the mechanical behavior of which can be successfully modeled by enhanced elastic theories. Although, in the works of Fatemi *et al.* (2002) and Yoon and Katz (1983), many higher-

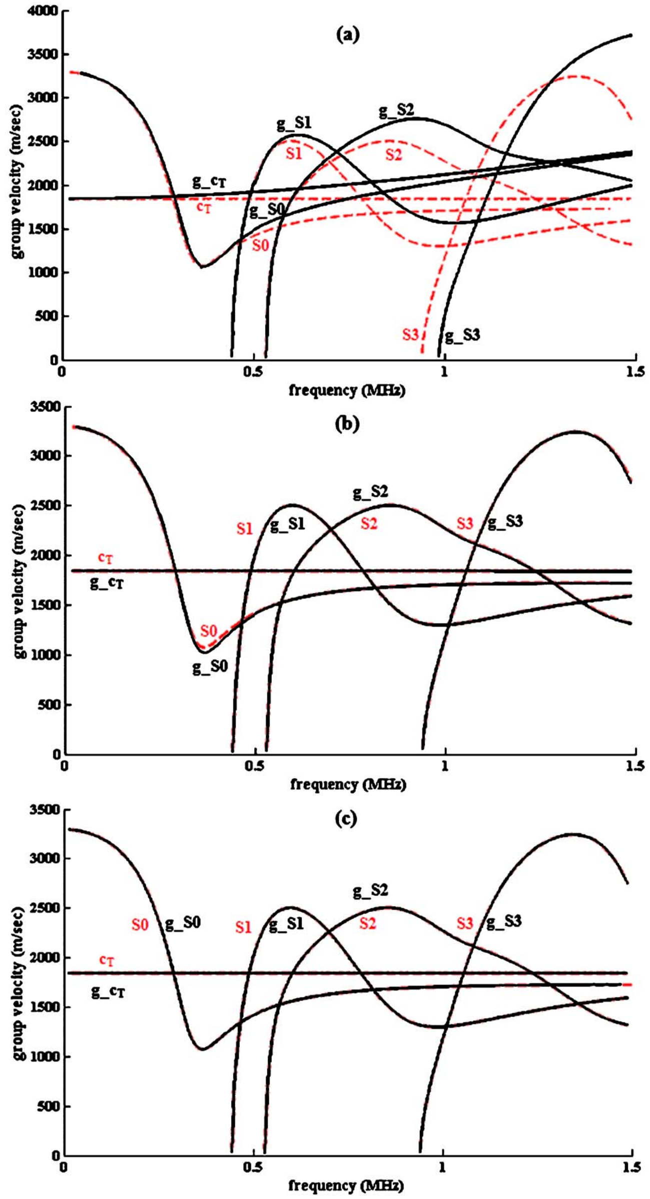


FIG. 4. (Color online) Group velocity dispersion curves of the symmetric modes for a free bone-mimicking plate for the case of the classical (dashed lines) and the gradient theory of elasticity (solid lines) for (a) Case-2a: $g > h$ ($g = 10^{-4}$, $h = 10^{-5}$), (b) Case-2b: $g < h$ ($g = 5 \times 10^{-6}$, $h = 10^{-5}$), and (c) Case-2c: $g = h$ ($g = 10^{-5}$, $h = 10^{-5}$).

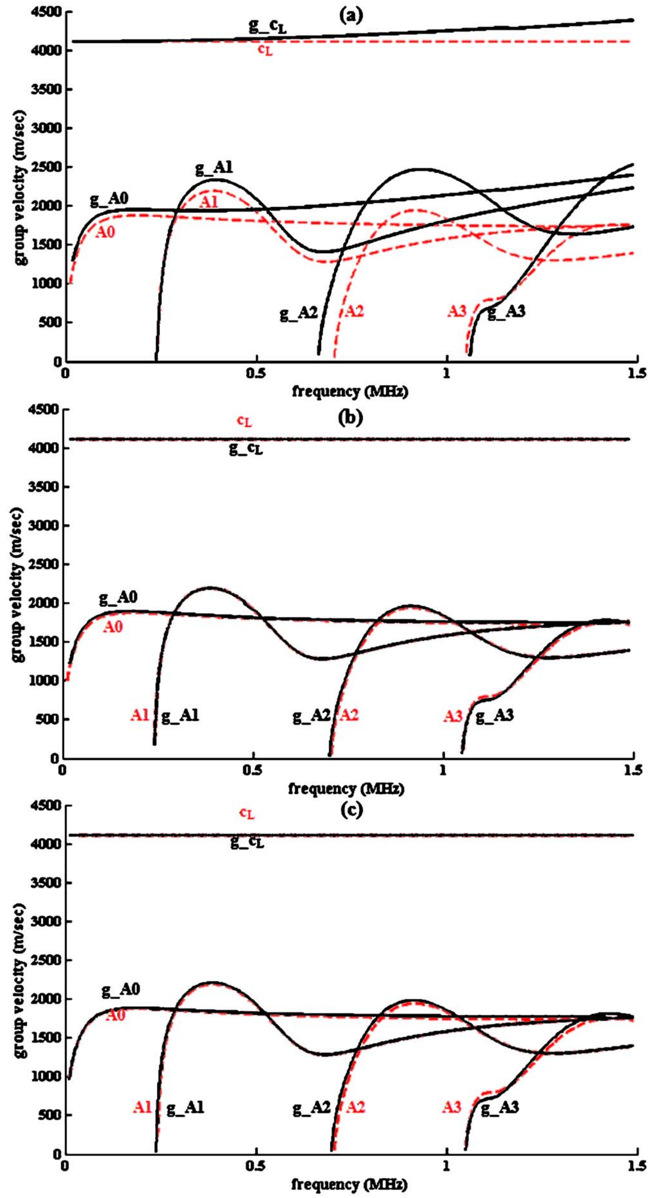


FIG. 5. (Color online) Group velocity dispersion curves of the antisymmetric modes for a free bone-mimicking plate for the case of the classical (dashed lines) and the gradient theory of elasticity (solid lines) for (a) Case-2a: $g > h$ ($g = 10^{-4}$, $h = 10^{-5}$), (b) Case-2b: $g < h$ ($g = 5 \times 10^{-6}$, $h = 10^{-5}$), and (c) Case-2c: $g = h$ ($g = 10^{-5}$, $h = 10^{-5}$).

order elastic theories are proposed for the description of the micromechanical effects in bones, only couple stress theories [mainly Cosserat and micropolar] have been utilized up to now for this purpose. The main reasons for this are as follows: (i) the use of Cosserat-micropolar theories in bending and torsion problems seems to be the most reasonable due to introduction of couple stresses, (ii) the higher-order gradient elastic theories of Mindlin (1964, 1965) as initially proposed were much more complicated than those of couple stresses, and (iii) for numerical solutions, the fourth order derivatives introduced in the differential operators of the higher-order gradient elastic equilibrium equations and equations of motion render the development of a direct finite element algorithm a difficult task since $C^{(1)}$-continuity elements are required. However, during the past decade the simplified versions of Mindlin's general elastic theory with microstructure (Georgiadis *et al.*,. 2004; Tsepoura *et al.*, 2002; Vardoulakis and Georgiadis, 1997) have gained much attention since (i) only one microstructural parameter for static problems and two for dynamic ones have to be determined instead of four required in couple stresses theories, (ii) in contrary to Cosserat and micropolar elasticity, all tensors involved in the aforementioned gradient elastic theories are symmetric being thus mathematically simpler and more understandable from a physical point of view, and (iii) for fracture mechanics problems (very important for applications in bones) gradient elastic theories lead to more realistic results than couple stresses ones (Amanatidou and Aravas, 2002; Karlis *et al.*, 2007, 2008; Stamoulis and Giannakopoulos,

2008) predicting phenomena associated with cusp-like crack profiles and development of process zone in front of crack tip observed experimentally.

The problem of wave propagation in plates with microstructure has been solved analytically only in the context of the Cosserat theory (Kulesh *et al.*, 2007). The same problem is treated here with the aid of the dipolar gradient elastic theory. Since the determination of the internal characteristic length scale parameters remains an open issue in the literature, we considered six different combinations between the gradient coefficient g and the inertia internal length h. In all subcases, the values of g and h (see Table I) were in accordance with previous experimental studies in bones (Lakes, 1982) and were assumed to be close to the size of the osteons, i.e., the microstructural level of bone's hierarchical structural organization (Rho, 1998). The value $g=10^{-3}$ m in Case-1 was deliberately ignored here since it leads to a rapid increase in the group velocities which would be rather undetectable in the plots. Also, in Case-2b for $g=5\times10^{-6}$ m, we observed that the dispersion curves are almost equal to those in the classical elasticity (the same also happened for 10^{-6} m, not shown herein). The obtained results make clear that the values of the two length scale parameters play an important role in the velocity trends of the guided modes, since our observations ranged from extreme to no differences at all. When the two internal material lengths g and h become zero or equal to each other, the presented theory provides velocity dispersion curves identical to those anticipated by the classical elasticity. This was numerically verified in our study, as shown in Figs. 2(c), 3(c), 4(c), and 5(c).

It becomes apparent that reasonable estimations for the relation between the material coefficients and the determination of their values can only be made by comparing experimental measurements with those predicted by the theory. In the analytical study of Georgiadis *et al.* (2004) dealing with Rayleigh wave dispersion in a gradient elastic half-plane, the value of $g=4\times10^{-5}$ m is proposed as the best value to describe sufficiently Rayleigh wave dispersion in a geomaterial. However, it is obvious that the values of the internal characteristic lengths vary according to the mechanical properties of the testing material.

Figures 2–4 reveal that the bulk longitudinal and shear waves propagating in the gradient elastic plate are dispersive, which was also predicted by the theoretically derived equation (28) and the results depicted in Fig. 1. For some combinations of g and h, the deviation of the obtained velocity dispersion of the bulk waves from the constant velocity of the classical elastic ones becomes significant. For instance, in the Case-1a at 1 MHz, i.e., for wavelength ≈ 4 mm, the velocity of the bulk longitudinal wave was changed as much as 52% between the classical and the gradient theory of elasticity. This may play an important role when interpreting axial-transmission velocity measurements along the long axis of a bone as it is reported in previous experimental studies (Njeh *et al.*, 1999; Protopappas *et al.*, 2005). In the considered plate, for different values of g and h and for frequencies from 0.5 MHz (i.e., for wavelength ≈ 8 mm) to 1.5 MHz (i.e., for wavelength ≈ 2.7 mm), which is the commonly used spectrum in the ultrasonic bone studies (Protopappas *et al.*, 2008), the velocity dispersion of the guided waves was significantly modified from that of the Lamb waves. In a previous study (Protopappas *et al.*, 2006), by superimposing the theoretical Lamb wave dispersion curves, computed for a bone-mimicking plate, on the time-frequency representation of the signal obtained from *ex vivo* measurements on an intact tibia, we found that the propagating guided waves could not be sufficiently characterized by the Lamb modes. Therefore, the Lamb wave theory has limited efficiency in predicting wave guidance phenomena in real bones. This was further supported by the findings of two subsequent three-dimensional computational studies (Bossy *et al.*, 2004; Protopappas *et al.*, 2007) showing that for the same frequency excitation, irregularities in the tubular geometry of the cortex as well as the anisotropy and inhomogeneity of the bone also give rise to major changes in the dispersion of the modes predicted by the classical tube theory. To this end, the results obtained in the present analysis clearly show that the material dispersion, i.e., the dependence of bone properties (such as the bulk longitudinal and shear velocities) on the frequency, induced by the bone's microstructure even in frequencies well below 1 MHz is an additional parameter, which significantly affects the characteristics of wave propagation in bone.

For very low frequencies, slight velocity differences were observed between the gradient and the classical elastic theory, fact that probably justifies why previous studies provided a good representation of guided wave propagation in bone using simple homogeneous plate or tube models (Nicholson *et al.*, 2002; Protopappas *et al.*, 2008).

In axial-transmission applications the bone is in contact with the overlying soft tissues which provide leakage paths for the ultrasonic energy, resulting thus in modified dispersion curves, i.e., the leaky Lamb modes (Moilanen, 2008; Vavva *et al.*, 2008). Accounting for the effect of the soft tissues on mode dispersion in the considered 2D plate with microstructure would imply the application of continuity conditions at the interfaces between the plate surfaces and the soft tissues as in the classical elastic case. However, in that case, continuity conditions must be assumed not only for displacements and tractions (as in the classical theory) but also for the normal derivatives of the tractions and double tractions. Another simplification in our study was that bone anisotropy was not considered. The introduction of an anisotropic constitutive law in the present theoretical model would necessitate a more complicated expression for the potential energy density $\hat{W}$, incorporating additional elastic parameters. Therefore, anisotropy issues could be addressed using the present model requiring, however, more complicated theoretical and numerical calculations.

Although the present analysis clearly shows that bone microstructure significantly affects guided wave propagation, the theoretical results should be interpreted with cautiousness and only in conjunction with measurements from real bones before they can be used in clinical practice.

VII. CONCLUSIONS

In this work we presented an analytical study on guided wave propagation in 2D bone-mimicking plates with microstructure. For the first time, the simple theory of gradient elasticity is proposed to incorporate bone's microstructural effects into the stress analysis. Hence, two additional elastic constants (i.e., g and h) associated with micro-elastic and micro-inertia effects were considered. We demonstrated that when the elastic constants have different values, microstructure plays a significant role in the propagation of the bulk longitudinal and shear waves by inducing material and geometrical dispersion. It was also shown that the insertion of the microstructural characteristics into the stress analysis gives rise to major changes in the dispersion of the guided modes predicted by the classical Lamb wave theory. Although previous studies (Georgiadis *et al.*, 2004) report that the microstructural effects are important only at high frequencies, in the present work it was made clear that they can be equally significant at medium frequencies, i.e., 0.7–1 MHz (i.e., for wavelengths from 2.8 to 4 mm), which are within the region of interest in ultrasonic bone studies. The effect was dependent on the absolute values of the coefficients and was less pronounced for the smallest value of h (i.e., Cases-2a and 2b). Our findings show that bone's microstructure is an important factor which should be taken into account both in theoretical and computational studies on wave propagation in bones. The gradient theory of elasticity could potentially provide more accurate interpretation of clinical measurements on intact and healing long bones. This study could be regarded as a step toward the ultrasonic evaluation of bone, although experimental research is further needed.

ACKNOWLEDGMENTS

This work is part funded by the General Secreteriat of Research and Development, Hellenic Ministry of the Development ("PENED 2003," 03ED140).

APPENDIX

The components of the determinant of the two systems which correspond to the symmetric modes:

$$A_{11} = (4\mu p^3 + 2\lambda k^2 p + 2\lambda p^3)\sin ph,$$

$$A_{12} = 4\mu ikq^2 \sin qh,$$

$$A_{13} = (4\mu r_p^3 - 2\lambda k^2 r_p + 2\lambda r_p^3)\sin r_p h,$$

$$A_{14} = -4\mu ikr_s^2 \sin hr_s h,$$

$$A_{21} = -4\mu ikp^2 \cos ph,$$

$$A_{22} = 2\mu(-q^3 + k^2 q)\cos qh,$$

$$A_{23} = 4\mu ikr_p^2 \cos r_p h,$$

$$A_{24} = 2\mu(r_s^3 + k^2 r_s)\cos r_s h,$$

$$A_{31} = \left(4\mu pik + 4\mu p^3 ikg^2 + 2\lambda p^3 ikg^2 + 2\lambda pik^3 g^2 + 8\mu pik^3 g^2 - \frac{2\rho h^2\omega^2}{3}\right) ikp \sin ph,$$

$$A_{32} = \left(2\mu q^2 - 2\mu k^2 + 2\mu q^4 g^2 + 4\mu k^2 q^2 g^2 - 2\mu k^4 g^2 - \frac{2\rho h^2\omega^2}{3} q^2\right)\sin qh,$$

$$A_{33} = \left(-4\mu r_p ik + 4\mu r_p^3 ikg^2 + 2\lambda r_p^3 ikg^2 - 2\lambda r_p ik^3 g^2 - 8\mu ik^3 r_p g^2 + \frac{2\rho h^2\omega^2}{3} ikp\right)\sin hr_p h,$$

$$A_{34} = \left(-2\mu r_s^2 + 2\mu r_s^4 g^2 - 2\mu k^4 g^2 - 2\mu k^2 - 4\mu k^2 r_s^2 g^2 + \frac{2\rho h^2\omega^2}{3} r_s^2\right)\sin hr_s h,$$

$$A_{41} = \left(-4\mu p^2 - 2\lambda p^2 - 2\lambda k^2 - 2\lambda p^4 g^2 - 4\mu p^4 g^2 - 2\lambda k^4 g^2 - 4\lambda k^2 p^2 g^2 - 8\mu k^2 p^2 g^2 + \frac{2\rho h^2\omega^2}{3} p^2\right)\cos ph,$$

$$A_{42} = \left(-4\mu qik - 2\mu q^3 ikg^2 - 6\mu ik^3 qg^2 + \frac{2\rho h^2\omega^2}{3} ikq\right)\cos qh,$$

$$A_{43} = \left(4\mu r_p^2 + 2\lambda r_p^2 - 2\lambda k^2 - 2\lambda r_p^4 g^2 - 4\mu r_p^4 g^2 - 2\lambda k^4 g^2 + 4\lambda k^2 r_p^2 g^2 + 8\mu k^2 g^2 r_p^2 - \frac{2\rho h^2\omega^2}{3} r_p^2\right)\cos hr_p h,$$

$$A_{44} = \left(-4\mu r_s ik + 2\mu r_s^3 ikg^2 - 6\mu ik^3 r_s g^2 + \frac{2\rho h^2\omega^2}{3} ikr_s\right)\cosh hr_s h, \qquad \text{(A1)}$$

and for the antisymmetric modes:

$$B_{11} = (-4\mu p^3 - 2\lambda k^2 p - 2\lambda p^3)\cos ph,$$

$$B_{12} = 4\mu ikq^2 \cos qh,$$

$$B_{13} = -4\mu ikr_s^2 \cos hr_s h,$$

$$B_{14} = (4\mu r_s^3 - 2\lambda k^2 r_p + 2\lambda r_p^3)\cos hr_p h,$$

$$B_{21} = 4\mu ikp^2 \sin ph,$$

$$B_{22} = -2\mu(q^3 - k^2 q)\sin qh,$$

$$B_{23} = -2\mu(r_s^3 + k^2 r_s)\sin hr_s h,$$

$$B_{24} = -4\mu ikr_p^2 \sin hr_p h,$$

$$B_{31} = (4\mu pik + 4\mu p^3 ikg^2 + 2\lambda p^3 ikg^2 + 2\lambda ik^3 pg^2 \\ + 8\mu ik^3 pg^2)\cos ph,$$

$$B_{32} = \left(-2\mu q^2 + 2\mu k^2 - 2\mu q^4 g^2 - 4\mu k^2 q^2 g^2 + 2\mu k^4 g^2 \right. \\ \left. + \frac{2\rho h^2\omega^2}{3} q^2\right)\cos qh,$$

$$B_{33} = \left(2\mu r_s^2 - 2\mu r_s^4 g^2 + 2\mu k^4 g^2 + 2\mu k^2 + 4\mu k^2 r_s^2 g^2 \right. \\ \left. - \frac{2\rho h^2\omega^2}{3} r_s^2\right)\cos hr_s h,$$

$$B_{34} = \left(4\mu r_p ik - 4\mu ikg^2 r_p^3 - 2\lambda r_p^3 ikg^2 + 2\lambda ik^3 r_p g^2 \right. \\ \left. + 8\mu ik^3 r_p g^2 - \frac{2\rho h^2\omega^2}{3} ikr_p\right)\cos hr_p h,$$

$$B_{41} = \left(4\mu p^2 + 2\lambda p^2 + 2\lambda k^2 + 2\lambda p^4 g^2 + 4\mu p^4 g^2 + 2\lambda k^4 g^2 \right. \\ \left. + 4\lambda k^2 p^2 g^2 + 8\mu k^2 p^2 g^2 - \frac{2\rho h^2\omega^2}{3} p^2\right)\sin ph,$$

$$B_{42} = \left(-4\mu ikq - 2\mu ikq^3 g^2 - 6\mu ik^3 qg^2 \right. \\ \left. + \frac{2\rho h^2\omega^2}{3} ikq\right)\sin qh,$$

$$B_{43} = \left(4\mu r_s ik - 2\mu r_s^3 g^2 ik + 6\mu ik^3 r_s g^2 \right. \\ \left. - \frac{2\rho h^2\omega^2}{3} ikr_s\right)\sin hr_s h,$$

$$B_{44} = \left(-4\mu r_p^2 - 2\lambda r_p^2 + 2\lambda k^2 + 2r_p^4 g^2 + 4\mu r_p^4 g^2 + 2\lambda k^4 g^2 \right. \\ \left. - 4\lambda k^2 r_p^2 g^2 - 8\mu k^2 r_p^2 g^2 + \frac{2\rho h^2\omega^2}{3} r_p^2\right)\sin hr_p h. \tag{A2}$$

Aggelis, D. G., Philippidis, T. P., Tsinopoulos, S. V., and Polyzos, D. (**2004**). "Wave dispersion in concrete due to microstructure," in CD-ROM Proceedings of the 2004 International Conference on Computational & Experimental Engineering & Sciences, Madeira, Portugal, 26–29 July.

Aifantis, E. C. (**1999**). "Strain gradient interpretation of size effects," Int. J. Fract. **95**, 299–314.

Altan, B. S., Evensen, H., and Aifantis, E. C. (**1996**). "Longitudinal vibrations of a beam: A gradient elasticity approach," Mech. Res. Commun. **23**, 35–40.

Amanatidou, E., and Aravas, N. (**2002**). "Mixed finite element formulations of strain-gradient elasticity problems," Comput. Methods Appl. Mech. Eng. **191**, 1723–1751.

Artan, R., and Altan, B. (**2002**). "Propagation of SV waves in a periodically layered media in nonlocal elasticity," Int. J. Solids Struct. **39**, 5927–5944.

Ben-Amoz, M. (**1976**). "A dynamic theory for composite materials," Z. Angew. Math. Phys. **27**, 83–99.

Bennett, T., Gitman, I. M., and Askes, H. (**2007**). "Elasticity theories with higher order gradients of inertia and stiffness for modelling of wave dispersion in laminates," Int. J. Fract. **148**, 185–193.

Bossy, E., Talmant, M., and Laugier, P. (**2002**). "Effect of cortical thickness on velocity measurements using ultrasonic axial transmission: A 2D simulation study," J. Acoust. Soc. Am. **112**, 297–307.

Bossy, E., Talmant, M., and Laugier, P. (**2004**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Camus, E., Talmant, M., Berger, G., and Laugier, P. (**2000**). "Analysis of the axial transmission technique for the assessment of skeletal status," J. Acoust. Soc. Am. **108**, 3058–3065.

Chakraborty, A. (**2007**). "Wave propagation in anisotropic media with nonlocal elasticity," Int. J. Solids Struct. **44**, 5723–5741.

Chakraborty, A. (**2008**). "Prediction of negative dispersion by a nonlocal poroelastic theory," J. Acoust. Soc. Am. **123**, 56–67.

Chang, C. S., Gao, J., and Zhong, X. (**1998**). "High-gradient modeling for Love wave propagation in geological materials," J. Eng. Mech. **124**, 1354–1359.

Chen, C. P., and Lakes, R. S. (**1989**). "Dynamic wave dispersion and loss properties of conventional and negative Poisson's ratio polymeric cellular materials," Cell. Polym. **8**, 343–369.

Chen, Y., Lee, J. D., and Eskandarian, A. (**2003**). "Examining the physical foundation of continuum theories from the viewpoint of phonon dispersion relation," Int. J. Eng. Sci. **41**, 61–83.

Chimenti, D. E. (**1997**). "Guided waves in plates and their use in material characterization," Appl. Mech. Rev. **50**, 247–284.

Cosserat, E., and Cosserat, F. (**1909**). *Théorie des Corps Déformables (Theory of Deformable Structures)* (Hermann et Fils, Paris).

Dodd, S. P., Cummingham, J. L., Miles, A. W., Gheduzzi, S., and Humphrey, V. F. (**2007**). "An in vitro study of ultrasound signal loss across simple fractures in cortical bone mimics and bovine cortical bone samples," Bone **40**, 656–661.

Eringen, A. C. (**1966**). "Linear theory of micropolar elasticity," J. Math. Mech. **15**, 909–923.

Eringen, A. C. (**2002**). *Nonlocal Continuum Field Theories* (Springer, New York).

Erofeyev, V. I. (**2003**). *Wave Processes in Solids With Micro-Structure* (World Scientific, Singapore).

Exadaktylos, G. E., and Vardoulakis, I. (**2001**). "Microstructure in linear elasticity and scale effects: A reconsideration of basic rock mechanics and rock fracture mechanics," Tectonophysics **335**, 81–109.

Fatemi, J., Van Keulen, F., and Onck, P. R. (**2002**). "Generalized continuum theories: Application to stress analysis in bone," Meccanica **37**, 385–396.

Gazis, D. C., Herman, R., and Wallis, R. F. (**1960**). "Surface elastic waves in cubic crystals," Phys. Rev. **119**, 533–544.

Georgiadis, H. G., Vardoulakis, I., and Velgaki, E. G. (**2004**). "Dispersive Rayleigh-wave propagation in microstructured solids characterized by dipolar gradient elasticity," J. Elast. **74**, 17–45.

Georgiadis, H. G., and Velgaki, E. G. (**2003**). "High-frequency Rayleigh waves in materials with micro-structure and couple-stress effects," Int. J. Solids Struct. **40**, 2501–2520.

Graff, F. K., *Wave Motion in Elastic Solids* (Oxford University Press, Oxford, **1975**).

Hsia, S., Chiu, S., and Cheng, J. (**2006**). "Wave propagation at the human muscle-compact bone interface," Theor Appl. Mech. **33**, 223–243.

Karlis, G. F., Tsinopoulos, S. V., Polyzos, D., and Beskos, D. E. (**2007**). "Boundary element analysis of mode I and mixed mode (I and II) crack problems of 2-D gradient elasticity," Comput. Methods Appl. Mech. Eng. **196**, 5092–5103.

Karlis, G. F., Tsinopoulos, S. V., Polyzos, D., and Beskos, D. E. (**2008**). "2D and 3D boundary element analysis of mode-I cracks in gradient elasticity," Comput. Model. Eng. Sci. **26**, 189–207.

Kondratev, A. I. (**1990**). "Precision measurements of the velocity and attenuation of ultrasound in solids," Sov. Phys. Acoust. **36**, 262–265.

Kulseh, M. A., Matveenko, V. P., and Shardakov, I. N. (**2007**). “Constructing an solution for Lamb waves using Cosserat continuum approach,” J. Appl. Mech. Tech. Phys. **48**, 119–125.

Kumar, R., and Partap, G. (**2006**). “Rayleigh Lamb waves in micropolar isotropic elastic plate,” Appl. Math. Mech. **27**, 1049–1059.

Lakes, R. S. (**1981**). “Dynamical study of couple stress effects in human compact bone,” J. Biomech. Eng. **104**, 6–11.

Lakes, R. S. (**1982**). “Dynamical study of couple stress effects in human compact bone,” J. Biomed. Eng. **104**, 6–11.

Lakes, R. S. (**1995**). “Experimental methods for study of Cosserat elastic solids and other generalized elastic continua,” in *Continuum Models for Materials With Microstructure*, edited by H. B. Muhlhaus (Wiley, Chichester).

Lam, D. C. C., Yang, F., Chong, A. C. M., Wang, J., and Tong, P. (**2003**). “Experiments and theory in strain gradient elasticity,” J. Mech. Phys. Solids **51**, 1477–1508.

Midya, G. K. (**2004**). “On Love-type surface waves in homogeneous micropolar elastic media,” Int. J. Eng. Sci. **42**, 1275–1288.

Mindlin, R. D. (**1964**). “Micro-structure in linear elasticity,” Arch. Ration. Mech. Anal. **16**, 51–78.

Mindlin, R. D. (**1965**). “Second gradient of strain and surface-tension in linear elasticity,” Int. J. Solids Struct. **1**, 417–438.

Moilanen, P. (**2008**). “Ultrasonic guided waves in bone,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1277–1286.

Nicholson, P. H. F., Moilanen, P., Karkkainen, T., Timonen, J., and Cheng, S. (**2002**). “Guided ultrasonic waves in long bones: Modeling, experiment and in vivo application,” Physiol. Meas. **23**, 755–768.

Njeh, C. F., Hans, D., Wu, C., Kantorovich, E., Sister, M., Fuerst, T., and Genant, H. K. (**1999**). “An in vitro investigation of the dependence on sample thickness of the speed of sound along the specimen,” Med. Eng. Phys. **21**, 651–659.

Ottosen, N. S., Ristinmaa, M., and Ljung, C. (**2000**). “Rayleigh waves obtained by the indeterminate couple-stress theory,” Eur. J. Mech. A/Solids **19**, 929–947.

Papargyri-Beskou, S., and Beskos, D. E. (**2008**). “Static, stability and dynamic analysis of gradient elastic flexural Kirchhoff plates,” Arch. Appl. Mech. **78**, 625–635.

Papargyri-Beskou, S., Tsepoura, K. G., Polyzos, D., and Beskos, D. E. (**2003**). “Bending and stability analysis of gradient elastic beams,” Int. J. Solids Struct. **40**, 385–400.

Park, H. C., and Lakes, R. S. (**1986**). “Cosserat micromechanics of human bone: Strain redistribution by a hydration sensitive constituent,” J. Biomech. **19**, 385–397.

Protopappas, V. C., Baga, D., Fotiadis, D. I., Likas, A., Papachristos, A. A., and Malizos, K. N. (**2005**). “An ultrasound wearable system for the monitoring and acceleration of fracture healing in long bones,” IEEE Trans. Biomed. Eng. **52**, 1597–1608.

Protopappas, V. C., Fotiadis, D. I., and Malizos, K. N. (**2006**). “Guided ultrasound wave propagation in intact and healing long bones,” Ultrasound Med. Biol. **32**, 693–708.

Protopappas, V. C., Kourtis, I. C., Kourtis, L. K., Malizos, K. N., Massalas, C. V., and Fotiadis, D. I. (**2007**). “Three-dimensional finite element modeling of guided ultrasound wave propagation in intact and healing long bones,” J. Acoust. Soc. Am. **121**, 3907–3921.

Protopappas, V. C., Vavva, M. G., Fotiadis, D. I., and Malizos, K. N. (**2008**). “Ultrasonic monitoring of bone fracture healing,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1243–1255.

Rho, J. Y. (**1998**). “Mechanical properties and the hierarchical structure of bone,” Med. Eng. Phys. **20**, 92–102.

Rose, J. L. (**1999**). *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge).

Ru, C. Q., and Aifantis, E. C. (**1993**). “A simple approach to solve boundary value problems in gradient elasticity,” Acta Mech. **101**, 59–68.

Sasso, M., Haiat, G., Yamato, Y., Naili, S., and Matsukawa, M. (**2008**). “Dependence of ultrasonic attenuation on bone mass and microstructure in bovine cortical bone,” J. Biomech. **41**, 347–355.

Sluys, L. J., de Brost, R., and Mulhlaus, H. B. (**1993**). “Wave propagation, localization and dispersion in gradient dependent medium,” Int. J. Solids Struct. **30**, 1153–1171.

Stamoulis, K., and Giannakopoulos, A. E. (**2008**). “Size effects on strength, toughness and fatigue crack growth of gradient elastic solids,” Int. J. Solids Struct. **45**, 4921–4935.

Stavropoulou, M., Exadaktylos, G., Papamichos, E., Larsen, I., and Ringstad, C. (**2003**). “Rayleigh wave propagation in intact and damaged geomaterials,” Int. J. Rock Mech. Min. Sci. **40**, 377–387.

Suiker, A. S. J., Metrikine, A. V., and de Borst, R. (**2001**). “Comparison of wave propagation characteristics of the Cosserat continuum model and corresponding discrete lattice models,” Int. J. Solids Struct. **38**, 1563–1583.

Tatarinov, A., Sarvazyan, N., and Sarvazyan, A. (**2005**). “Use of multiple acoustic wave modes for assessment of long bones: Model study,” Ultrasonics **43**, 672–680.

Tekoglu, C. (**2007**). “Size effects in cellular solids,” Ph.D. thesis, University of Groningen, Groningen, The Netherlands.

Tekoglu, C., and Onck, P. R. (**2005**). “Size effects in the mechanical behavior of cellular materials,” J. Mater. Sci. **40**, 5911–5917.

Tiersten, H. F., and Bleustein, J. L. (**1974**). “Generalized elastic continua,” in *R.D. Mindlin and Applied Mechanics*, edited by G. Herman (Pergamon, New York).

Tomar, S. K., and Gogna, M. L. (**1995**). “Reflection and refraction of longitudinal waves at an interface between two micropolar elastic media in welded contact,” J. Acoust. Soc. Am. **97**, 822–830.

Tsepoura, K. G., Papargyri-Beskou, S., and Polyzos, D. (**2002**). “A boundary element method for solving 3D static gradient elastic problems with surface energy,” Comput. Mech. **29**, 361–381.

Vardoulakis, I., and Georgiadis, H. G. (**1997**). “SH surface waves in a homogeneous gradient elastic half-space with surface energy,” J. Elast. **47**, 147–165.

Vardoulakis, I., and Giannakopoulos, A. E. (**2006**). “An example of double forces taken from structural analysis,” Int. J. Solids Struct. **43**, 4047–4062.

Vardoulakis, I., and Sulem, J. (**1995**). *Bifurcation Analysis in Geomechanics* (Blackie,London, Chapman and Hall, London).

Vavva, M. G., Protopappas, V. C., Gergidis, L. N., Charalampopoulos, A., Fotiadis, D. I., and Polyzos, D. (**2008**). “The effect of boundary conditions on guided wave propagation in two-dimensional models of healing bone,” Ultrasonics **48**, 598–606.

Wear, K. A. (**2007**). “Group velocity, phase velocity, and dispersion in human calcaneus in vivo,” J. Acoust. Soc. Am. **121**, 2431–2437.

Wolfram Research, Inc. (**2004**). MATHEMATICA, Wolfram Research, Inc., Champaign, IL.

Yamato, Y., Matsukawa, M., Otani, T., Yamazaki, K., and Nagano, A. (**2006**). “Distribution of longitudinal wave properties in bovine cortical bone in vitro,” Ultrasonics **44**, e233–e237.

Yang, J. F. C., and Lakes, R. S. (**1981**). “Transient study of couple stress effects in compact bone: Torsion,” J. Biomech. Eng. **103**, 275–279.

Yang, J. F. C., and Lakes, R. S. (**1982**). “Experimental study of micropolar and couple stress elasticity in compact bone in bending,” J. Biomech. **15**, 91–98.

Yerofeyev, V. I., and Sheshenina, O. A. (**2005**). “Waves in a gradient-elastic medium with surface energy,” J. Appl. Math. Mech. **69**, 57–69.

Yim, H., and Sohn, Y. (**2000**). “Numerical simulation and visualization of elastic waves using mass-spring lattice model,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 550–558.

Yoon, H., and Katz, L. J. (**1983**). “Is bone a Cosserat solid?,” J. Mater. Sci. **18**(5), 1297–1305.

Field recordings of Gervais' beaked whales *Mesoplodon europaeus* from the Bahamas

Douglas Gillespie[a)]
Sea Mammal Research Unit, University of St Andrews, St Andrews KY16 8LB, United Kingdom

Charlotte Dunn
Bahamas Marine Mammal Research Organisation, P.O. Box AB-20714, Marsh Harbour, Abaco, Bahamas and Sea Mammal Research Unit, University of St Andrews, St Andrews KY16 8LB, United Kingdom

Jonathan Gordon
Sea Mammal Research Unit, University of St Andrews, St Andrews KY16 8LB, United Kingdom

Diane Claridge
Bahamas Marine Mammal Research Organisation, P.O. Box AB-20714, Marsh Harbour, Abaco, Bahamas and Sea Mammal Research Unit, University of St Andrews, St Andrews KY16 8LB, United Kingdom

Clare Embling
School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, United Kingdom

Ian Boyd
Sea Mammal Research Unit, University of St Andrews, St Andrews KY16 8LB, United Kingdom

(Received 23 December 2008; revised 3 March 2009; accepted 10 March 2009)

The first recordings from free-ranging Gervais' beaked whale (*Mesoplodon europaeus*) are presented. Nine Gervais' beaked whales were observed visually for over 6 h. Clicks were only detected over a 15 min period during the encounter, which coincided with an 88 min period during which no whales were observed at the surface. Click lengths were typically around 200 μS and their dominant energy was in the frequency range 30–50 kHz. While these clicks were broadly similar to those of Cuvier's and Blainville's beaked whales, the Gervais' beaked whale clicks were at a slightly higher frequency than those of the other species. *© 2009 Acoustical Society of America.* [DOI: 10.1121/1.3110832]

PACS number(s): 43.80.Ka [WWA] Pages: 3428–3433

I. INTRODUCTION

Beaked whales are a group of medium sized, deep diving toothed whales. They inhabit oceanic waters and make extremely deep and long foraging dives (Tyack *et al.*, 2006). These elusive and poorly known animals have been the focus of particular interest in recent years as a result of repeated incidents during which the use of mid-frequency military sonar has resulted in mass stranding and multiple mortalities of beaked whales. A number of well investigated incidents over the past decade, as well as an analysis of historical stranding databases, have now firmly established the link between military sonar exercises and beaked whale mortality events (Frantzis, 1998; Evans *et al.*, 2001; Jepson *et al.*, 2003; Fernandez *et al.*, 2005; Cox *et al.*, 2006). The mechanism by which these stranding occur and the reasons for beaked whale's unusual sensitivity remains unresolved, although it is now felt likely that the process is the secondary result of altered behavior rather than a direct acoustic impact (Cox *et al.*, 2006). Incidents so far have involved Blainville's beaked whale (*Mesoplodon densirostris*), Cuvier's beaked whale (*Ziphius cavirostris*), and Gervais' beaked whale (*Mesoplodon europaeus*), but this may reflect densities and distributions of beaked whales in sonar exercise areas as much as the susceptibility of different species to sonar.

Beaked whales are difficult animals to observe at sea (Barlow and Gisiner, 2006) and an ability to detect them acoustically using passive acoustic monitoring (PAM) could assist in attempts to mitigate sonar impacts. PAM could provide better information on the animals' distribution, it could allow real time monitoring before and during exercises, and it may have a role in facilitating research aimed at understanding the mechanisms behind stranding events (Cox *et al.*, 2006). Effective PAM requires a better understanding of the signal type and the animals' acoustic behavior.

Beaked whales are an unusually diverse group with 21 genetically confirmed species (Dalebout *et al.*, 2004). Wide bandwidth recordings have been made and reported from only a handful of species, however. Recent work, using Digital recording TAGS (DTAG—Johnson and Tyack, 2003) attached to individual whales using suction cups, has provided extremely detailed information on the acoustic characteristics of the sounds produced by two species: Cuvier's and Blainville's beaked whales (Johnson *et al.*, 2004; Madsen *et al.*, 2005; Zimmer *et al.*, 2005; Johnson *et al.*, 2006).

The dominant energy in Cuvier's beaked whale clicks was shown to be between 30 and 45 kHz and click lengths

[a)]Author to whom correspondence should be addressed. Electronic mail: dg50@st-andrews.ac.uk were around 200 μs. The clicks of Blainville's beaked whales were broadly similar to those of the Cuvier's. The recording bandwidth of the tag used in some of the earlier studies was 48 kHz, leading to speculation that there may be energy at even higher frequencies. However, later work using DTAGs with sampling rates of 192 kHz and a cabled hydrophone system with sensitivity up to 180 kHz (Zimmer *et al.*, 2005) have confirmed that the dominant frequency for Cuvier's beaked whales is indeed in the 30–40 kHz region.

By combining tracking data and recordings from DTAGs on two Cuvier's beaked whales tagged concurrently, Zimmer *et al.* (2005) were able to calculate a peak to peak source level for clicks of 214 dB re 1 μPa at 1 m and a directionality index of 25dB.

Clicks of similar frequency have also been recorded from other beaked whale species including Baird's beaked whale (*Berardius bairdii*), which had frequency peaks between 23 and 42 kHz (Dawson *et al.*, 1998) and northern bottlenose whales (*Hyperoodon ampullatus*) which had frequency peaks at a mean frequency of 24 kHz while foraging (Hooker and Whitehead, 2002).

Whistles or tonal vocalizations have rarely been recorded from beaked whales. Rogers and Brown (1999) reported on recordings made from Arnoux's beaked whales (*Berardius arnuxii*) made using audiocassette tape with an upper frequency response of ~16 kHz. As well as clicks, these authors reported whistles with a mean length of 0.65 s which were highly frequency modulated with multiple harmonics in the 2–6 kHz range. They also recorded amplitude modulated pulsed tones of similar duration and dominant frequencies in the 1–8.5 kHz range.

DTAG recordings have provided detailed information on the acoustic behavior of Blainville's and Cuvier's beaked whales (Madsen *et al.*, 2005; Tyack *et al.*, 2006; Zimmer *et al.*, 2005; Johnson *et al.*, 2006). In both species clicks have only been detected during deep foraging dives at depths between 222 and 1885 m. There is strong evidence that clicks are used for echolocation (Madsen *et al.*, 2005). Search clicks are produced in regular trains, with typical inter-click intervals (ICIs) of around 0.3–0.4 s and these are interspersed with buzz clicks, rapid bursts of higher frequency clicks produced as whales approach their prey (Johnson *et al.*, 2006).

Caldwell and Caldwell (1991) described clicks and a tonal sound recorded from a male Gervais beaked whale held in captivity following a live stranding. Sounds from the animal were recorded using a Uher 4400-Report tape recorder with an upper frequency limit of 20 kHz. During recording sessions over several days, clicks described as having a "high amplitude" as well as a tonal sound at 6 kHz were recorded. ICIs evident in the spectrograms in the article are consistent with those of other beaked whale species. Click energy is, however, at a much lower frequency than has been reported from other species, the dominant energy being generally below 3 kHz. The tonal sound is approximately 0.1 s long and is modulated in frequency, first sweeping down from 6 to 5 kHz, and then back up to slightly over 6 kHz.

Gervais' beaked whales have rarely been sighted in the wild, and as such very little is known about their ecology. They have been described as widely distributed in deepwater habitats in warm temperate and tropical waters of the North and South Atlantic (Jefferson *et al.*, 2008). Most knowledge about their distribution comes from strandings, most of which have been reported between Cape Cod Bay and Florida on the eastern sea board of the United States, and it is the most frequently stranded *Mesoplodon* in this region (Norman and Mead, 2001; MacLeod *et al.*, 2006; Waring *et al.*, 2007). The species is also found in the Gulf of Mexico and through the Caribbean. South of the equator, strandings have been reported along the Atlantic coast of South America as far south as Brazil and Ascension Island (MacLeod *et al.*, 2006). Strandings have been less frequently reported on the eastern side of the Atlantic and Mediterranean (Podesta *et al.*, 2005) but seem to occur over roughly the same range of latitudes.

Gervais' beaked whales are known from the Bahamas from six single stranding events and one confirmed sighting at sea in March 2001 (Balcomb, 1981; Balcomb and Claridge, 2001; BMMRO unpublished data). A second sighting of Gervais' beaked whales occurred off Andros Island in the Bahamas in October 2007, during which the first recordings of free-ranging Gervais beaked whales, reported here, were collected.

II. METHODS

A. Visual

Recordings were made during a visual line transect and photo-identification survey for beaked whales in the northern Bahamas conducted in October 2007 from a 26 m converted shrimp trawler. While surveying, the research vessel followed a pre-determined transect at a speed of approximately 8 kn. Survey effort was restricted to Beaufort sea state 4 or less and most tracks were completed in Beaufort sea state 3 or less. During visual surveys three observers searched for cetaceans from an observation platform at a height of 7 m. Two observers scanned from 90° to ahead on each side of the vessel using 25 × 150 Big-Eye binoculars while a third searched with naked eye and 7 × 50 binoculars. Ranges to sightings were measured using reticules in the binocular eye pieces to measure angle of dip from the horizon. Information on vessel track, survey effort, and environmental conditions (sea state, swell, visibility, wind speed, etc.) were collected using the IFAW LOGGER software (www.ifaw.org/sotw). Effort and environment data were entered every 30 min or when conditions changed.

Once a beaked whale was sighted, the line transect survey was suspended and a 5.5 m rigid hulled inflatable boat (RHIB) was launched in order to make close approaches for photo-identification and biopsy. During these periods, the main vessel remained stationary or moved slowly to stay in the vicinity of the RHIB and whales. A team of four observers remained on the main vessel, two of them continuing to search with Big-Eye binoculars while the others monitored the acoustic data collection and assisted with visual data logging and communications.

B. Acoustic data collection

Acoustic recordings were made continuously through the whole survey using a 400 m long towed hydrophone array (Seiche Measurement UK Ltd). The array consisted of four hydrophone elements arranged as two pairs. Hydrophone pairs were located at 200 and 400 m and the spacing between elements within pairs was 3 m. Each hydrophone was a spherical ceramic connected to a 35 dB preamplifier with a high pass filter configured to be -3 dB at 2 kHz. Hydrophone sensitivity was approximately -165 dB re 1 V/1 μPa at 40 kHz and its response was approximately flat from 2 to 200 kHz. Signals from the hydrophone were recorded using an RME Fireface 800 sound card (Audio AG, Haimhausen, Germany) sampling at 192 kHz. The effective recording bandwidth was therefore from 2 to over 90 kHz. Recordings were made using IFAW LOGGER software and written to disk as four channel, 16 bit wav files. Recordings were made continuously whenever the hydrophone was deployed—both while on transect and during photo-identification periods.

C. Ancillary data

Depth sensors were incorporated into the array close to each hydrophone pair. Hydrophone depth and Global Positioning System (GPS) data were logged every 10 s by the LOGGER software.

D. Acoustic analysis

Clicks were detected offline using click detector modules in the PAMGUARD software (www.pamguard.org, Gillespie *et al.*, 2008). The PAMGUARD click detector was configured to first filter the data using a high pass second order Butterworth filter with a corner frequency at 4 kHz to remove low frequency noise. Data were then passed through a 25–40 kHz fourth order band-pass filter. The output of this band-pass filter then went to a threshold trigger to select sounds with significant energy (>8 dB above background noise) in the 25–40 kHz band. In the event of a trigger, short sound clips (2–3 ms) were made, using data from the output of the first filter. This allowed the broader band data (4–90 kHz) to be used for the next stage of the analysis, classification.

During the Gervais' beaked whale encounter, the hydrophone was hanging near vertical in the water. Although clicks were detected on all hydrophones, only data from the two hydrophones furthest from the boat, which had a better signal to noise ratio, were analyzed. Angles to detected clicks, relative to the array, were calculated using time of arrival differences between the two hydrophones.

Click files were viewed with the RAINBOWCLICK software (Gillespie and Leaper, 1996). This allows the user to easily view groups of clicks on a plot of angles to detected clicks against time and the waveforms and power spectra of individual clicks can be scrutinized. Clicks were selected for the latter stages of the analysis if they satisfied the following criteria:

(1) had significant energy in the 25–50 kHz energy band compared to lower and higher frequencies,
(2) had a waveform resembling that of published data for other beaked whale species, and
(3) formed part of a click train, i.e., they were arriving from the same angle as other clicks and time intervals between clicks appeared regular.

To search for tonal sounds of the type described by Caldwell and Caldwell (1991), all sound files from the encounter were also monitored carefully using high quality headphones (Sennheiser HD 280 pro) while the operator (C Dunn) simultaneously viewed a scrolling spectrogram display of data. An Fast Fourier Transform (FFT) length of 4096 samples was used with a 50% overlap and Hanning window, giving time and frequency resolutions on the spectrogram of approximately 21 ms and 47 Hz, respectively. Only spectral data from 2 to 16 kHz were viewed to improve screen resolution at those frequencies. One channel of the hydrophone pair at 200 m and one channel from the pair at 400 m were selected both for listening and for viewing to maximize the chances of picking up sounds from animals at the surface or at depth.

III. RESULTS

A. Visual

Nine Gervais' beaked whales were encountered between 08:14 and 14:50 local time (12:14–18:50 GMT[1]) on 4 October 2007. Sea conditions at this time were flat calm offering excellent sighting conditions. The initial sighting was made using Big-Eye binoculars at an estimated distance of 5870 m. The survey vessel remained in the vicinity of the animals for over 6 h while the RHIB maneuvered to approach animals during surfacing bouts. The encounter ended when it was judged that sufficient photo-identification and biopsy data had been collected.

The assemblage was encountered as three sub-groups consisting of three, two, and four individuals, including adults and sub-adults, with no calves noted. Sighting times and distances to the three groups are shown in Fig. 1. Sub-group A remained in the immediate area of the research vessel until 11:50 and was then not seen thereafter, sub-group B left the area during the encounter and was last seen at 10:21, and sub-group C appeared at 13:18. Photo-identification showed that the animals observed in the three sub-groups were different individuals. Sub-group A completed a number of short dives that were visually recorded, with the shortest duration being 9 min, the longest 28 min, and the mode being 18 min.

The species identification was based on a combination of visual cues as well as by genetic analysis of a biopsy sample. This was analyzed using standard extraction and genetic sequencing procedures, and the resulting sequence was checked against three reference libraries for identification consistency (personal communication, K. Robertson, SWFSC Molecular Genetics Laboratory, La Jolla, CA).

No cetaceans were sighted that morning during an hour of surveying prior to the Gervais' beaked whale encounter

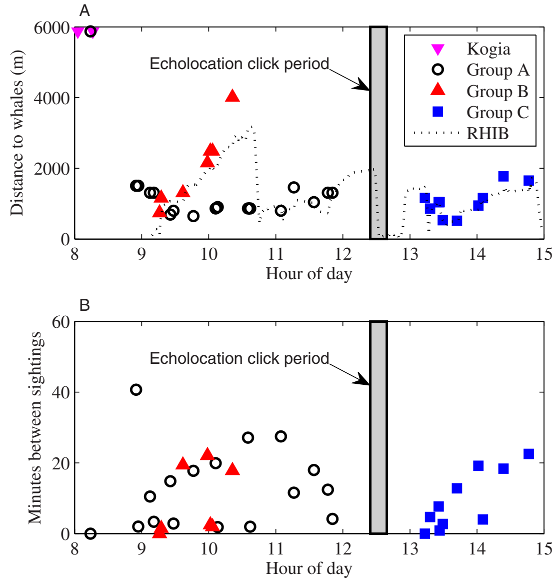


FIG. 1. (Color online) (a) Distances from the ship to sightings of Kogia, the three groups of Gervais' beaked whales, and to the RHIB and (b) time interval between visual sightings of beaked whales.

apart from an unknown *Kogia* species sighted at 08:02 and a dwarf sperm whale (*Kogia sima*) at 08:16, both close to the start of the Gervais' beaked whale encounter. Both *Kogia brevicips* and *Kogia Sima* are known to produce clicks at frequencies well above 100 kHz, similar to those of porpoises (Madsen *et al.*, 2005; V. Janik, personal communication) and therefore would not have been detected by the equipment used in this study. Both were approximately 5.9 km from the vessel and were 6 and 2 km from the Gervais' beaked whales, respectively. The *Kogia* were not re-sighted and no other species were seen until well after the encounter when a group of Cuvier's beaked whales were spotted at 16:19 at a distance of approximately 15 km from the Gervais' beaked whales. The sea conditions were calm and sighting conditions excellent throughout the day.

Figure 1(a) shows ranges (based on reticule measurements) between the recording vessel and whales observed at the surface during the encounter. The sightings immediately before and after the acoustic contact (see below) were at distances of 1309 and 1160 m from the vessel. Figure 1(b) shows times between sightings. For most of the encounter, sightings were quite regular, with one of the sub-groups sighted at the surface at least once every 10–15 min. However, a single long period with no whale sightings occurred between 11:50 and 13:18.

B. Acoustic

The hydrophone was deployed and recordings were made continuously from 09:00 until the end of the encounter. For most of the encounter, from 09:20 until 14:52, the vessel was stationary, allowing the hydrophone to hang near vertically in the water column. Depth sensors close to the front and rear pairs of hydrophones read 188 and 384 m, respectively, throughout this period.

During the entire encounter a total of 124 beaked whale clicks were detected on the lowest hydrophone pair. All of these detections occurred within a short time window between 12:24:30 and 12:39:19 ($\sim$15 min) which was within the longest (88 min) interval between sightings, the first click being detected 34 min after the most recent sighting (Fig. 1). Angles to the clicks were more or less constant during the encounter, varying by no more than 6° and at no time were clicks detected simultaneously at different angles, therefore giving no indication that the clicks came from more than one individual. Angles to clicks all indicated that the animal was deeper than the hydrophone, i.e., at a depth of over 384 m.

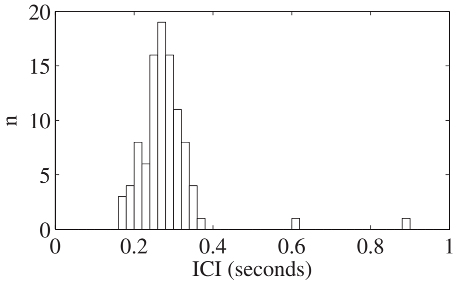


FIG. 2. Histogram of ICIs <1 s.

The clicks were detected in short sequences followed by gaps which varied in time from a few seconds to 336 s. Intervals between clicks within sequences are shown in Fig. 2. The mean ICI for intervals <0.5 s (the dominant peak in Fig. 2) was 0.27 s. There were no regular clicks with an ICI greater than 0.4 s and it is probable that the ICIs in Fig. 2 at 0.6 and 0.9 s are due to one and two missed clicks, respectively. A typical click waveform, power spectrum, and time-frequency (Wigner) distribution are shown in Fig. 3. For comparative purposes, the mean power spectrum for all of the Gervais' beaked whale clicks is shown with similar averaged spectra for clicks from Blainville's and Cuvier's beaked whales encountered during the same cruise and analyzed in the same way in Fig. 4. The click waveform is similar to that of Cuvier's and Blainville's beaked whales, having a duration of about 200 μs and energy concentrated

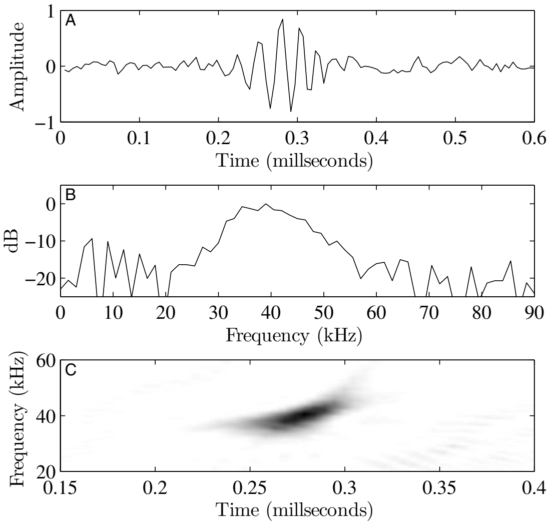


FIG. 3. (a) Waveform, (b) normalized power spectrum, and (c) time-frequency (Wigner) distribution for a typical detected click from a Gervais' beaked whale.

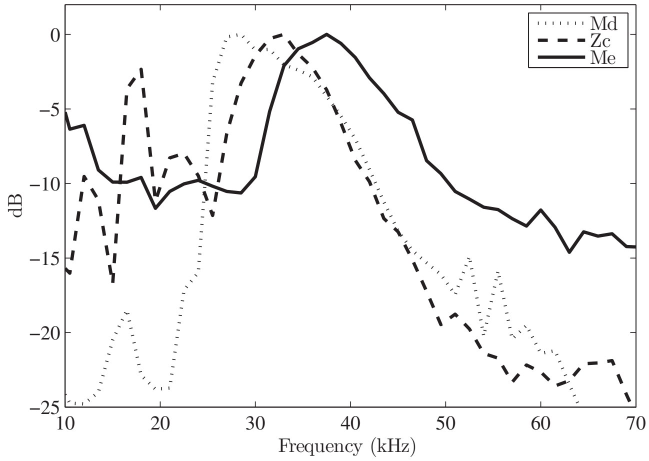


FIG. 4. Averaged power spectra for three beaked whale species (Md =Blainville's beaked whale, Zc=Cuvier's beaked whale, and Me=Gervais' beaked whale).

in the 30–50 kHz band. The spectral data indicated that Gervais' beaked whale clicks were at a slightly higher frequency than other species. Like these other species the Gervais' beaked whale click appeared to sweep up in frequency.

A single tonal sound was both detected on the spectrogram and heard with headphones. The sound was only detected on the hydrophone at 200 m depth. It was approximately 30 ms in length and was at a very narrow band frequency centered at 6 kHz. From a time of arrival measurement of the signal on the two hydrophones at 200 m, it was found that the sound was coming from the direction of the vessel and it is our belief that it is mechanical in nature.

IV. DISCUSSION

Data from DTAGs (Johnson *et al.*, 2004, 2006; Madsen *et al.*, 2005) show that other beaked whale species only echolocate when undertaking deep foraging dives. Our data indicate that the same may be true for Gervais' beaked whales since they were only detected acoustically during the longest interval between visual sightings of whales. However, since the individuals observed after the bout of clicking were different from those observed before it, we cannot be certain that the animals were engaged in a long foraging dive and had not simply departed from or arrived into the area.

The potential for underwater recordings to be contaminated by sounds from unseen animals other than the focal species is always a concern especially with species like beaked whales which seem to vocalize mainly at depth when they cannot be directly observed at the surface. In this case several pieces of information give us confidence that the clicks we detected were those of Gervais' beaked whales.

(1) The low density of cetaceans in the area allied to the good sighting conditions, high level of visual effort, and lack of detections of any other cetacean within several hours of the acoustic detection.
(2) The substantial spatial and temporal gap between this encounter and sightings of any other beaked whales.
(3) The fact that sound production appeared to be synchronized with a long period without visual sightings of the targeted group, as has been documented during deep foraging dives for other beaked whale species.
(4) The observation that the sounds have the same general characteristics of those of other beaked whale species recorded by this team using the same equipment, yet demonstrate sufficient differences to be distinctive.

Sighting and photo-identification records indicate that there were three individuals present in the surface encounter before the clicks were detected and four different animals during the surface encounter immediately afterward. The number of clicks detected (124) was considerably less than the 4000–5000 clicks per dive reported for Cuvier's and Blainville's beaked whales (Madsen *et al.*, 2005). Zimmer *et al.* (2005) showed that beaked whale clicks are highly directional and are therefore only likely to be detected if an animal happens to be orientated toward the hydrophone. Zimmer *et al.* (2008) modeled detection probability for Cuvier's beaked whales and showed that the probability of detecting any individual click with a remote hydrophone is low, but that the overall probability of detecting at least some clicks from a group is much higher. Our results are consistent with these findings.

The only previous recording of this species of which we are aware was made from a captive animal in a small tank by Caldwell and Caldwell (1991). The recording equipment used was not sensitive to the higher frequency sounds recorded in this study and no sounds of the type that they recorded could be found in our recordings. Caldwell and Caldwell (1991) described the sounds they recorded as having a high amplitude. One possible explanation for the clicks they reported is that the high frequency clicks were saturating their recording equipment, resulting in a broad band distortion that would have been within the bandwidth of their system.

It is not known whether the surfacing and diving behavior observed here is typical for all Gervais' beaked whales and they are certainly insufficient for drawing many conclusions. Clicks were only heard once during a period of over 6 h spent in the proximity of several animals. If this behavior is typical, then the probability of acoustic detection during an acoustic survey or for mitigation is likely to be low. This is not necessarily a severe problem for survey applications where additional survey effort can be expended. For mitigation applications though this would suggest that only a small degree of risk reduction could be provided by PAM.

Clicks were detected in a number of short sequences, followed by gaps of varying durations. All of the clicks were relatively quiet and were very close to the limits of detectability with the hardware used. It is therefore highly likely that many clicks were missed and that clicking was much more continuous than indicated by these data.

The waveform and spectrum of Gervais' beaked whale clicks are similar to those of Cuvier's and Blainville's beaked whales, but are at a slightly higher frequency. It would be unwise to draw too many conclusions from a recording of what may be a single animal, but if this difference in frequency is genuinely characteristic of the species, it may in the future be possible to tell some beaked whale species apart based on acoustic data alone.

ACKNOWLEDGMENTS

We would like to thank the other field scientists: Edward Adderley, Kiya Gornik, Leigh Hickmott, and Olivia Patterson, as well as Captain Scott Smith and the crew of the *Dolphin Dream*. The field effort was supported by a research grant from the U.S. Office of Naval Research (Grant No. N000140710120). Acoustic analysis was funded by the UK Defense Science and Technology Laboratory Joint Grants Scheme and the Natural Environment Research Council. LOGGER and RAINBOW CLICK softwares were made available courtesy of the International Fund for Animal Welfare (IFAW). This work was conducted under the Bahamas Department of Marine Resources Research Permit No. 1, issued under authorization of the Bahamas Marine Mammal Protection Act (2005) and under guidance from BMMRO's Institutional Animal Care and Use Committee.

[1] All subsequent times in the paper are local Bahamas time=GMT−4.

Balcomb, K. C. (**1981**). "Ziphiid whales from the Bahamas," Bahamas Naturalist, 19–22.

Balcomb, K. C., and Claridge, D. E. (**2001**). "A mass stranding of cetaceans caused by naval sonar in the Bahamas," Bahamas Journal of Science **2**, 2–12.

Barlow, J., and Gisiner, R. (**2006**). "Mitigating, monitoring and assessing the effects of anthropogenic sound on beaked whales," J. Cetacean Res. Manage. **7**, 239–249.

Caldwell, D., and Caldwell, C. (**1991**). "A note describing sounds recorded from 2 cetacean species, Kogia breviceps and Mesoplodon europaeus stranded in northeastern Florida," Marine Mammal Strandings in the United States: Proceedings of the Second Marine Mammal Stranding Workshop, Miami, FL, **1987**, United States Department of Commerce, National Oceanic and Atmospheric Administration, Vol. **98**, pp. 151–154.

Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., D'Amico, A., D'Spain, G., Fernandez, A., Finneran, J., Gentry, R., Gerth, W., Gulland, F., Hildebrand, J., Houser, D., Hullar, T., Jepson, P. D., Ketten, D., MacLeod, C. D., Miller, P., Moore, S., Mountain, D. C., Palka, D., Ponganis, P., Rommel, S., Rowles, T., Taylor, B., Tyack, P., Wartzok, D., Gisiner, R., Mead, J., and Benner, L. (**2006**). "Understanding the impacts of anthropogenic sound on beaked whales," J. Cetacean Res. Manage. **7**, 177–187.

Dalebout, M. L., Baker, C. S., Mead, J. G., Cockcroft, V. G., and Yamada, T. K. (**2004**). "A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae," J. Hered. **95**, 459–473.

Dawson, S., Barlow, J., and Ljungblad, D. (**1998**). "Sounds recorded from Baird's beaked whale, Berardius Bairdil," Marine Mammal Sci. **14**, 335–344.

Evans, D. L., England, G. R., Lautenbacher, C. C., Hogarth, W. T., Livingstone, S. M., and Johnson, H. T. (**2001**). "Joint interim report Bahamas marine mammal stranding event of 15–16 March 2000," NOAA and Department of the Navy, Vol. **59**.

Fernandez, A., Edwards, J. F., Rodriguez, F., Espinosa de los Monteros, A., 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.

Frantzis, A. (**1998**). "Does acoustic testing strand whales?," Nature (London) **392**, 29.

Gillespie, D., Gordon, J., Mchugh, R., Mclaren, D., Mellinger, D., Redmond, P., Thode, A., Trinder, P., and Deng, X. Y. (**2008**). "PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans," Proceedings of the Institute of Acoustics, Vol. **30**.

Gillespie, D., and Leaper, R. (**1996**). "Detection of sperm whale (Physeter macrocephalus) clicks and discrimination of individual vocalizations," *Eur. Res. Cetaceans* [Abstracts], pp. 10:87–91.

Hooker, S. K., and Whitehead, H. (**2002**). "Click characteristics of northern bottlenose whales (hyperoodon ampullatus)," Marine Mammal Sci. **18**, 69–80.

Jefferson, T., Webber, M., and Pitman, R. (**2008**). *Marine Mammals of the World: A Comprehensive Guide to Their Identification* (Elsevier, London, UK).

Jepson, P. D., Arbelo, M., Deaville, R., Patterson, I. A. P., Castro, P., Baker, J. R., Degollada, E., Ross, H. M., Herraez, P., and Pocknell, A. M. (**2003**). "Gas-bubble lesions in stranded cetaceans," Nature (London) **425**, 575–576.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., and Tyack, P. L. (**2006**). "Foraging Blainville's beaked whales (Mesoplodon densirostris) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038.

Johnson, M. P., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.

MacLeod, C. D., Perrin, W. F., Pitman, R., Barlow, J., Ballance, L., D'Amico, A., Gerrodette, T., Joyce, G., Mullin, K. D., and Palka, D. L. (**2006**). "Known and inferred distributions of beaked whale species (Cetacea: Ziphiidae)," J. Cetacean Res. Manage. **7**, 271–286.

Madsen, P. T., Johnson, M., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (Mesoplodon densirostris)," J. Exp. Biol. **208**, 181–194.

Norman, S. A., and Mead, J. G. (**2001**). "Mesoplodon europaeus," Mamm. Species **688**, 1–5.

Podesta, M., Cagnolaro, L., and Cozzi, B. (**2005**). "First record of a stranded Gervais' beaked whale, Mesoplodon europaeus (Gervais, 1855), in the Mediterranean waters," Atti della Società italiana di scienze naturali e del museo civico di storia naturale di Milano **146**, 109–116.

Rogers, T. L., and Brown, S. M. (**1999**). "Acoustic observations of Arnoux's beaked whale (berardius arnuxii) off Kemp Land, antarctica," Marine Mammal Sci. **15**, 192–198.

Tyack, P. L., Johnson, M., Aguilar de Soto, N., Sturlese, A., and Madsen, P. T. (**2006**). "Extreme diving of beaked whales," J. Exp. Biol. **209**, 4238–4253.

Waring, G. T., Josephson, E., Fairfield, C. P., and Maze-Foley, K. (**2007**). "US Atlantic and Gulf of Mexico marine mammal stock assessments—2006," NOAA Tech Memo NMFS NE **201**, 378.

Zimmer, W., Harwood, J., Tyack, P., Johnson, M., and Madsen, P. (**2008**). "Passive acoustic detection of deep diving beaked whales," J. Acoust. Soc. Am. **124**, 2823–2832.

Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (Ziphius cavirostris)," J. Acoust. Soc. Am. **117**, 3919–3927.

The acoustics and acoustic behavior of the California spiny lobster (*Panulirus interruptus*)

S. N. Patek,[a] L. E. Shipp, and E. R. Staaterman
Department of Integrative Biology, University of California, Berkeley, California 94720-3140

(Received 15 October 2008; revised 9 February 2009; accepted 18 February 2009)

Numerous animals produce sounds during interactions with potential predators, yet little is known about the acoustics of these sounds, especially in marine environments. California spiny lobsters (*Panulirus interruptus*) produce pulsatile rasps when interacting with potential predators. They generate sound using frictional structures located at the base of each antenna. This study probes three issues—the effect of body size on signal features, behavioral modification of sound features, and the influence of the ambient environment on the signal. Body size and file length were positively correlated, and larger animals produced lower pulse rate rasps. Ambient noise levels (149.3 dB re 1 μPa) acoustically obscured many rasps (150.4 ± 2.0 dB re 1 μPa) at distances from 0.9–1.4 m. Significantly higher numbers of pulses, pulse rate, and rasp duration were produced in rasps generated with two antennae compared to rasps produced with only one antenna. Strong periodic resonances were measured in tank-recorded rasps, whereas field-recorded rasps had little frequency structure. Spiny lobster rasps exhibit flexibility in acoustic signal features, but their propagation is constrained, perhaps beneficially, by the noisy marine environment. Examining the connections between behavior, environment, and acoustics is critical for understanding this fundamental type of animal communication. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097760]

PACS number(s): 43.80.Ka [WWA] Pages: 3434–3443

I. INTRODUCTION

Surprisingly few studies have examined the acoustics of antipredator sounds in the marine environment even though these sounds have been observed across a wide range of taxa in an array of marine habitats. For example, when interacting with intruders or potential predators, nephropid lobsters vibrate antennal muscles (Mendelson, 1969; Henninger and Watson, 2005), ocypodid and pagurid crabs stridulate (Guinot-Dumortier and Dumortier, 1960; Field *et al.*, 1987), astacid crayfish squeak with their abdomen (Sandeman and Wilkens, 1982), mantis shrimp rumble using muscles attached to their carapace (Patek and Caldwell, 2006), and fish produce a myriad of sounds using mechanisms from muscle contractions to stridulatory jaws (Fish and Mowbray, 1970). Antipredator sounds have been shown experimentally to deter predators (Alexander, 1958; Masters, 1979; Lewis and Cane, 1990; Sargent, 1990) and function through a variety of mechanisms, including the effects of startle, warning or predator memory enhancement (Gittleman and Harvey, 1980; Gamberale and Tullberg, 1996; Speed, 2000; Sherratt and Beatty, 2003; Ruxton *et al.*, 2004; Caro, 2005).

There are two key factors when examining antipredator sounds: signal propagation through the environment and the effect of, or information contained in, the signal features. In terms of signal propagation, the emission of sound has the potential to attract other predators to the scene; this may make the situation for the prey item even more dangerous or, alternatively, it may increase conflict between the predators and thereby increase the odds that the prey item escapes (reviewed in Chivers *et al.*, 1996). In either case, the propagation of antipredator signals through the environment—given the ambient background noise and the effects of the physical structure of the habitat—is an important factor in how the signal functions.

Antipredator signal features are relevant to their function and performance in deterring predators. Most antipredator signals, whether acoustic, chemical, or visual, capitalize on being noxious or generally startling to either deter the predator or trigger a predator's memory that the prey item is not palatable (Edmunds, 1974; Ruxton *et al.*, 2004). Furthermore, it is generally advantageous for the prey to appear as threatening as possible by exaggerating size (e.g., eye spots). Although these general principles of antipredator signal design are widely accepted, the features of acoustic antipredator signals are rarely characterized.

Most spiny lobster species (Palinuridae), including the California spiny lobster (*Panulirus interruptus*), generate antipredator sounds called "rasps" (Parker, 1878; 1883; Lindberg, 1955; Moulton, 1957; George and Main, 1967; Meyer-Rochow and Penrose, 1974; Smale, 1974; Meyer-Rochow and Penrose, 1976; Mulligan and Fischer, 1977; Patek, 2001; Patek, 2002; Patek and Oakley, 2003; Latha *et al.*, 2005; Patek *et al.*, 2006). Documented for over a thousand years (Athenaeus, 300), these sounds are produced when spiny lobsters are handled by potential predators. The rasps function to deter predators; spiny lobsters that have been silenced (i.e., the sound-generating apparatus has been disabled) are attacked more frequently and with greater success than spiny lobsters with intact sound-producing structures (Bouwma and Herrnkind, 2004; Bouwma, 2006; Bouwma and Herrnkind, 2007).

[a] Author to whom correspondence should be addressed. Electronic mail: patek@berkeley.edu

0001-4966/2009/125(5)/3434/10/$25.00
© 2009 Acoustical Society of America

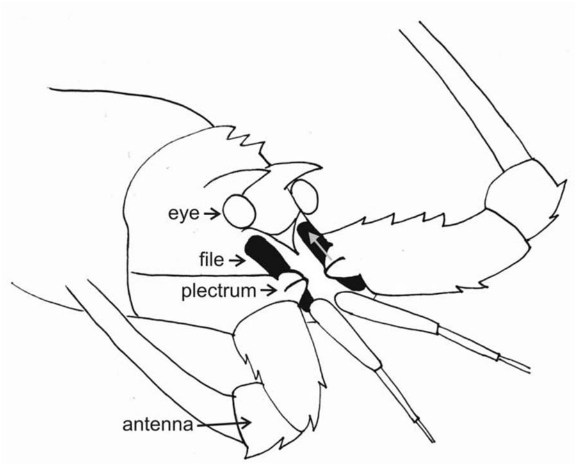


FIG. 1. The sound-producing anatomy of a California spiny lobster (*Panulirus interruptus*). A plectrum is found at the base of each antenna and rubs over a file beneath each eye. Sound is produced when the plectrum slides posteriorly (arrow) over the file. Adapted from (Summers, 2001).

The stick-slip frictional mechanism of spiny lobster sound production is unusual in the biological world, and the paired structures potentially yield flexibility in signal features. Analogous to bowed-stringed instruments, spiny lobsters produce pulses of sound through stick-slip frictional interactions between the plectrum and file surfaces such that the plectrum sticks and slips due to friction as it is pulled posteriorly over the file; a pulse of sound is produced during each "slip" (Patek, 2001, 2002; Patek and Baio, 2007). The plectrums are located at the base of each antenna and traverse the oblong files located on each side of the antennular plate (Fig. 1). Each rasp sound is produced when the plectrum is pulled posteriorly and generates a series of sound pulses as it sticks and slips over the surface of the file (Fig. 2). Because there is a pair of plectrum/file units, sounds can be produced by rubbing only one plectrum over the file, both plectrums in series, or both plectrums concurrently. This flexibility in the deployment of the sound-producing structures potentially offers spiny lobsters additional variation in the range of signal features.

While the above studies have addressed the functional morphology, evolutionary history, and behavioral context of sound production, the acoustics of these sounds have been examined exclusively in laboratory settings and, thus, the amplitude and frequency structure of these sounds in nature are not currently known. Furthermore, little is known about the influence of using paired structures on the signal features and the scaling of signal features with body size in adult lobsters (Meyer-Rochow and Penrose, 1976; Patek and Oakley, 2003). Examination of the sounds in the laboratory and field offers important insights into the use of these signals by spiny lobsters and, more generally, the role and function of antipredator sounds in the marine environment. In this study, as in all previous analyses of spiny lobster antipredator sounds, we measure the rasps generated during handling, simulating the lobster's experience once a predator has successfully caught the prey and is attempting to process it; this leaves open the possibility that a different suite of acoustic signals are used during predator approach or for signaling to distant predators, although no such sounds have been documented to date in spiny lobsters.

Thus, the goals of this study were to examine the acoustics and acoustic behavior of the California spiny lobster (*Panulirus interruptus*) from the following three perspectives: (1) *Body size and signal features*: Which acoustic parameters are correlated with body size? Do spiny lobsters vary rasp duration by increasing number of pulses or decreasing pulse rate? (2) *Plectrum activation and rasp variation*: Are single, sequential and concurrent plectrum activation patterns correlated with specific rasp features, such as greater rasp duration, higher pulse rate or greater number of pulses, and particular behaviors, such as tail flipping or leg movements? (3) *Rasps and their acoustic environment*: How does the acoustic environment influence rasp signal features, specifically when comparing recordings made in a tank *versus* in the field?

II. METHODS

A. Animal collection and care

California spiny lobsters, *Panulirus interruptus* (Crustacea, Decapoda, Palinuridae), were collected at the University of Southern California, Wrigley Institute for Environmental Studies (WIES, Santa Catalina Island, CA,) in baited lobster

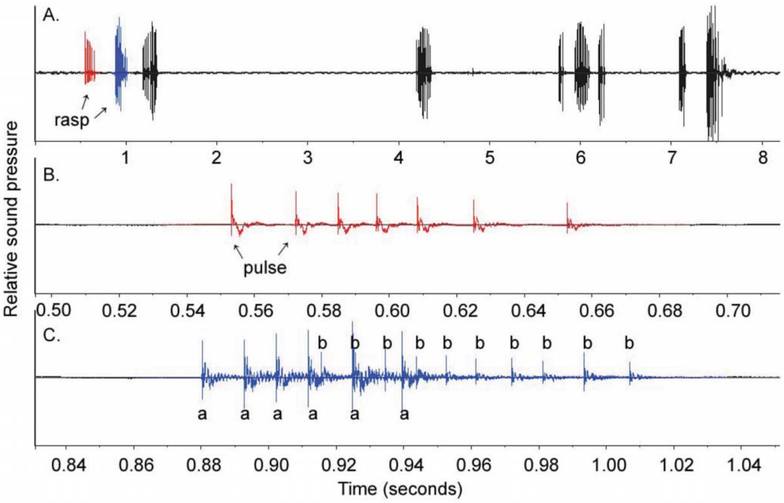


FIG. 2. (Color online) Spiny lobsters typically produce a series of rasps (A) each consisting of a series of pulses (B). In a rasp produced by a single plectrum (the first rasp in panel A), seven pulses are visible (B). The second rasp in A was generated with two plectrums activated concurrently (C), beginning with the first plectrum producing a series of pulses labeled "a" and the second plectrum generating the overlapping series of pulses labeled "b." Series a and series b are distinguishable by differences in both amplitude and temporal spacing.

traps or by hand (CA Department of Fish and Game Permit No. SC-5751). One small lobster was borrowed from the WIES "touch tank." Lobsters that were brought back to the laboratory were maintained in cylindrical holding tanks (1.5 m diameter, 0.8 m height) with a continuous supply of sea water (14–16 °C). They were fed bait fish daily.

Experiments were conducted during three different field seasons, and the lobsters were captured shortly before each set of experiments. In the first field season (2005), we conducted the temporal acoustic analysis experiments in which we recorded 24 individuals (2 males and 22 females; 44–102 mm carapace length; 14–15 °C water temperature). In the second field season (2006), we conducted audio-video experiments of acoustic behavior and comparisons of rasp acoustics in tank versus field conditions. In these experiments, we recorded 20 individuals (6 males and 14 females; 50–113 mm carapace length; 8–18 mm file length; 14–15 °C water temperature). In the third field season (2008), we recorded 13 more individuals (12 females, 1 male; 65–93 mm carapace length; 19.8 °C water temperature) in the field and measured pressure levels of the rasps and ambient field environment. The specific sample sizes used in each experiment are detailed below.

B. Temporal acoustic analyses

We tested for the presence of correlations between body size and temporal components of the rasp, including rasp duration (s), number of pulses (total number of pulses per rasp), and average pulse rate (pulses s^{-1}: number of pulses per rasp divided by rasp duration) (Fig. 2). A hydrophone (1 Hz–170 kHz, TC4013, Reson, Slangerup, Denmark) was connected to a band-pass filter (high-pass: 10 Hz, low-pass: 15 kHz; 1 Hz–1 MHz VP2000 voltage preamplifier, Reson, Slangerup, Denmark) and a digital audio recorder (48 kHz sample rate, maximum 20 kHz frequency response [−0.5 dB], PMD670, Marantz, NJ). Individuals were hand-held at variable depths in a fiberglass tank (1.5 m diameter, 0.8 m height) with the hydrophone suspended approximately 60 cm from the anterior end of the lobster.

Rasp waveforms were measured using acoustic software (RAVEN 1.2.1, Cornell Laboratory of Ornithology, NY). We defined rasps as consisting of at least two pulses that occurred within 45 ms of each other. Sometimes, a single pulse of sound was produced in isolation; these pulses were not included in the analyses. Pulse duration was not measured, because previous studies have shown that tank reverberations obscure the ending time of each pulse (Patek and Baio, 2007). When the rasps were difficult to resolve against excessive background noise, they were omitted from the analyses.

In many recordings, the sounds generated by the two plectrums were distinguishable from each other, either temporally or through amplitude differences. We split the dataset into rasp waveforms unambiguously produced by one plectrum only and compared them to rasp waveforms clearly showing activity from two plectrums (Fig. 2); when this distinction was not clear, the rasps were not included in this particular analysis. Using these distinct patterns, we documented three modes of plectrum activation: (1) "single plectrum" in which only one plectrum was used to generate one rasp, (2) "sequential plectrums" in which the same plectrum was used repeatedly or two plectrums were used in sequence to generate multiple rasps, and (3) "concurrent plectrums" in which two plectrums concurrently generated one or more rasps (see Fig. 2 and further explanation in Sec. III). We limited this dataset to individuals that produced at least three rasps each of single plectrum and dual plectrum activation.

We analyzed the relationships among these acoustic variables and between these variables and body size. Least-squares linear regressions were used to examine the correlation between carapace length and mean values for acoustic features across individuals. A general linear model [analysis of covariance (ANCOVA)] was applied to examine the effects of plectrum activation on the temporal acoustic features, as well as the effects of individual, and individual by plectrum usage on the resulting correlations. Similarly, an ANCOVA was used when examining the correlation between pulse number and rasp duration within and across individuals. Statistics were performed with JMP v. 5.0.1 software.

C. Comparison of acoustic frequencies in tank *versus* field conditions

The acoustic frequencies of *P. interruptus* rasps were compared between field and tank recordings. Each individual lobster was recorded in the field and then recorded in the tank so that the spectral characteristics could be compared both within and among individual lobsters.

Field recordings were taken in 7.3 m water (14 °C) with the lobsters hand-held at 42 cm depth. The distance of the hydrophone from the anterior end of the focal lobster was held at a constant 31 cm in the field and ranged from 31–66 cm in the tank. The tank recordings were performed in a cylindrical, fiberglass tank (1.5 m diameter, 0.8 m height) at 15 °C. Calibrated recordings were taken with a hydrophone (0.1 Hz–10 kHz ± 1.5 dB, sensitivity: −206.1 dB ± 0.25 dB re 1 V/μPa, Type 8104 hydrophone, Brüel and Kjaer, Nærum, Denmark) and amplifier (set at high-pass filter 2 Hz and low-pass filter 10 kHz; 0.2 Hz–200 kHz, Type 2635 charge amplifier, Brüel and Kjaer, Nærum, Denmark) which were connected to a digital data acquisition system (50 kHz sample rate, NIDAQ 6062E PCMCIA data acquisition card, National Instruments, TX; custom data acquisition software, MATLAB, The Mathworks, Natick, MA). Using a custom MATLAB program, the data were converted to ".wav" files by scaling the voltage amplitude by a factor of 0.1 and running a 20 Hz high-pass Butterworth filter.

The dominant frequencies (the two frequencies with greatest acoustic power) were identified for each rasp and compared to ambient background noise in each recording (RAVEN v. 1.2.1 and 1.3, Cornell Laboratory of Ornithology, Ithaca, NY). Temporal measurements were calculated from the acoustic waveforms; frequency analyses were measured from power spectra using a discrete Fourier transform (settings: Hanning window, 2000 sample window size, 3 dB filter bandwidth at 36 Hz resolution).

We examined correlations between acoustic features and body size using least-squares linear regressions. We performed a t-test to examine whether recording conditions significantly affected the dominant frequencies. Statistics were performed using JMP software (v. 7.0, SAS Institute, Inc., NC).

D. Rasps in the field environment

The pressure levels of the rasps and background noise were measured in the field. Lobsters were held by hand in 7.3 m water at 45 cm depth with the hydrophone positioned at 97 cm depth. Thus, the effective diagonal distances between the lobster and hydrophone were 0.9, 1.1, 1.3, and 1.5 m. The equipment and settings were the same as in Sec. II C. Absolute average power (dB) was calculated by converting RAVEN software's dimensionless units to pascals using the calibration provided by the hydrophone and amplifier manufacturer and the conversion factors provided by RAVEN software (version 1.4, Hanning window, 2000 sample window size, 3 dB filter bandwidth at 36 Hz resolution). These calibration methods are explained in the RAVEN software support documents and are also available upon request from the authors. The average power (dB) was calculated relative to 1 μPa (the standard for aquatic measurements) and also calculated relative to the baseline noise level measured in each recording.

E. Audio-video analyses of acoustic behavior

In order to test whether rasp features and plectrum activation were correlated with specific behaviors, we used synchronous audio and video to record spiny lobsters producing rasps. Each individual was held approximately 36 cm deep, 1.5 m from the camera, and recorded until it produced 5–10 rasps (20–15 000 Hz; HTI-94-SSQ hydrophone, High Tech, Inc., Gulfport, MS; Sony DCR-VX2100 Handycam video camera, Tokyo, Japan; Amphibico VLAL0010 underwater housing, Montreal, Canada). Rasps were elicited by holding and gently squeezing or tickling the lobster.

We identified behavioral units typically associated with escape or arousal in lobsters (Atema and Cobb, 1980). The two most consistent and identifiable behaviors were leg movements and tail flips. Tail flips are an escape response in which the tail is rapidly tucked under the body causing the animal to rapidly jet backwards. Leg movements were noted if they were vigorous and continuous (as distinguished from the slow or small movements associated with resting behavior).

Sound production and behavioral units were counted and binned over 10 s intervals. The onset of each 10 s bin occurred when the spiny lobster first started to produce rasps. Sound production and behavior were measured for 10 s in all individuals with the exception of one individual for which only 6 s were recorded. We tested whether including this individual affected the results by running the analyses with and without it. We logged the time at which each behavioral unit and sound occurred and noted the identity of the plectrum(s) (right plectrum, left plectrum, or both) producing the sound. We then calculated the rate of rasp production (the number of rasps divided by the 10 s bin during which they occurred) and the proportion of rasps produced by a single plectrum or both plectrums (including both sequential and concurrent movement) out of the total number of rasps produced during the 10 second time period.

TABLE I. Temporal features of rasps. Sample size was 19 individuals with 5–21 rasps recorded per individual. The minimum number of pulses in a rasp sequence was set at two pulses; single pulses were not included in the analysis. A one-way analysis of variance tested for differences across individuals. ** indicates $p < 0.0001$.

	Minimum–maximum	Mean ± std. dev.	F-ratio
Pulse rate (pulses s^{-1})	24–192	71 ± 20	**7.9982
Rasp duration (ms)	15–303	108 ± 35	**6.4656
Number of pulses	2–19	7 ± 3	**8.5686

We tested whether the behavioral units (tail flip and antennal movement) were correlated with the rate of rasp production and the number of rasps produced using both plectrums concurrently. These data were not normally distributed (Shapiro-Wilks Goodness-of-Fit test; $p < 0.0001$), therefore the nonparametric Kruskal–Wallis test was used in place of a t-test (JMP 5.0.1, SAS Institute, Inc., NC).

Results are presented as mean ± one standard deviation.

III. RESULTS

A. Acoustic features and body size correlations

The temporal rasp features varied substantially both within and across individuals (Table I). File length (Fig. 1) was positively correlated with carapace length (Fig. 3) ($N = 18$; $R^2 = 0.6178$, $F = 25.8664$, $p = 0.0001$). Carapace length was negatively correlated with mean pulse rate, but not correlated with mean rasp duration or mean number of pulses per rasp (Fig. 3; Table II). Rasp duration was positively correlated with number of pulses (Fig. 4) ($df = 18$; whole model: $R^2 = 0.6998$, $F = 24.7785$, $p < 0.0001$; number of pulses: $F = 181.4677$, $p < 0.0001$; individuals: $F = 12.5963$, $p < 0.0001$). Because pulse rate was calculated using values from rasp duration and pulse number, it was not statistically valid to examine the relationships among the three variables. The rasps produced with two plectrums concurrently had a significantly greater number of pulses than rasps produced by one plectrum alone or two plectrums sequentially; rasp duration and pulse rate were also greater in rasps produced with two plectrums (Table II).

Dominant frequencies were not correlated with body size in the tank nor in the field (Fig. 3) (least-squares linear regression, $p > 0.4$ in all tests). Body size was not correlated with average power when pooled across all recording distances (Fig. 3) ($R^2 = 0.302$, $df = 12$, $F = 4.76$, $p = 0.05$).

B. Acoustic frequencies in the tank and the field

The frequency characteristics and background noise levels of the rasp recordings were different in the field and the

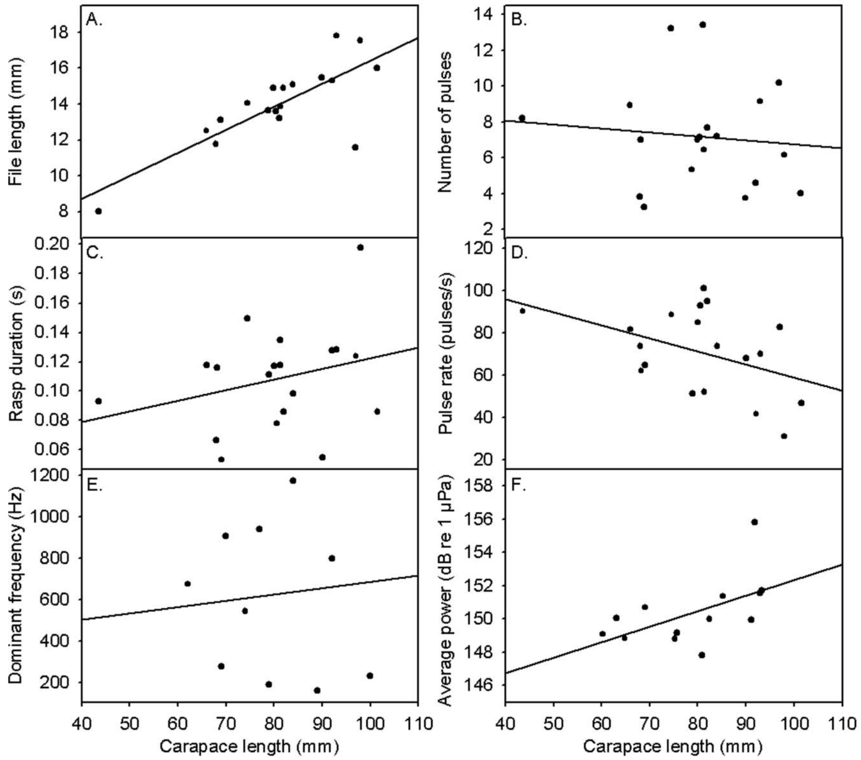


FIG. 3. The relationships between file length (A), temporal signal features [(B)–(D)], spectrographic features [(E)–(F)], and body size. Each data point represents the mean value for an individual lobster.

tank (Table III). The rasps from the field recordings typically had one distinct narrow peak below 500 Hz and another broader peak around 1.5–2 kHz (Fig. 5). Tank recordings lacked this predictable structure and exhibited a pattern of evenly spaced narrow peaks (Fig. 5). The dominant frequencies in the tank and field were significantly different (t-ratio=7.569, $p<0.0001$), but the second most powerful frequency was not significantly different (t-ratio=1.212, $p=0.24$).

C. Rasps in the field acoustic environment

The average power of the rasps was 150.4 ± 2.0 dB re 1 μPa ($N=13$ lobsters; 281 rasps). The average background noise level was 149.3 ± 3.3 dB re 1 μPa ($N=36$ recordings) and the rasps exceeded the background noise level by an average 1.6 dB (range: −7.5 to 9.5 dB) (Table IV; Figs. 6 and 7). 31% of the rasps had power less than the average ambient noise with the majority of the negative decibel ref-

TABLE II. The correlation between temporal rasp features, body size, and plectrum activation. Least-squares linear regressions were used to analyze the relationship between pulse rate, rasp duration, number of pulses, and carapace length. A general linear model (ANCOVA) was used to analyze the correlation between the use of one or both plectrums to generate sound and the temporal features of the rasp. This second analysis was restricted to individuals producing at least three rasps each of single plectrum and double plectrum activation, resulting in a dataset of five individuals. * indicates $p<0.05$; ** indicates $p<0.001$.

	Carapace length $N=19$ individuals	Plectrum activation $df=1,4$	
Pulse rate	$^{*}R^2=0.2170$	Whole model	$^{**}R^2=0.6501$, $F=10.9405$
	$F=4.7110$	plectrum activation	$^{*}F=4.1794$
		Individual	$^{**}F=18.6654$
		plectrum activation × individual	$F=0.2576$
Rasp duration	$R^2=0.0517$	Whole model	$^{**}R^2=0.5289$, $F=6.6105$
	$F=1.0308$	plectrum activation	$^{**}F=13.5364$
		Individual	$^{**}F=9.4418$
		plectrum activation × individual	$F=1.1452$
Number of pulses	$R^2=0.03826$	Whole model	$^{**}R^2=5826$, $F=8.2215$
	$F=0.6763$	plectrum activation	$^{**}F=45.2203$
		Individual	$^{*}F=4.4095$
		plectrum activation × individual	$F=0.\ 1.9512$

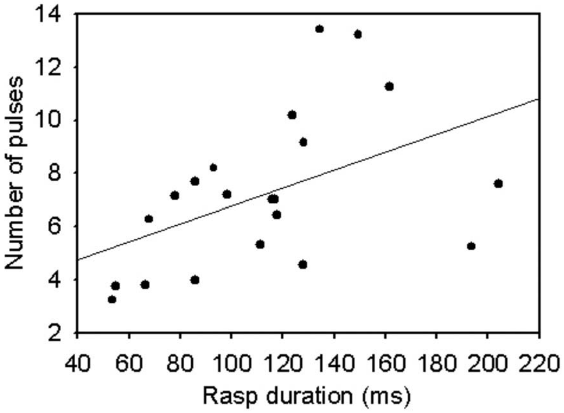


FIG. 4. The number of pulses scales positively and significantly with rasp duration. Each data point represents the mean value for an individual.

erenced to background noise rasps occurring at greater recording distances (Fig. 7). However, when the proportion of rasps below zero dB re background noise was calculated within each individual and then pooled across individuals for each distance, this pattern was less evident and was nonsignificant (least-squares linear regression: $R^2=0.5532$, $df=1,12$, $F=1.905$, $p=0.09$).

D. Audio-video analyses of acoustic behavior

Two datasets were analyzed, one including all data and one excluding a short video sequence. The results were consistent whether or not the short video clip was included in the dataset; thus, the statistical results presented here include all available data. The number of rasps produced by both plectrums concurrently [Fig. 2(c)] was positively correlated with tail-flip behavior (Table V). Regardless of whether rasps were produced with a single plectrum or both, the rate of sound production increased significantly when an individual's legs were moving (Table V).

TABLE III. Comparison of frequencies in the field versus tank recordings and background noise versus rasps. Data are in the following format: mean frequency ± s.d.; tank: $N=13$ individuals (3–22 rasps per individual); field: $N=11$ individuals (5–34 rasps per individual).

	Tank		Field	
	Background noise	Rasp	Background noise	Rasp
Dominant frequency (Hz)	126 ± 322	1794 ± 338	366 ± 706	633 ± 374
Second dominant frequency (Hz)	n.a.	1796 ± 303	n.a.	1590 ± 483

IV. DISCUSSION

The acoustics of the California spiny lobster's rasp were tied to the ambient environment, the individual behavior of the lobsters, and, to a lesser extent, the size of the lobsters. As we discuss below, the interconnections between the rasp characteristics and the environment may be central to the rasp's function as an antipredator signal.

A. Body size and signal features

Although the size of the sound-producing apparatus was tightly correlated with the body size of these animals, the acoustic features were less so (Fig. 3). Pulse rate was correlated with body size, such that larger animals produced rasps with a slower pulse rate [Fig. 3(d)]. However, dominant frequency and power were not strongly associated with body size [Figs. 3(e) and 3(f)]. Given that the rasps are broadband signals with little tonal definition, it is perhaps unsurprising that a significant correlation between body size and dominant frequency was not observed. Future studies should examine a broader range of body sizes and examine the effect of motivation on signal features. For example, the stick-slip mechanism of sound production may permit greater power output when individuals pull the plectrum more tightly against the file thereby generating a higher normal force and louder

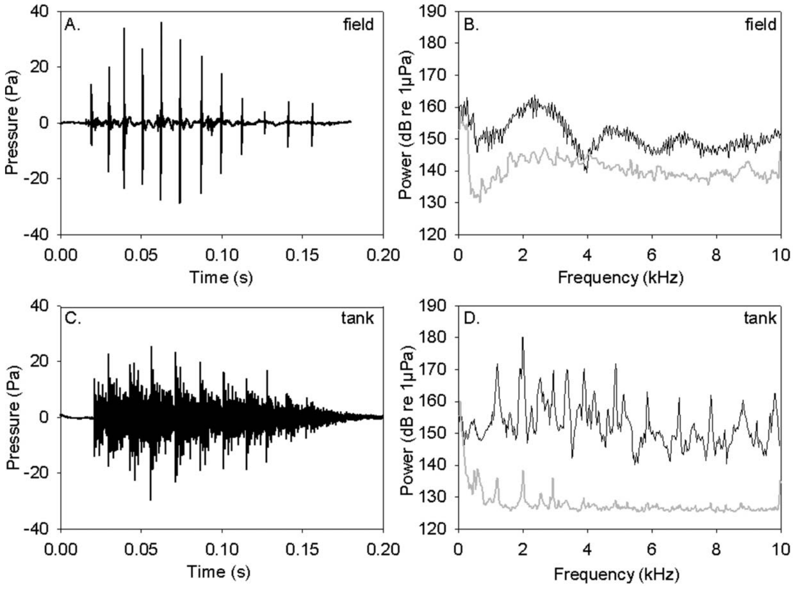


FIG. 5. Recordings of the same lobster producing a rasp in two different environments. In the field, the pulse structure of the rasp is evident (A) and the sound shows little resonant structure [(B) power spectrum settings as described in Sec. II]. In a tank, the reverberations obscure the pulse structure (C) and a series of harmonics are apparent (D). The grayed spectra in (B) and (D) indicate the signature of the ambient background noise. This particular lobster was positioned 53 cm from the hydrophone in the tank and 30 cm from the hydrophone in the field, which may also have caused spectral differences (Akamatsu *et al.*, 2002).

TABLE IV. Average power of rasps at varying distances in the field. Power is reported as dB re 1 μPa and dB re background noise ± standard deviation. Samples sizes are number of individuals (N) followed by number of rasps recorded per individual.

	0.9 m $N=8$ (3–18)	1.1 m $N=11$ (3–22)	1.3 m $N=9$ (1–22)	1.5 m $N=5$ (1–9)
dB re 1 μPa	150.2 ± 1.4	150.2 ± 2.1	149.2 ± 1.6	149.2 ± 1.7
dB re background noise	1.8 ± 1.6	1.3 ± 2.2	0.45 ± 2.1	1.0 ± 2.4

sound (Patek and Baio, 2007). The fact that the lobsters were hand-held in this study may have elicited different signaling behavior than in freely-moving individuals (although their rapid escape responses preclude measuring calibrated power levels at known distances in freely-moving individuals). Also, repeated stimulation of the same individuals may have yielded habituation, again influencing signal feature patterns over the time-course of these experiments. This might explain the unexpected variation in power levels across the four field recording distances (Table IV).

Scaling of sound with body size has been examined previously in several spiny lobster species and across the family as a whole (Meyer-Rochow and Penrose, 1974; 1976; Patek, 2002; Patek and Oakley, 2003; Patek and Baio, 2007). Across the palinurid family, and within *Panulirus argus*, pulse rate and number of pulses were positively correlated with the length of the file while rasp duration was negatively correlated with file length (Patek and Oakley, 2003). At early developmental stages, correlations between dominant frequencies and size were not found (Meyer-Rochow and Penrose, 1974); however, across juvenile and adult *Panulirus longipes*, a positive correlation between body size and rasp duration was found, and, similar to the results of this study, there was a negative correlation between size and pulse rate (Meyer-Rochow and Penrose, 1976).

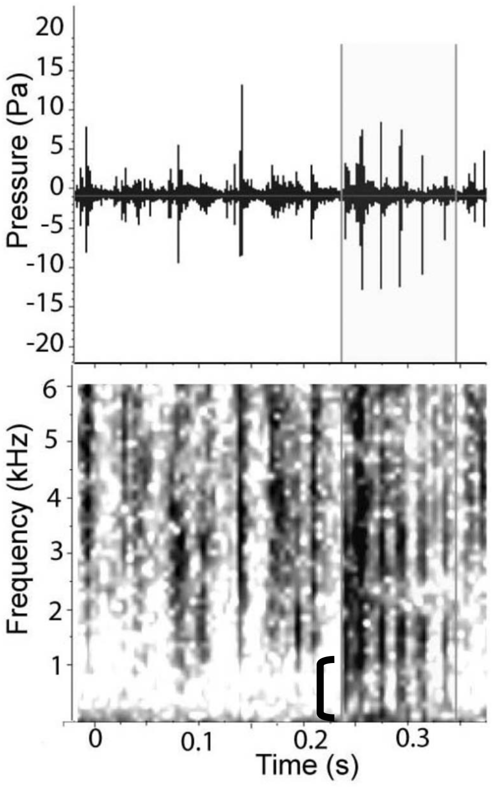


FIG. 6. A rasp produced by a lobster in the field at 1.1 m from the hydrophone. The rasp is highlighted and is shown as a waveform (upper) and spectrogram (lower; RAVEN PRO software v. 1.4, Hanning window, 512 sample window size, 3 dB filter bandwidth at 140 Hz resolution). The bracket indicates the energy extending below 1 kHz from the rasp, whereas the ambient background noise is less powerful in this frequency range.

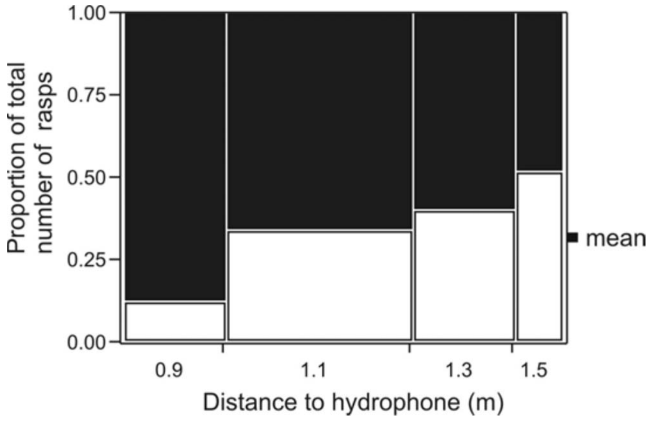


FIG. 7. The proportion of rasps with power levels above the background noise (black bars) and below background noise level (white bars) at four distances from the hydrophone. The width of the horizontal bars represents the relative number of rasps recorded (0.9 m: 65 rasps; 1.1 m: 118 rasps; 1.3 m: 65 rasps; 1.5 m: 29 rasps). The overall mean is indicated to the right.

B. Plectrum activation and rasp variation

While the lack of body size and signal feature correlation in spiny lobsters might be explained by the study's limited body size range, another key factor is the behavioral use of the dual sound-producing apparatuses. For example, when examining temporal features of rasps, a larger number of pulses yielded a longer duration rasp (Fig. 4). While this was not explained by body size variation, it may instead be attributed to the lobster's use the pair of plectrums rather than a single plectrum to generate sounds.

Behavior, particularly the use of one or both plectrums had strong influences on the temporal features of the rasp. When two plectrums were used concurrently, the number of pulses, rasp duration, and pulse rate were greater (although these results may be confounded by individual differences; Table II). Furthermore, concurrent activation of both plectrums was correlated with the attempt to escape by tail-flipping and the overall activity of the animal. Thus, the be-

TABLE V. Correlation between plectrum activation and behavioral units (tail flip and leg movement). A nonparametric, two-sample Kruskal–Wallis test was used to test whether these behaviors were associated with rasp rate and plectrum use. Sample size was 20 individuals each sampled once. * $p<0.05$.

	Overall rate of rasp production	Number of rasps when both plectrums were concurrently active
Tail flip	$Z=0.6622$; $p=0.51$	$Z=2.3437$; $p=0.019$*
Leg movement	$Z=-2.2955$; $p=0.022$*	$Z=-1.7812$; $p=0.075$

havioral motivation of the animal may more directly influence signal characteristics than the body size even though body size corresponds closely with the size of the sound-producing apparatus.

The relevance of behavior to signal features has been suggested previously in spiny lobsters (Patek and Oakley, 2003) and demonstrated in other systems with multiple signal-generating devices. For example, the searobin (*Prionotus carolinus*) has a pair of sonic muscles which it can contract simultaneously to generate greater amplitude or sequentially to produce a higher fundamental frequency (Connaughton, 2004). The California mantis shrimp (*Hemisquilla californiensis*) may also use its paired sonic muscles to vary signal features (Patek and Caldwell, 2006).

The behavioral manipulation of the signal features may be important for tailoring an acoustic response to particular predators. For example, multiple studies have shown that vertebrates produce signal features specific to the predator (e.g., Templeton *et al.*, 2005). Thus, it will be important in future studies to present a range of predators to spiny lobsters and assess whether they respond differently depending on the relative size, risk, and hearing capabilities of that particular predator.

C. Rasps and their acoustic environment

Consistent with previous studies (Parvulescu, 1967; Meyer-Rochow and Penrose, 1976; Akamatsu *et al.*, 2002), there were significant effects of the tank and field on the frequency characteristics of the sound (Fig. 5; Table III). The tank recordings yielded average dominant frequencies of 1794 Hz, whereas in the field, the dominant frequencies averaged 633 Hz. The second most powerful frequency was similar in both settings—1796 Hz in the tank and 1590 Hz in the field—suggesting that the tank resonated the higher frequencies in the rasp or damped the lower dominant frequency. These substantial differences in frequencies and temporal structure between the tank and field strongly suggest that tank-based aquatic recordings should be interpreted with caution and are not useful for comparisons and characterizations of frequency-spectra.

The high intensity collapse of cavitation bubbles dominated the acoustic landscape around Santa Catalina Island, the site of this study. The majority of these sounds in other, similar environments have been attributed to snapping shrimp (Johnson *et al.*, 1947; Au and Banks, 1998; Versluis *et al.*, 2000), although it is likely that cavitation sounds are being produced by other organisms as well (Colson *et al.*, 1998; Patek *et al.*, 2004; Patek and Caldwell, 2005; Simon *et al.*, 2005). Consistent with our measurements of field background noise averaging 149.3 dB re 1 μPa, snapping shrimp (*Synalpheus paraneomeris*) generate signals at 183–189 dB re 1 μPa at 1 m from a hydrophone in a tank (Au and Banks, 1998).

The average power level of the rasps, 150 dB re 1 μPa, is quite loud compared to measurements of marine acoustic signals from similar sized organisms (excluding the sound of cavitation). A study of two spiny lobster species, *Panulirus homarus* and *Palinustus waguensis* (misspelled in the original paper), documented power levels of 50–143 dB (Latha *et al.*, 2005); however, the reference level and distance from the recording device were not specified, so it is difficult to draw comparisons with the present data. The damselfish (*Abudefduf abdominalis*) generates courtship calls at 105–119 dB re 1 μPa at 0.5–1 m (Maruska *et al.*, 2007). Toadfish (*Halobatrachus didactylus*) acoustic power scales with body size, with pressure levels ranging from approximately 108–140 dB re 1 μPa (Vasconcelos and Ladich, 2008). *Opsanus tau* toadfish produce boatwhistle calls of an average 126 dB re 1 μPa at 1 m (Barimo and Fine, 1998).

Thus, the spiny lobster's rasp is loud, but so is the background noise (Figs. 6 and 7). A primary consequence of the loud background noise is that the rasps are obscured by the ambient background noise even though they attenuate minimally over the distances in which a predator encounter might occur. Given that the rasps are similar in power to the ambient background noise, the probability that they will be obscured is quite high—approximately 31% of the rasps recorded had a negative decibel level relative to the background noise (Fig. 7). This confers an advantage in the context of the antipredator function—the sounds are both loud and local, and perhaps less likely to attract additional nearby predators to the scene.

The frequency structure of the background noise relative to the rasp may also be important for propagation (Fig. 6). A quiet window is present below 1 kHz, a region in which the rasp's power output is relatively high. It is possible that spiny lobsters make use of such a "window" similar to gobies shown to communicate in the quiet low-frequency region in a noisy stream environment (Lugli *et al.*, 2003; Lugli and Fine, 2007). Like the gobies, it is also possible that antipredator communication is occurring in the near-field, thus measurements of particle velocity at close-ranges would yield a more accurate portrait of the rasp's acoustic landscape. While many of the spiny lobster's fish predators can detect pressure waves, most marine organisms are also sensitive to particle vibrations in the near-field. Characterizing the near-field of these local antipredator sounds is necessary both to understand the propagation of these signals and to determine the relevant signal features to attacking predators.

Various lobster species have been shown to detect vibrations in the near-field and at low frequencies (less than 200 Hz), yet the presence of pressure-sensitive hearing structures in crustaceans remains contentious (Cohen, 1955; Offutt, 1970; Tazaki and Ohnishi, 1974; Goodall *et al.*, 1990; Budelmann, 1992; Popper *et al.*, 2001; Lovell *et al.*, 2005; Lovell *et al.*, 2006). Previous research suggested that the rasps could function in the near-field to warn neighboring conspecifics (Lindberg, 1955; Meyer-Rochow *et al.*, 1982) *via* an "alarm signal." However, given that palinurid larvae cycle for many months before settling (Phillips *et al.*, 2006), it is unlikely that they are genetically related and thus the fundamental assumption that alarm calls aid close relatives (Caro, 2005) would not be met.

In conclusion, there is a web of interconnections between the basic mechanism of sound production, the behavioral deployment, and the ambient environment in which these sounds are produced; each component is essential to the production and propagation of the signal. Across the size ranges of the lobsters included in this study, there are no obvious signals to potential predators about body size. However, the overall behavior of the lobster strongly impacts the rasp features produced, suggesting that there may be a more important association between the stimulus and acoustic response than we addressed in these particular experiments. The loud natural environment plays a key role in masking the rasps even though the aquatic environment minimally attenuates sound over these small distances. This first analysis of antipredator acoustics and behavior in the California spiny lobster suggests that much remains to be learned in this rich frontier of marine bioacoustic research.

ACKNOWLEDGMENTS

The authors thank J. Baio, M. deVries, R. Jaciento, B. Nowroozi, M. Patek, V. Patek, A. Summers, N. Valencia, and W. Van Trump for helping to collect the sound and video recordings. They thank two anonymous reviewers for their constructive comments. They appreciate the staff support by the Wrigley Institute for Environmental Studies. Funding to SNP was provided by the UC Berkeley Undergraduate Research Apprenticeship Program, the Committee on Research Junior Faculty Research Grant and the Hellman Family Faculty Fund.

Akamatsu, T., Okumura, T., Novarini, N., and Yan, H. (**2002**). "Empirical refinements applicable to the recording of fish sounds in small tanks," J. Acoust. Soc. Am. **112**, 3073–3082.

Alexander, A. J. (**1958**). "On the stridulation of scorpions," Behaviour **12**, 333–352.

Atema, J., and Cobb, J. S. (**1980**). "Social behavior," in *The Biology and Management of Lobsters: Physiology and Behavior*, edited by J. S. Cobb and B. F. Phillips (Academic, New York), pp. 409–450.

Athenaeus (**300**). *Le Banquet des Savants [The Banquet of Scientists]* (Lamy Ed., Paris).

Au, W. W. L., and Banks, K. (**1998**). "The acoustics of the snapping shrimp *Synalpheus parneomeris* in Kaneohe Bay," J. Acoust. Soc. Am. **103**, 41–47.

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.

Bouwma, P. E. (**2006**). "Aspects of antipredation in *Panulirus argus* and *Panulirus guttatus*: Behavior, morphology and ontogeny," in *Biological Science* (Florida State University, Tallahassee).

Bouwma, P., and Herrnkind, W. (**2004**). "The antipredator function of sound production by the spiny lobster *Panulirus argus*," in *Seventh International Conference and Workshop on Lobster Biology and Management* (Hobart, Tasmania).

Bouwma, P., and Hernnkind, W. (**2007**). "Aspects of antipredation in *Panulirus argus* and *P. guttatus*: Behavior, morphology and ontogeny," The Lobster Newsletter **20**, 4–6.

Budelmann, B. U. (**1992**). "Hearing in Crustacea," in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 131–139.

Caro, T. (**2005**). *Antipredator Defenses in Birds and Mammals* (The University of Chicago Press, Chicago).

Chivers, D. P., Brown, G. E., and Smith, R. J. F. (**1996**). "The evolution of chemical alarm signals: Attracting predators benefits alarm signal senders," Am. Nat. **148**, 649–659.

Cohen, M. J. (**1955**). "The function of receptors in the statocyst of the lobster *Homarus americanus*," J. Physiol. (London) **140**, 9–34.

Colson, D. J., Patek, S. N., Brainerd, E. L., and Lewis, S. M. (**1998**). "Sound production during feeding in *Hippocampus* seahorses (Syngnathidae)," Environ. Biol. Fishes **51**, 221–229.

Connaughton, M. A. (**2004**). "Sound generation in the searobin (*Prionotus carolinus*), a fish with alternate sonic muscle contraction," J. Exp. Biol. **207**, 1643–1654.

Edmunds, M. (**1974**). *Defence in Animals: A Survey of Anti-Predator Defenses* (Longman Group Ltd., Essex).

Field, L. H., Evans, A., and MacMillan, D. L. (**1987**). "Sound production and stridulatory structures in hermit crabs of the genus *Trizopagurus*," Journal of Marine Biology, U.K. **67**, 89–110.

Fish, M. P., and Mowbray, W. H. (**1970**). *Sounds of Western North Atlantic Fishes* (The Johns Hopkins Press, Baltimore).

Gamberale, G., and Tullberg, B. S. (**1996**). "Evidence for a peak-shift in predator generalization among aposematic prey," Proc. R. Soc. London, Ser. B **263**, 1329–1334.

George, R. W., and Main, A. R. (**1967**). "The evolution of spiny lobsters (Palinuridae): A study of evolution in the marine environment," Evolution (Lawrence, Kans.) **21**, 803–820.

Gittleman, J. L., and Harvey, P. H. (**1980**). "Why are distasteful prey not cryptic?," Nature (London) **286**, 149–150.

Goodall, C., Chapman, C., and Neil, D. (**1990**). "The acoustic response threhold of the Norway lobster, *Nephrops norvegicus* (L.) in a free sound field," in *Frontiers in Crustacean Neurobiology*, edited by K. Wiese, W. D. Krenz, J. Tautz, and H. Reichert (Birkhauser, Boston), pp. 106–113.

Guinot-Dumortier, D., and Dumortier, B. (**1960**). "La stridulation chez les Crabes [Stridulation by crabs]," Crustaceana **2**, 117–155.

Henninger, H. P., and Watson, W. H., III (**2005**). "Mechanisms underlying the production of carapace vibrations and associated waterborne sounds in the American lobster, *Homarus americanus*," J. Exp. Biol. **208**, 3421–3429.

Johnson, M. W., Everest, F. A., and Young, R. W. (**1947**). "The role of snapping shrimp (*Crangon* and *Synalpheus*) in the production of underwater noise in the sea," Biol. Bull. **93**, 122–138.

Latha, G., Senthilvadivu, S., Venkatesan, R., and Rajendran, V. (**2005**). "Sound of shallow and deep water lobsters: Measurements, analysis and characterization (L)," J. Acoust. Soc. Am. **117**, 2720–2723.

Lewis, E. E., and Cane, J. H. (**1990**). "Stridulation as a primary antipredator defense of a beetle," Anim. Behav. **40**, 1003–1004.

Lindberg, R. G. (**1955**). "Growth, population dynamics, and field behavior in the spiny lobster, *Panulirus interruptus* (Randall)," University of California Publications in Zoology **59**, 157–248.

Lovell, J. M., Findlay, M. M., Moate, R. M., and Yan, H. Y. (**2005**). "The hearing abilities of the prawn *Palaemon serratus*," Comp. Biochem. Physiol. A **140**, 89–100.

Lovell, J. M., Moate, R. M., Christiansen, L., and Findlay, M. M. (**2006**). "The relationship between body size and evoked potentials from the statocysts of the prawn *Palaemon serratus*," J. Exp. Biol. **209**, 2480–2485.

Lugli, M., and Fine, M. L. (**2007**). "Stream ambient noise, spectrum and propagation of sounds in the goby (*Padogobius martensii*): Sound pressure and particle velocity," J. Acoust. Soc. Am. **122**, 2881–2892.

Lugli, M., Yan, H. Y., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: The relationship between ambient noise, hearing thresholds and sound spectrum," J. Comp. Physiol. **189**, 309–320.

Maruska, K. P., Boyle, K. S., Dewan, L. R., and Tricas, T. C. (**2007**). "Sound production and spectral hearing sensitivity in the Hawaiian sergeant damselfish, *Abudefduf abdominalis*," J. Exp. Biol. **210**, 3990–4004.

Masters, W. M. (**1979**). "Insect disturbance stridulation: Its defensive role," Behav. Ecol. Sociobiol. **5**, 187–200.

Mendelson, M. (**1969**). "Electrical and mechanical characteristics of a very fast lobster muscle," J. Cell Biol. **42**, 548–563.

Meyer-Rochow, V. B., and Penrose, J. D. (**1974**). "Sound production and sound emission apparatus in puerulus and postpuerulus of the western rock lobster (*Panulirus longipes*)," J. Exp. Zool. **189**, 283–289.

Meyer-Rochow, V. B., and Penrose, J. D. (**1976**). "Sound production by the western rock lobster *Panulirus longipes* (Milne Edwards)," Journal of Experimental Marine Biology and Ecology **23**, 191–209.

Meyer-Rochow, V. B., Penrose, J. D., Oldfield, B. P., and Bailey, W. J. (**1982**). "Phonoresponses in the rock lobster *Panulirus longipes* (Milne Edwards)," Behav. Neural Biol. **34**, 331–336.

Moulton, J. (**1957**). "Sound production in the spiny lobster *Panulirus argus* (Latreille)," Biol. Bull. **113**, 286–295.

Mulligan, B. E., and Fischer, R. B. (**1977**). "Sounds and behavior of the spiny lobster *Panulirus argus*," Crustaceana **32**, 185–199.

Offutt, C. G. (**1970**). "Acoustic stimulus perception by the American lobster *Homarus americanus* (Decapoda)," Experientia **26**, 1276–1279.

Parker, T. J. (**1878**). "Note on the stridulation organ of *Panulirus vulgaris*," Proceedings of the Zoological Society of London **1878**, 442–444.

Parker, T. J. (**1883**). "On the structure of the head in *Palinurus* with special reference to the classification of the genus," Nature (London) **29**, 189–190.

Parvulescu, A. (**1967**). "The acoustics of small tanks," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 7–9.

Patek, S. N. (**2001**). "Spiny lobsters stick and slip to make sound," Nature (London) **411**, 153–154.

Patek, S. (**2002**). "Squeaking with a sliding joint: Mechanics and motor control of sound production in palinurid lobsters," J. Exp. Biol. **205**, 2375–2385.

Patek, S. N., and Baio, J. E. (**2007**). "The acoustic mechanics of stick-slip friction in the California spiny lobster (*Panulirus interruptus*)," J. Exp. Biol. **210**, 3538–3546.

Patek, S. N., and Caldwell, R. L. (**2005**). "Extreme impact and cavitation forces of a biological hammer: Strike forces of the peacock mantis shrimp (*Odontodactylus scyllarus*)," J. Exp. Biol. **208**, 3655–3664.

Patek, S. N., and Caldwell, R. L. (**2006**). "The stomatopod rumble: Sound production in *Hemisquilla californiensis*," Mar. Freshwater Behav. Physiol. **39**, 99–111.

Patek, S. N., and Oakley, T. H. (**2003**). "Comparative tests of evolutionary tradeoffs in a palinurid lobster acoustic system," Evolution (Lawrence, Kans.) **57**, 2082–2100.

Patek, S. N., Korff, W. L., and Caldwell, R. L. (**2004**). "Deadly strike mechanism of a mantis shrimp," Nature (London) **428**, 819–820.

Patek, S. N., Feldmann, R. M., Porter, M., and Tshudy, D. (**2006**). "Phylogeny and evolution of lobsters," in *Lobsters: Biology, Management, Aquaculture and Fisheries*, edited by B. F. Phillips (Blackwell, Ames, IA).

Phillips, B. F., Booth, J. D., Cobb, J. S., Jeffs, A. G., and McWilliam, P. (**2006**). "Larval and postlarval ecology," in *Lobsters: Biology, Management, Aquaculture and Fisheries*, edited by B. F. Phillips (Blackwell, Oxford), pp. 231–262.

Popper, A. N., Salmon, M., and Horch, K. W. (**2001**). "Acoustic detection and communication by decapod crustaceans," J. Comp. Physiol., A **187**, 83–89.

Ruxton, G. D., Sherratt, T. N., and Speed, M. P. (**2004**). *Avoiding Attack: The Evolutionary Ecology of Crypsis, Warning Signals and Mimicry* (Oxford University Press, Oxford).

Sandeman, D. C., and Wilkens, L. A. (**1982**). "Sound production by abdominal stridulation in the Australian Murray River crayfish, *Euastacus armatus*," J. Exp. Biol. **99**, 469–472.

Sargent, T. D. (**1990**). "Startle as an anti-predator mechanism, with special reference to the underwing moths, (Catocala)," in *Insect Defenses: Adaptive Mechanisms and Strategies of Prey and Predators*, edited by D. L. Evans and J. O. Schmidt (State University of New York Press, Albany), pp. 229–249.

Sherratt, T. N., and Beatty, C. D. (**2003**). "The evolution of warning signals as reliable indicators of prey defense," Am. Nat. **162**, 377–389.

Simon, M., Wahlberg, M., Ugarte, F., and Miller, L. A. (**2005**). "Acoustic characteristics of underwater tail slaps used by Norwegian and Icelandic killer whales (*Orcinus orca*) to debilitate herring (*Clupea harengus*)," J. Exp. Biol. **208**, 2459–2466.

Smale, M. (**1974**). "The warning squeak of the Natal rock lobster," South African Association Marine Biology Research Bulletin **11**, 17–19.

Speed, M. P. (**2000**). "Warning signals, receiver psychology and predator memory," Anim. Behav. **60**, 269–278.

Summers, A. (**2001**). "The lobster's violin," Natural History **110**, 26–27.

Tazaki, K., and Ohnishi, M. (**1974**). "Responses from tactile receptors in the antenna of the spiny lobster *Panulirus japonicus*," Comp. Biochem. Physiol. A **47A**, 1323–1327.

Templeton, C. N., Greene, E., and Davis, K. (**2005**). "Allometry of alarm calls: Black-capped chickadees encode information about predator size," Science **308**, 1934–1937.

Vasconcelos, R. O., and Ladich, F. (**2008**). "Development of vocalization, auditory sensitivity and acoustic communication in the Lusitanian toadfish *Halobatrachus didactylus*," J. Exp. Biol. **211**, 502–509.

Versluis, M., Schmitz, B., von der Heydt, A., and Lohse, D. (**2000**). "How snapping shrimp snap: Through cavitating bubbles," Science **289**, 2114–2117.

Relationship between sperm whale (*Physeter macrocephalus*) click structure and size derived from videocamera images of a depredating whale (sperm whale prey acquisition)

Delphine Mathias[a)] and Aaron Thode
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Jan Straley
University of Alaska Southeast, Sitka, Alaska 99835

Kendall Folkert
Alaska Longline Fishermen's Association, P.O. Box 6497, Sitka, Alaska 99835

(Received 12 August 2008; revised 10 February 2009; accepted 18 February 2009)

Sperm whales have learned to depredate black cod (*Anoplopoma fimbria*) from longline deployments in the Gulf of Alaska. On May 31, 2006, simultaneous acoustic and visual recordings were made of a depredation attempt by a sperm whale at 108 m depth. Because the whale was oriented perpendicularly to the camera as it contacted the longline at a known distance from the camera, the distance from the nose to the hinge of the jaw could be estimated. Allometric relationships obtained from whaling data and skeleton measurements could then be used to estimate both the spermaceti organ length and total length of the animal. An acoustic estimate of animal length was obtained by measuring the inter-pulse interval (IPI) of clicks detected from the animal and using empirical formulas to convert this interval into a length estimate. Two distinct IPIs were extracted from the clicks, one yielding a length estimate that matches the visually-derived length to within experimental error. However, acoustic estimates of spermaceti organ size, derived from standard sound production theories, are inconsistent with the visual estimates, and the derived size of the junk is smaller than that of the spermaceti organ, in contradiction with known anatomical relationships. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097758]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 3444–3453

I. INTRODUCTION

The question of whether an organism's anatomical dimensions can be inferred from features of its acoustic signal has a long history in bioacoustics. The bulk of this research has focused on inferring the length of the mammalian vocal tract via formant analysis or other spectral measures.[1–5]

Attempts to derive a cetacean's body size from its acoustic signal structure face even greater challenges than terrestrial studies, due to the difficulty of obtaining independent and accurate measurements of animal dimensions in the wild. The most detailed theory linking a cetacean sound to an individual's anatomical dimensions concerns sperm whales (*Physeter macrocephalus*), which produce a 25–30 ms transient sound called a click. Researchers have divided sperm whale sounds into a variety of categories, based on the timing between subsequent clicks in a sound sequence. This timing pattern, called the inter-click interval (ICI), has been used to classify clicks into usual clicks, slow clicks, creaks, and codas.[6] Clicks are believed to serve a variety of functions, including echolocation and communication, depending on the click pattern. Usual clicks typically have an ICI of 0.5–1.0 s, while creaks display a shorter ICI between 0.2 s and 0.5 s.[7,8] ICI values within a given creak also tend to decrease with time, and orientation measurements from tagged animals have demonstrated that sudden changes in orientation are associated with creak sounds.[8,9] These observations, among others, suggest that creaks are used as an echolocation signal.[6,7,10–14]

A click displays internal structure in the form of several local maxima, or "pulses," with the time interval between pulses within a click labeled the "inter-pulse interval" (IPI). Note that this quantity is different from the ICI defined earlier.

A 1972 paper by Norris and Harvey[10] hypothesized that a click is initially generated at a pair of fatty tissues called the *museau de singe*, at the front of the animal's nose.[42] Under this hypothesis, most acoustic energy escapes the nose directly, while the remaining fraction propagates backwards through the spermaceti organ, gets reflected forward by the frontal air sac, and finally escapes into the water either via the anterior end of the spermaceti organ or the rostrum (Fig. 1).

Modifications to this model have been made by Mohl[15–17] and supported by Zimmer *et al.*[18,19] Under this "bent-horn" interpretation, the initial omnidirectional pulse P_0 transmitted directly into the water is actually only a small portion of the sound generated at the museau de singe, whereas the bulk of the energy propagates backwards through the spermaceti organ and reflects off the frontal air sac. The largest component of this reflection is transmitted

[a)] Author to whom correspondence should be addressed. Electronic mail: delphine.mathias@gmail.com

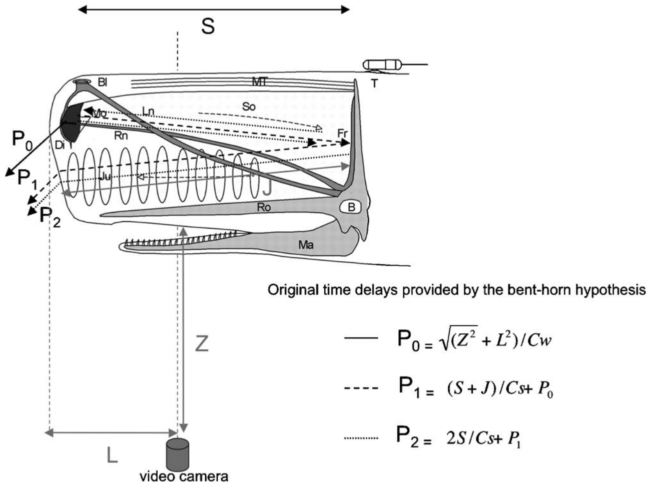


FIG. 1. Bent-horn model of sound production (adapted from Fig. 1 of Ref. 43) and associated formulas for acoustic path lengths. Anatomical labels: B: brain, Bl: blow hole, Di: distal air sac, Fr: frontal air sac, Ju: junk, Ln: left naris, Ma: mandible, Mo: monkey/phonics lips or museau de singe, MT muscle/tendon layer, Rn: right naris, Ro: rostrum, So: spermaceti organ. Propagation path variables: S: length of the spermaceti organ, J: length of the junk, Z: distance between the videocamera and the jaw of the animal, L: lateral distance between the nose and the videocamera, Cw: sound speed in the water = 1500 m/s, Cs: sound speed in the spermaceti organ = 1370 m/s.

into the water via the junk, creating a highly directional main pulse P_1. Recent measurements[19] also indicate that a portion of the energy reflected from the frontal air sac (just over the skull) escapes directly into the water, creating a $P_{1/2}$ pulse that can be detected between the P_0 and P_1 pulse in recordings made off the longitudinal axis of the animal. Finally, under the revised theory, a portion of the P_1 pulse energy propagates back into the junk and/or spermaceti organ, reflects off the frontal sac, and passes a second time through the junk into the water, creating a P_2 pulse. These propagation paths are illustrated in Fig. 1.

According to both the "Norris–Harvey" and bent-horn theories, the nominal IPI, or IPI between P_1 and P_2,[20] is proportional to the two-way acoustic travel time between the *museau de singe* and frontal sac, provided that the measurements are made either directly in front of or behind the animal, such that the aspect-dependent $P_{1/2}$ pulse merges with the P_0 or P_1 pulse. In 2005, Zimmer *et al.*[19] showed that measurements made off the longitudinal axis yield off-axis "distortion," in the sense that the appearance of a $P_{1/2}$ pulse obfuscates the interpretation of the timing measurements.

A natural consequence of these hypotheses is that the nominal IPI could be used to estimate the length of the spermaceti organ or junk. Allometric relationships between nose size and body length[21] then suggest that the nominal IPI should be correlated with body length. Indeed, functional regressions between the nominal IPI and body length have been published.[22–24] These polynomials were derived using analyzed IPIs from whales whose length had been independently measured while they had surfaced. However, no independent means of estimating the animals' spermaceti organ size were available.

Various automated methods have been explored for estimating the nominal IPI, but cepstral analysis[25] has been used in several published papers,[20,23,26] under the assumption that a click can be modeled as a convolution of a "source" and a "reflection" function. Teloni *et al.*[20] showed that aspect-dependent features of the IPI estimates could be effectively removed by averaging a large number of cepstra derived from clicks recorded during an animal's foraging dive, during which the animal presumably presents a wide variety of orientations with respect to the recording hydrophone, thus permitting aspect-dependent features of the clicks to be averaged down and the nominal IPI to be enhanced.

This paper uses an unusual dataset to directly compare visual estimates of the size of a sperm whale's head with acoustic estimates of its total length, spermaceti organ length, and junk length. Under normal foraging conditions sperm whales typically dive to depths greater than 300 m,[8,9,27–29] making it impossible to acquire video recordings of prey acquisition attempts. However, sperm whales have learned how to depredate fishing gear, particularly demersal longline operations, in a number of locations around the globe since 2002,[30–34] including Norway, Greenland, eastern Canada (Labrador and Newfoundland), Chile, and the Falkland Islands. It is the largest marine mammal known to depredate on human fishing activities, and these activities have received increasing coverage in the scientific literature.[31–39]

In 2003 the Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) was established by scientists, managers, and fishermen to characterize the severity of sperm whale depredation activity on black cod (*Anoplopoma fimbria*) off Sitka, Alaska. Passive acoustic measurements collected during SEASWAP discovered that the animals occasionally dove below the fishing vessels at depths less than 100 m, depths presumably shallow enough to permit visual observations of this activity. In 2006, videocameras were deployed for the first time to capture this behavior.

Section II describes the video and acoustic recording equipment, discusses how the camera was deployed from a fishing vessel during an active longline haul, and then outlines the procedures used to derive visual and acoustic estimates of the size of an animal captured on-camera. Section III describes a specific depredation encounter recorded on May 31, 2006, during which the whale touches the fishing line at a known distance from an underwater videocamera lowered to 108 m depth. The visual and acoustic estimates of the animal's size are compared, and the timing between a particular set of pulses within a click yields an acoustic length estimate that corresponds well with the visually-derived length estimate. The results confirm earlier studies that the IPI can be correlated with total body length, but finds inconsistencies in the visual and acoustic estimates of spermaceti organ size.

II. PROCEDURE

A. Video and audio recording equipment

Visual data were collected by a Sony HVR-1AU videocamera, housed in a Gates Underwater Products HC1/A1U housing using a WP-25 lens port (80° field of view, 63.5 mm diameter lens). The data were recorded onto Sony PHDVM-

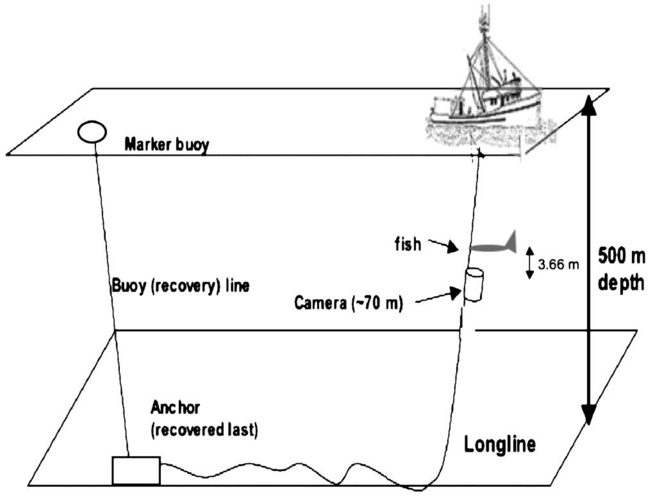


FIG. 2. Schematic of camera deployment from fishing vessel.

63DM DigitalMaster tapes in DVCAM format, recording over 60 min of uncompressed audio and video per tape. The housing and camera port were depth-rated to 152 m (500 ft). The widest angle field of view was used for the recordings (minimum zoom).

The Gates housing also contained a Kobitone PN 252-LM049 Electret Condenser Microphone, recorded to an audio channel on the videotape simultaneously with the videostream. The official sensitivity range of the microphone was 20–12 kHz, with −162 dB re 1 V/μPa sensitivity. The data were sampled at 48 kHz.

B. Deployment procedure

All deployments took place from the F/V Cobra, a 59 ft fishing vessel mastered by one of the authors (Folkert), who also designed most of the camera deployment system (Fig. 2). A standard demersal longline is comprised of 200 m long lengths of 1/4 in. groundline called “skates,” and at every meter along a skate a baited hook is attached by a small line called a “ganion,” typically about 20 cm long. Under normal operations the longline is typically deployed by dropping a surface buoy over the side, deploying sufficient line to account for the water depth and desired scope of the buoy, and then deploying a 30 kg anchor overboard. As the fishing vessel moves over the desired site, the baited longline pays out the back of the vessel. After a typical set has been cast (about 4–6 km long) a second anchor is deployed overboard, attached to a second surface buoy. After a 3–12 h “soak,” the upstream surface buoy is recovered, and the anchor line is passed over a “roller” mounted on the side of the vessel and through a hydraulically-operated pot-puller to haul the rope up and recover the fish. It is at this point that whales like to depredate the line, as fish are accessible throughout the water column.

The deployment technique was designed to avoid substantial changes in fishing procedures, which would have run into regulatory issues. Thus “blank” skates, marked at 50 m intervals, were inserted between every two baited skates, permitting a normal deployment and haul. During the recovery, a blank skate would be hauled until the start of the next hooked skate began to emerge from the water. If whales were present, the camera assembly was activated, sealed, and attached to the start of the hooked skate, with the lens port oriented so that it would be facing the ocean surface. The camera was attached to the line so that two to six fish (already present on the line) would be visible above the camera. The distance of each fish from the camera and the length of each fish was recorded and noted. The assembly would then be lowered down into the water until tape marks on the line indicated the desired deployment depth had been reached by the camera (Fig. 2). The true deployment depth could be measured precisely by attaching a commercial dive computer to the line just underneath the camera. Thus the entire procedure minimized alterations to standard fishing procedures.

C. Audio data analysis

Acoustic data from the camera were extracted from the video and stored as a 48 kHz 16-bit WAV file. A peak detector was used to locate echolocation clicks. For every local temporal maximum detected, a 30 ms click sample was extracted and saved, centered around the arrival time of the signal peak. This length of time window was chosen to avoid contamination by preceding or subsequent clicks.

The time difference between subsequent clicks was used to estimate the ICI. By plotting ICI as a function of time, click detections that shared the same ICI “trajectory,” and thus presumably belonging to the same whale, could be selected for further analysis, removing clicks generated by other whales in the vicinity.

Once the click samples from the camera sequence were isolated, the IPI was then extracted from each time sample using two different methods: an incoherent peak detection method, applied to the Hilbert transform of the signal, and via the signal cepstrum,[26] which presumes that the click can be expressed as the convolution between a scattering/reflection function and an impulsive “excitation” function.

To obtain the first estimate of the click envelope, a Hilbert transform was applied to each click time series, and then the transforms were stacked on top of each other to permit a view of the click evolution. The time origin was defined as the arrival time of the most intense pulse within the click, which will be labeled as pulse B. The envelopes were averaged over every 0.6 s of the recording to produce estimates evenly sampled in time. To determine pulse intervals, the local maxima in the [5–10] ms interval before and after pulse B were flagged, using methods nearly identical to the original peak detection method used to detect clicks in the data. These intra-pulse maxima were then labeled A and C. (The conventional notation of P_0, P_1, etc., used by the benthorn hypothesis is avoided here in order to avoid a particular interpretation of the pulse structure). The precision in the pulse estimate was typically 0.05 ms.

The second measurement method used cepstral analysis. Heuristically speaking, the cepstrum measures any periodic oscillation of the signal spectrum with respect to frequency, which can be interpreted in the time domain as the time delay between an original signal and its reflection. Thus for the case of a sperm whale click, the Norris–Harvey hypothesis and later modifications predict that the cepstrum, which is measured in units of time, should display a peak at the time corresponding to the two-way travel time between the *museau de singe* and the frontal sac.[23].

After band-pass filtering between 2 and 18 kHz, cepstral estimates were made over the entire 30 ms window, the −15 to 0 ms window, and the 0–15 ms window, relative to the arrival time of pulse B. This was done in order to estimate separate IPI values for the A-B and B-C pulse intervals. Caution must be exercised when band-pass filtering a time series with a finite impulse response (FIR) filter before computing the signal cepstrum, since the application of the filter produces ripples in the output spectrum, which then produces an artificial peak in the cepstrum. It was found that by specifying a wide filter stopband of 1000 Hz, cepstral artifacts were restricted to 0–3 ms band in the cepstral output, outside the region of interest. Cepstral segments were then averaged over every 0.6 s of the recording (as with the Hilbert transform plots) to produce estimates evenly sampled in time.

Finally, the whale body length can be estimated from the IPI using the Gordon polynomial relating sperm whale body length and extracted IPI:[23]

$$\text{Total length (m)} = 4.833 + 1.453\text{IPI(s)} - 0.001\text{IPI}^2\text{(s)}. \tag{1}$$

Rhinelander and Dawson also derived a relationship between sperm whale size and IPI using photogrammetric length estimates, and measurements of the IPI for 12 individuals:[24]

$$\text{Total length (m)} = 17.120 - 2.189\text{IPI(s)} + 0.251\text{IPI}^2\text{(s)}. \tag{2}$$

Gordon used measurements from 11 individuals in the tropical Indian Ocean, while the Rhinelander and Dawson dataset contained 66 whales from Kaikoura (New Zealand). Gordon's regression contained only one individual longer than 12 m, while all individuals from Rhinelander and Dawson's dataset were larger than 12 m, possibly explaining the large difference between the polynomial coefficients between Eqs. (2) and (3). Thus, Rhinelander and Dawson's polynomial would be expected to be more suited for sperm whales living in the Gulf of Alaska, which are generally males greater than 12 m.

D. Video analysis

The dataset discussed here involves video images of a whale contacting a fishing line at a known distance of 3.66 m (12 ft) from the videocamera, with its body oriented roughly perpendicular to the camera. The underexposed features of the whale were enhanced by performing a power law gray scale transformation ($\gamma = 0.5$). The animal was judged to be perpendicular to the camera plane when the teeth on the left side of the lower jaw lined up in such a way as to block the view of corresponding teeth in the right side of the jaw.

On shore the camera was attached to a rope and oriented so that when it was lowered to 3.66 m depth in a pool, the camera pointed toward the surface. A trellis of known size (1.22 m on a side) was slid on the surface across the camera's field of view to permit a conversion of pixel separation to physical distance.

The camera was further calibrated for image distortion using a printed checkboard pattern of 1.2 m width, 1 m height, with 10 cm squares. The board was placed 3 m from the camera, and was moved in many directions in order to get as many angular views as possible. Then 20 images were extracted and standard camera calibration procedures were used to extract the system's focal length, principal point, and image distortion.[40] The net result is a distortion model that maps the radial and tangential distortion of every pixel in the image. It was found that pixels that lie within half the image width or height from the image center suffered less than 1% distortion, while pixels on the image borders experienced 4% distortion. The second estimate of distortion was made by placing the camera at 3.66 m water depth and measuring at the size variation of the treillis when moved across the camera's field of vision. In this case, it was found that the size of the treillis suffered less than 3% distortion when placed on the image border.

The combined scaling and distortion calibrations permitted three physical measurements to be extracted from the image: the distance between the tip of the snout to the point where the jaw hinges to the skull, or "gape angle" (SG), the distance between the tip of the nose to the start of the upper jaw (SJ), and the mean spacing between the animal's teeth (TS). Ideally, the distance between the blowhole and the center of the eye should have been measured (BE), since this distance is a good estimate of the spermaceti organ length,[41] and thus the IPI estimate could be directly related to spermaceti organ size. However, due to the low level of ambient illumination the eye cannot be located in the silhouette.

Allometric relationships published in Ref. 21 relate the SG and the SJ to total body length, thus permitting an estimate of the size of the skull and the length of the whale, even if the entire animal is not visible in the video.[42] Unfortunately, the SG measurement is not precise, because the actual location of the gape angle lies under the animal's skin, resulting in a possible error of 30%–50% in animal size. Thus the third measurement, TS, was made.

From the video it is possible to measure the distance between teeth along portions of the jaw, including the locations of the first four teeth near the tip of the jaw, and the rearmost nine teeth. To determine whether an allometric relationship exists between total body length and mean tooth spacing, tooth measurements were obtained from three male sperm whale skeletons with original total body lengths 18.3, 11.9, and 14 m. Two skeletons were from beached animals stored by the Natural History Museum of Los Angeles County in Southern California, and thus originated from the Pacific Ocean, and the third was from the Nantucket Historical Association Whaling Museum in Nantucket, Massachusetts. The tooth separation varies with distance along the jaw, so the mean value of the spacing was computed using the teeth that correspond to the teeth visible in the image. The ratio between the total measured length and mean tooth spacing (MLTL) was then computed. A corresponding MLTL can be estimated from the image, using the TS measured from

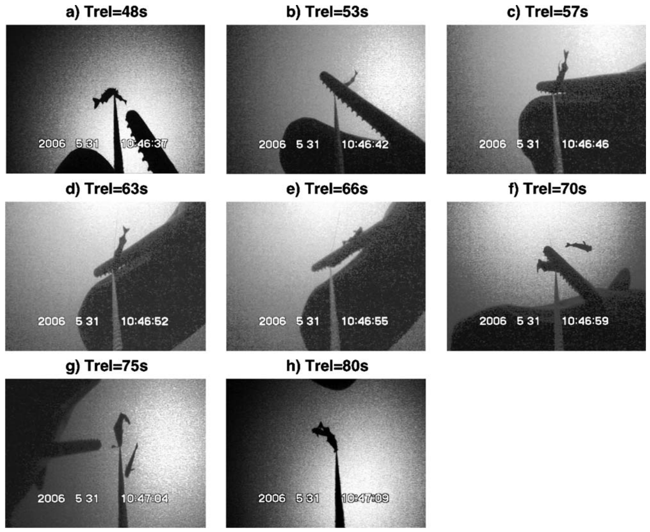

a) Trel=48s
b) Trel=53s
c) Trel=57s
d) Trel=63s
e) Trel=66s
f) Trel=70s
g) Trel=75s
h) Trel=80s

FIG. 3. Snapshots from the May 31 sperm whale encounter video, with the relative time in seconds counted from 10:45:59. Figures are labeled (a) through (h), starting from upper left and moving left-to-right across columns. (f) was used for the visual length estimate.

the image and the body length derived from the SJ and TS measurements, to determine whether the video MLTL lies within the ranges of the skeletons' MLTL.

The BE dimension, and thus the spermaceti organ size, was interpolated from the SJ, SG, and TS measurements, using the data of Nishiwaki *et al.* Anatomically speaking, the SG dimension is always larger than the BE dimension, which means an upper bound can be placed on the size of the spermaceti organ from the image analysis.

III. RESULTS

A. General description of the May 31st 2006 encounter

The first videotaped encounter of a depredating sperm whale took place on May 23 2006; however, the whale never contacted the rope, precluding a visual estimate of its body size. A second encounter was recorded on May 31, 2006, during one of the last attempts to obtain such a recording. The deployment depth of the camera was 108 m, and two black cod were present at distances of 3.66 and 5.49 m (12 and 18 ft) from the camera, respectively. The camera was activated at 10:25 local time, and by 10:45 had been lowered to the target depth. Significant sperm whale acoustic activity was recorded on the videotape from the moment the camera entered the water.

In this and all following sections, the relative timing of events will be expressed as seconds elapsed from 10:45:49, the time at which the acoustic activity from the animal initiated off camera. At 48 s [Fig. 3(a)], the jaw of a sperm whale appears and contacts the longline at an estimated distance of 1.8 m from the camera. The animal then slides along the longline until its jaw lies adjacent to the fish nearest the camera (3.66 m) at 53 s [Fig. 3(b)]. The animal then seems to completely close its jaw around the longline at that point by 57 s [Fig. 3(c)], deflecting the rope a considerable amount. At 63 s [Fig. 3(d)], the animal performs a slight barrel roll, and the fish attached a distance of 5.49 m from the camera breaks off the longline and floats away. The fish immediately adjacent to the jaw does not detach. As the animal opens its jaw, it stops producing creak clicks and works itself free of the line between 66, 70, and 75 s [Figs. 3(e)–3(g)]. Figure 3(f) is the key image used in subsequent analysis; it is the moment when the whale's head is judged perpendicular to the camera. Once free, the animal and floating fish are seen on-camera until 80 s, during which the whale seems to be orienting its head toward the loose fish. Unfortunately whatever happens next occurs after both the whale and fish float out of view.

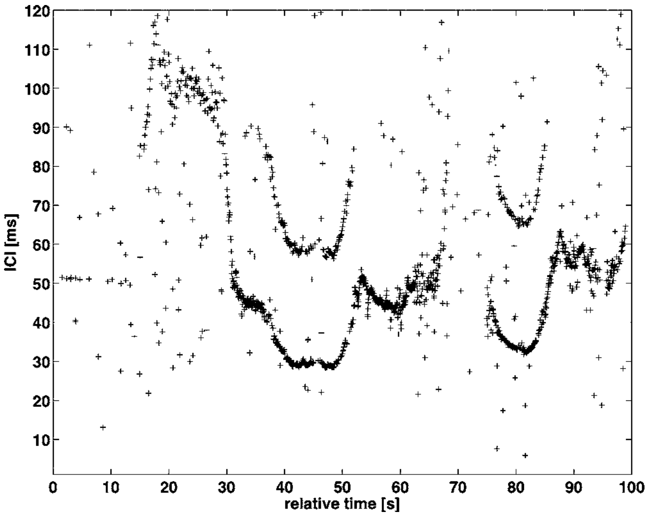


FIG. 4. ICI for the May 31 camera sequence, with relative time referenced in seconds from 10:45:49.

B. Acoustic analysis

The click detection procedure in Sec. II C generated 1723 click samples from the May 31 sequence over roughly 100 s. Figure 4 displays the ICI between individual clicks, expressed in the relative time scale. The ICI was estimated simply by measuring the time interval between successive detections in the detection function. Note that the time between 30 s and 85 s corresponds to ICI values of under 50 ms, traditionally considered characteristic of "creak" activity (e.g., Refs. 6 and 9), and thus most of the sequence analyzed here involves high SNR creak clicks.

As an aside, note that even at times when the whale is biting down on the line, the animal still produces creak clicks, and these clicks indicate a minimum ICI of 30 ms, or about 33/s, considerably lower than the maximum click rate of 90.9 click/s in other accounts.[7] The timing of these pulses is consistent with a two-way travel time from a target 22 m away, although the whale is clearly interested in targets just a couple of meters away. Thus Fig. 4 suggests the whale may be reaching a physiological limit in its click production rate. The lack of a visible ICI between 68 and 75 s corresponds to the time when the whale is working its jaw free of the line, and the ICI sequence reappears a fraction of a second after the jaw snaps free from the line. This correspondence between acoustic and visual events provides confidence that the clicks analyzed here were produced by the whale viewed on

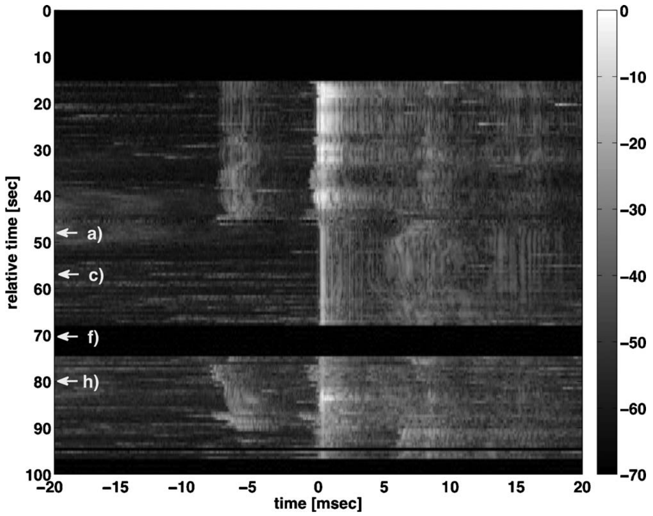


FIG. 5. Stacked click structure of the May 31 sequence, created by filtering each click sample between 2 and 18 kHz, applying the Hilbert transform, and averaging transforms in 0.6 s bins. Time on the *y*-axis is expressed relative to 10:45:49.

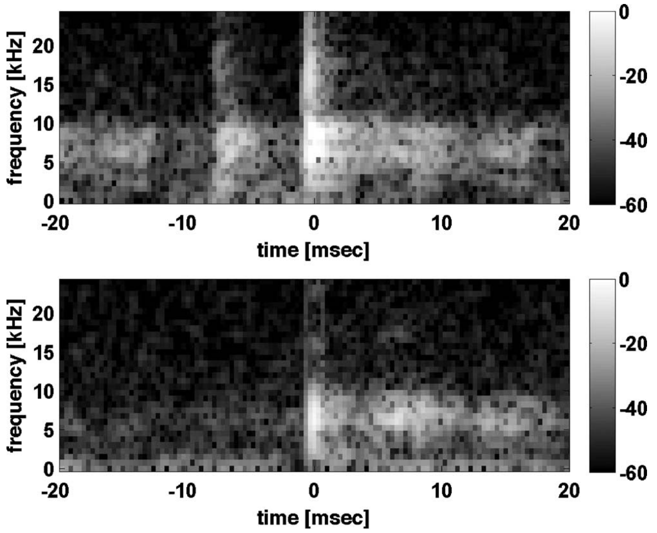


FIG. 6. Representative spectrogram of (a) standard click (42 s in Fig. 5) and (b) ambiguous click (55 s in Fig. 5).

the camera. Figure 4 was thus used to window clicks to those that fit this ICI pattern, and thus the on-camera whale, leaving 1178 clicks.

Figure 5 displays the resulting stacked plot of the absolute values of the Hilbert transform of the filtered click time series for the May 31st sequence, following the procedure in Sec. II C, with key events labeled using overlying letters that correspond to the images from Fig. 3. As previously discussed, each transform has been time-shifted so that the signal maximum (B pulse) aligns with the time origin.

One aspect of Fig. 5 that attracts instant attention is the sudden disappearance of either the A or B pulse from the train of clicks between 45 and 75 s, which corresponds closely to the times that the depredating whale appears on-camera. In the rest of this discussion, click samples from within this time range are dubbed “ambiguous” clicks, because of the ambiguity in associating the missing pulse with either the P_0 or P_1 pulse of the bent-horn hypothesis. All other clicks outside this time range are dubbed “standard” clicks. Figure 6 shows representative spectrograms of ambiguous and standard clicks. From this point on, the body length analysis uses only standard click samples.

As stated in Sec. II C, two cepstra were computed from each standard click: one only containing the A and B pulses, and one only containing the B and C pulses. The individual click cepstra then needed to be averaged to produce a clear IPI peak, as was also found by Teloni *et al.*[20] Cepstral averages over 5 and 10 s time intervals were examined, but reliable cepstral estimates were only obtained by averaging all standard cepstra.

Figure 7(a) displays the stacked cepstra generated from the entire click, Fig. 7(b) displays the average cepstrum using only the [A-B] portion of the standard clicks, and Fig. 7(c) displays the average cepstrum using only the [B-C] portion of the standard clicks. In Fig. 7(b), a weak local maximum at 6.3 msec is discernable, while in Fig. 7(c) a strong local maximum at 8.15 ms is clearly visible.

Table I summarizes all IPI estimates extracted from the standard clicks, derived using both Hilbert transform and cepstral methods. Bootstrap methods were used to estimate the variance of the cepstral estimates, by randomly selecting 850 individual cepstra from the appropriate set of click types, averaging, and measuring the peak.

C. Body length estimation using visual and acoustic methods

In Fig. 3(f) the entire whale’s head is visible in the image, oriented perpendicularly with respect to the camera, from the tip of the nose to the point where the lower jaw enters the head. The point where the jaw touches the line is 3.66 m from the camera.

Figure 8 displays the superposition of Fig. 3(f) and an image from the pool calibration, from which the distance between the snout tip to the best estimate of gape angle (SG) is 3.8(±0.1) m, and the distance between the snout tip to the lower jaw tip (SJ) is 1.07(±0.04) m. The mean tooth separation is 0.100(±0.02) m with an uncertainty of 0.005 m on the measurements. These estimate uncertainties arise from both uncertainties in the exact location of the gape in the image, as well as distortion effects. To check how these measurements were affected by the specific image chosen, the measurements were repeated on two other video images extracted 0.5 s before and after Fig. 3(f), when the whale is not exactly perpendicular to the camera. The resulting total length estimates lie within 0.12 m of the results obtained from Fig. 3(f).

Figure 9, derived using whaling data from Ref. 21, relates SG, SJ, and BE measurements as a function of total whale length. As mentioned above, the BE measurement provides an estimate of spermaceti organ length, and the SG measurement provides an upper bound on this organ’s length. A 3.8(±0.1) m SG length translates into a total body length estimate of 15.2(±0.3) m, while a 1.07(±0.04) m SJ length translates into a total body length estimate of 15.25(±0.06) m, a good agreement. The resulting BE length

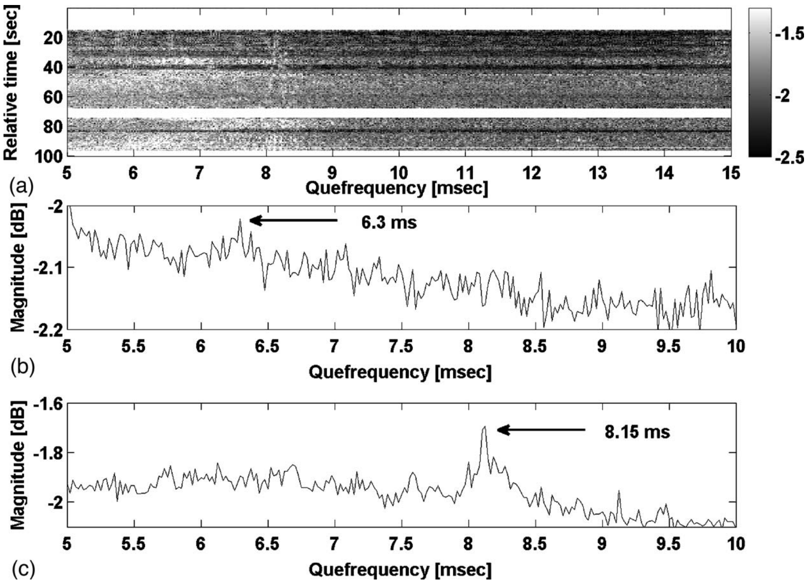


FIG. 7. (a) Stacked cepstra of May 31st sequence, computed over the set of standard clicks (0–48 s in Fig. 5); (b) averaged cepstrum of the [A-B] portion of the standard click samples; (c) averaged cepstrum of the [B-C] portion of the standard click samples.

estimate corresponds to a 3.4(±0.1) m spermaceti organ. As mentioned in Sec. II D, concerns about the accuracy of the SG measurement prompted additional measurements of the ratio between the derived body length and mean tooth separation. This dimensionless ratio was 152(±38), and the ratios derived from the skeletons described in Sec. II D are 179(±27), 160(±23), and 168(±16). Thus the SJ and SG measurements yield virtually identical results for total length, and the teeth measurements produce the same length estimate to within experimental error, alleviating concerns that the SG measurement might yield length estimates that are 40%–50% below the actual value.

Both the Gordon and the Rhinelander and Dawson polynomials, Eqs. (1) and (2), are combined with the two IPI estimates (Table I) to yield four acoustic estimates of the body length. The acoustic estimate obtained by applying the Rhinelander and Dawson formula to the B-C measurements fits the visual estimate to within a meter, as would be expected from previous literature. An acoustic estimate of the spermaceti organ size can be derived, assuming that the B-C interval corresponds to the P_1-P_2 interval of the bent-horn hypothesis, which in turn represents the two way travel time within the spermaceti organ. An estimate of the junk size can then be derived by assuming the A-B interval corresponds to the P_0-P_1 interval, which yields the combined acoustic path length of the spermaceti organ and junk.

TABLE I. Body length estimation for May 31st sequence: comparison between visual and acoustic data.

	IPI (ms)	Body length Eq. (1) (m)	Body length Eq. (2) (m)	Anatomical dimension (m)	Body length Fig. 9 (m)	Anatomical description
			Acoustic measurements			
B-C component, Hilbert	7.6 (±0.9)	15.9 (±1.1)	15.1 (±1.0)	5.23 (±0.6)		Spermaceti organ
B-C component, cepstral	8.1 (±0.2)	16.5 (±0.3)	15.8 (±0.3)	5.55 (±0.3)		Spermaceti organ
A-B component, Hilbert	6.4 (±0.8)	14.1 (±1.0)	13.4 (±1.1)	8.8 (±1.0)		Spermaceti+junk
A-B component, cepstral	6.3 (±0.5)	13.9 (±0.5)	13.3 (±0.6)	8.6 (±0.6)		Spermaceti+junk
			Video measurements			
SG estimate				3.8 (±0.1)	15.2 (±0.3)	Tip of snout to angle of gape
SJ estimate				1.07 (±0.04)	15.25 (±0.06)	Tip of snout to tip of lower jaw
BE estimate				3.4 (±0.1)		Blowhole to center of eye
Teeth estimate				0.100 (±0.02)	15.2 (±0.3)	Mean teeth separation

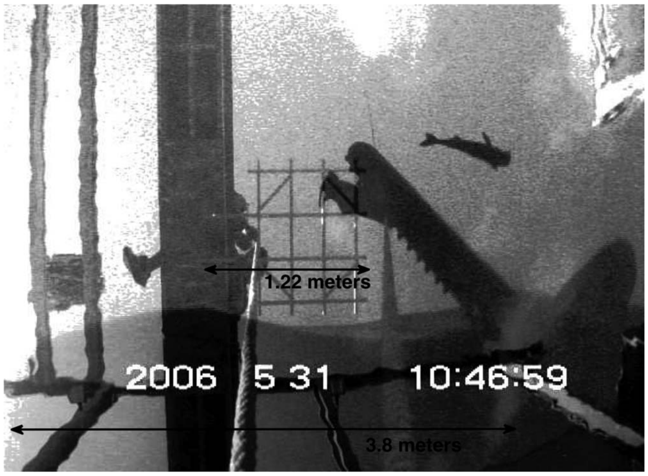


FIG. 8. Superimposition of the videocamera image [Fig. 3(f)] with a pool calibration image.

Table I compares the derived body lengths for both the visual and acoustic methods (the latter using standard clicks only). A 8.1 ms B-C interval and 6.3 ms A-B interval yields respective spermaceti and junk lengths of 5.55 and 3.1 m.

IV. DISCUSSION

A. Relationship between IPI and anatomical structure

Table I shows that both the cepstral and Hilbert IPI estimation procedures produce similar results, in that they find that the A-B IPI is at least 1 ms shorter than the B-C interval. When this IPI difference is applied to either Eq. (1) or (2), the different IPI values yield a 2 m difference in whale length. The body length computed from the B-C time interval, using the Rhinelander and Dawson polynomial [Eq. (2)] matches the body length estimated from the video (15.2 m) to within the 0.6 m experimental uncertainty, which indicates that the B-C "nominal" IPI (P_1-P_2 interval) must have dominated the IPI measurements used to derive the original polynomial fit. Equation (2) would be expected to provide a better fit than Eq. (1), as all individuals from Rhinelander and Dawson dataset were larger than 12 m, whereas Gordon's regression contained only one individual longer than 12 m. Thus to this point, the observations are consistent with the bent-horn hypothesis.

However, the videocamera data raise some questions about the physical interpretation of the IPI in the animal. Under the bent-horn hypothesis, the P_1-P_2 interval should represent the two-way travel time of a sound pulse inside the spermaceti organ; thus the acoustic measurements indicate a spermaceti organ on the order of 5.5 m [assuming a propagation speed of 1370 m/s (Refs. 13 and 26)]. However, the visual estimate of the spermaceti organ length, using the BE measurement from the image, is only 3.4 m (Table I and Fig. 9), which is only 60% of the acoustic estimate. One simple explanation for the discrepancy is that the BE measurement (from the blowhole to the center of the eye) is a biased estimate of spermaceti organ length. Whether that bias would be so large to yield almost a factor of 2 error is unknown.

Another anatomical puzzle is that the 3.1 m length of the junk derived by Sec. III C is shorter than the 5.55 m spermaceti organ length, which is in contradiction with known anatomical relationships.[41] Were the museau de singe physically further from the camera than the acoustic exit point from the junk, this discrepancy might be explained; however, the physical orientation of the whale relative to the camera indicates that the relative distance between the two expected acoustic exit points should not be a large effect.

In summary, the combined video/acoustic measurements suggest that the IPI of the sperm whale can be related to the total length of the animal, as has been demonstrated empirically elsewhere many times. However, the video observations are inconsistent with standard interpretations of the propagation paths through the animal's head; namely, that the P_1-P_2 interval is a direct measure of the size of the spermaceti organ.

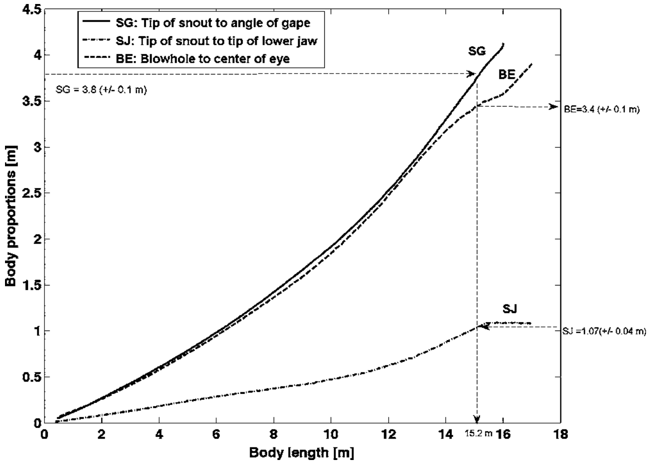


FIG. 9. Allometric relationships between body proportions of male sperm whales caught in the North Pacific (modified from Ref. 21) Solid line: SG, distance from tip of snout to angle of gape; dot-dash line: SJ, distance from tip of snout to tip of lower jaw; dashed line: BE, distance from blowhole to center of eye. Both SG and SJ were measured directly from the image. The BE measurement is a proxy for spermaceti organ length.

B. Unresolved questions

The dataset shown in Fig. 5 shows some additional puzzling features that are worth mentioning briefly. The first is the lack of a so-called $P_{1/2}$ pulse, observed in detailed measurements elsewhere.[19] The orientation of the animal captured by the videocamera would suggest a clear time separation between the P_0, $P_{1/2}$, and P_1 pulses, yet our results indicate no trace of this additional propagation path.

Even more puzzling is the temporal structure of the so-called ambiguous clicks, visible between the 48 and 80 s time window in Fig. 5. Out of the two longline encounters captured by videocamera, this type of click structure only appears in the May 31 encounter. The ambiguous clicks are intriguing because they are produced when the whale is biting the line next to a fish, and have a highly variable ICI (Fig. 4). Why would a whale continue to creak or "buzz" while biting the line, when the presumed targets of interest are off-axis of the presumed sonar beam and actually behind the museau de singe?

Furthermore, the ambiguous clicks are missing a pulse, when compared with the standard clicks that are the focus of this study. However, a cepstral analysis of this click subset shows weak peaks at 6.37 and 8.1 ms, the same IPIs present in the standard clicks. A close visual inspection of the Hilbert transforms of Fig. 5 does seem to indicate weak local maxima arriving at about 6 and 8 ms after the arrival of the intense, temporally compact main pulse at 0 ms.

There a variety of explanations as to what could be happening, but given a sample size of 1, such speculation is premature. Additional data will be needed to determine if these ambiguous clicks appear consistently during an actual depredation event.

V. CONCLUSION

Video and audio recordings of a prey acquisition attempt from depredating sperm whales in the Gulf of Alaska have been processed to compare visual and acoustic methods for estimating animal size. The IPI between the B and C pulses (interpreted as the P_1 and P_2 pulses in the bent-horn hypothesis) yields a size estimate that matches visual estimates to within experimental error, a result consistent with previous empirical studies. However, the video data also permit bounds to be placed on the size of the spermaceti organ of the animal, and those results suggest that the IPI might not be a direct measure of the size of this organ. Furthermore, the size of the junk derived from the acoustic data is smaller than the estimated size of the spermaceti organ, which is inconsistent with anatomical fact.

This paper has provided a glimpse of a possible new approach for investigating the biosonar of large marine mammals in the wild, which permits close-range measurements of the acoustic structure of the terminal buzz or creak of the animal, from both a broadside and on-axis orientation that lies within the main beam of the animal. These measurement locations are unavailable to bioacoustic tags, which can only measure sounds from orientations behind the animal. Only future work will tell whether close-range observations of depredating whales can yield sample sizes sufficient to draw additional conclusions about the biosonar properties of sperm whales, and whether echoes from prey items might be identified.

ACKNOWLEDGMENTS

The authors thank Marty, Tamantha and Jesse for allowing us to dedicate a portion of their black cod quota to the experiment. We also thank the videocamera analysis efforts of Amanda Koltz, the feedback of Ted Cranford and two anonymous reviewers, whose comments have substantially improved the paper's quality and focus. This work was supported by Grant No. 7973-06 from the National Geographic Society and North Pacific Research Board grant 626. With regard to potential regulatory issues, the authors consulted with the Permits, Conservation and Education Division, Office of Protected Resources, National Marine Fisheries Service (NMFS), which is processing the pending amendment to NMFS research Permit No. 473-1700-02, to cover future camera work that may involve more substantial modifications of longlining fishing techniques. The authors also thank the Mammalogy Collections Manager Jim Dines and the Curator Dave Janiger from the Natural History Museum of Los Angeles County as well as the Collections Manager Tony Dumitru and the Curator Ben Simons from the Nantucket Historical Association Whaling museum for their help with collecting the tooth spacing measurements.

[1] W. T. Fitch, "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222 (1997).

[2] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).

[3] T. Nearey, *Phonetic Features for Vowels* (Indiana University Linguistics Club, Bloomington, 1979).

[4] P. Lieberman, *The Biology and Evolution of Language* (Harvard University Press, Cambridge, MA, 1984).

[5] G. E. Petersonand and H. L. Barney, "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **24**, 175–184 (1952).

[6] H. Whitehead, *Sperm whales: Social Evolution in the Ocean* (University of Chicago Press, Chicago, IL, 2003).

[7] N. Jaquet, S. Dawson, and L. Douglas, "Vocal behavior of male sperm whales: Why do they click?," J. Acoust. Soc. Am. **109**, 2254–2259 (2001).

[8] S. L. Watwood, P. J. O. Miller, M. P. Johnson, P. T. Madsen, and P. L. Tyack, "Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*)," J. Anim. Ecol. **75**, 814–825 (2006).

[9] P. J. O. Miller, M. P. Johnson, and P. L. Tyack, "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey," Proc. R. Soc. London, Ser. B **271**, 2239–2247 (2004).

[10] K. S. Norris and G. W. Harvey, "A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon L.*)," in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville, NASA Special Publication No. **262**, Washington, DC, pp. 397–417, 1972.

[11] J. C. D. Gordon, "Behavior and ecology of sperm whales of Sri Lanka," Ph.D. thesis, University of Cambridge, Cambridge, UK (1987).

[12] H. Whitehead and L. Weilgart, "Patterns of visually observable behavior and vocalizations in groups of female sperm whales," Behaviour **118**, 275–296 (1991).

[13] J. C. Goold and S. E. Jones, "Time and frequency domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**, 1279–1291 (1995).

[14] A. Thode, D. K. Mellinger, S. Stienessen, A. Martinez, and K. Mullin, "Depth-dependant features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," J. Acoust. Soc. Am. **116**, 245–253 (2002).

[15] B. Mohl, E. Larsen, and M. Amundin, "Sperm whale size determination: Outlines of an acoustic approach," Fisheries Series, FAO Rome, 1981, pp. 327–332.

[16] B. Mohl, "Sound transmission in the nose of the sperm whale, *Physeter Catodon*. A post mortem study," J. Comp. Physiol. **187**, 335–340 (2001).

[17] B. Mohl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund, "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154 (2003).

[18] W. M. X. Zimmer, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **118**, 3337–3345 (2005).

[19] W. M. X. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson, and P. L. Tyack, "Off-axis effects on the multi-pulse structure of sperm whale usual clicks with implications for the sound production," J. Acoust. Soc. Am. **118**, 3337–3345 (2005).

[20] V. Teloni, W. M. X. Zimmer, M. Wahlberg, and P. T. Madsen, "Consistent acoustic size estimation of sperm whales using clicks recorded from unknown aspects," J. Cetacean Res. Manage. **9**, 127–136 (2007).

[21] N. Nishiwaki, S. Oshumi, and Y. Maeda, "Changes in form of the sperm whale accompanied with growth," Sci. Rep. Whales Res. Inst. **17**, 1–13 (1963).

[22] M. R. Clarke, "Structure and proportions of spermaceti organ in sperm whale," J. Mar. Biol. Assoc. U.K. **58**, 1213–1222 (1978).

[23] J. C. D. Gordon, "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*), from their vocalisations," J. Zool. (London) **224**, 301–314 (1991).

[24] M. Q. Rhinelander and S. M. Dawson, "Measuring sperm whales from

their clicks: Stability of interpulse intervals and validation that they indicate whale length," J. Acoust. Soc. Am. **115**, 1826–1831 (2004).

[25] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The frequency analysis of time series for echoes; cepstrum pseud-autocovariance cross-cepstrum and shape-cracking," Symposium on Time Series Analysis, edited by M. Rosenblatt (Wiley, New York, 1963), Vol. **15**, pp. 209–243.

[26] J. C. Goold and S. E. Jones, "Sound velocity measurements in spermaceti oil under the combined influences of temperature and pressure," Deep-Sea Res., Part I **43**, 961–969 (1996).

[27] M. Wahlberg, "The acoustic behaviour of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**, 53–62 (2002).

[28] W. A. Watkins, M. A. Daher, N. A. DiMarzio, A. Samules, D. Wartzok, K. M. Fristrup, P. W. Howey, and R. R. Maiefski, "Sperm whale dives tracked by radio tag telemetry," Marine Mammal Sci. **18**, 55–78 (2002).

[29] M. R. Clarke and N. Macleod, "Cephalopod remains from sperm whales caught off Iceland," J. Mar. Biol. Assoc. U.K. **56**, 733–750 (1976).

[30] D. W. Rice, "Sperm Whales," in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London, 1989), Vol. **4**, pp. 177–233.

[31] J. R. Ashford, P. S. Rubilar, and A. R. Martin, "Interactions between cetaceans and longline fishery operations around South Georgia," Marine Mammal Sci. **12**, 452–457 (1996).

[32] D. Capdeville, "Interaction of marine mammals with the longline fishery around the Kerguelen Island Division, 58.5.1 during the 1995/96 cruise," Ccamlr Sci. **4**, 171–174 (1997).

[33] C. P. Nolan and G. M. Liddle, "Interactions between killer whales (*Orcinus orca*) and sperm whales (*Physeter macrocephalus*) with a longline fishing vessel," Marine Mammal Sci. **16**, 658–664 (2000).

[34] E. F. Gonzalez, presented at the XXI Congreso de Ciencias del Mar, Chile (2001).

[35] P. S. Hill, J. L. Laake, and E. Mitchell, "Results of a pilot program to document interactions between sperm whales and longline vessels in Alaska waters," U.S. Department of Commerce, Alaska Fisheries Science Center, Report No. NOAA, TM-NMFS-AFSC-108, 42 pp., 1999.

[36] R. Hucke-Gaete, C. A. Moreno, J. Arata, and Blue Whale Ctr, "Operational interactions of sperm whales and killer whales with the Patagonian toothfish industrial fishery off Southern Chile," Ccamlr Sci. **11**, 127–40 (2004).

[37] M. G. Purves, D. J. Agnew, E. Balguerias, C. A. Moreno, and B. Watkins, "Killer whale *Orcinus orca* and sperm whale *Physeter macrocephalus* interactions with longline vessels in the patagonian toothfish fishery at South Georgia, South Atlantic," Ccamlr Sci. **11**, 111–126 (2004).

[38] M. F. Sigler, C. R. Lunsford, J. M. Straley, and J. B. Liddle, "Sperm whale depredation of sablefish longline gear in the northeast Pacific Ocean," Marine Mammal Sci. **24**, 16–27 (2008).

[39] A. Thode, J. Straley, C. O. Tiemann, K. Folkert, and V. O'Connell, "Observations of potential acoustic cues that attract sperm whales to longline fishing in the Gulf of Alaska," J. Acoust. Soc. Am. **122**, 1265–1277 (2007).

[40] J. Bouguet, Camera Calibration Toolbox for Matlab (2008).

[41] T. Cranford, "The sperm whale's nose: Sexual selection on a grand scale?," Marine Mammal Sci. **15**, 1133–1157 (1999).

[42] M. Amundin, "Sound production in odontecetes with emphasis on the harbour porpoise *Pbocoena pbocoena*," Ph.D thesis, University of Stocklom, Stocklom, Sweden (1991).

[43] P. T. Madsen, M. Wahlberg, and B. Mohl, "Male sperm whale (*Physeter macrocephalus*) acoustics in a high latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41 (2002).

Model predicts bat pinna ridges focus high frequencies to form narrow sensitivity beams

Roman Kuc[a)]
Intelligent Sensors Laboratory, Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520-8284

(Received 28 December 2007; revised 19 December 2008; accepted 18 February 2009)

The pinnae of bats contain ridges whose function was previously thought to be structural. This paper suggests that ridges form a reflecting Fresnel lens that focuses high-frequency acoustic signals into the ear canal to form a narrow elevation sensitivity beam. *E. fuscus* ridges are modeled as a series of four paraboloidal strips and the tragus is considered to act as a secondary reflecting element analogous to a Cassegrain system. A diffraction grading having the equivalent spacing suggests frequencies above 150 kHz. Using an example 170 kHz, a random search for ridge dimensions that minimize side-lobes in the frequency magnitude response yields the tapered ridge structure observed in *E. fuscus* and produces an 18° (full width half energy) beam width. We speculate that the possible high-frequency sources are ecologically (prey) generated and/or the third harmonic of the call. The attenuation at such high frequencies requires that the source be close by. Passive prey localization in the postbuzz stage, when echoes overlap call transmissions and the prey is within 8 cm, could improve prey capture efficiency. An experiment using 40 kHz ultrasound with human observers verifies that frequencies beyond the audiometric range, when sufficiently intense, can still be perceived. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3097500]

PACS number(s): 43.80.Lb, 43.80.Gx, 43.66.Gf, 43.60.Fg [JAS] Pages: 3454–3459

I. INTRODUCTION

Many bat species have pinnae that contain prominent ridge patterns.[1] For example, Fig. 1 shows the pinnae of *E. fuscus*, the big brown bat that uses echolocation for aerial pursuit of prey. Among the obvious features in the pinna are four ridges along the inside pinna surface and the tragus, a small flap in front of the ear canal (ec). Initial models considered the pinna as an obliquely-truncated horn[2] that has a frequency-dependent directionality.[3] But these models do not typically include ridges. It was previously thought that these ridges provide structural support.[4] Recent numerical models of bat pinnae[5] consider ridges to form a diffraction grating. This paper extends the diffraction grating model by widening diffracting lines into more efficient reflecting surfaces, suggesting a Fresnel lens. The ridges then focus an incoming acoustic signals onto the ec.

A diffraction grating analysis is a natural first-choice when observing a series of periodic parallel structures. For first-order diffraction along angle θ, the spacing d of the grating is related to wavelength λ as follows:[6]

$$\lambda = d \sin \theta. \tag{1}$$

Hence, the maximum wavelength at which diffraction occurs is $\lambda_{\max}=d$, providing a clue as to the frequencies that can be employed by a ridge structure. Published data report the average pinna height of *E. fuscus* to be approximately 14 mm.[7] Each pinna typically contains four ridges spanning approximately half the pinna height. A diffraction grading spanning 7 mm, with lines at the limits, has spacing $d=7\text{ mm}/3 =2.3\text{ mm}$. Applying the diffraction grating criterion and using the speed of sound $c=343\text{ m/s}$, we find the minimum frequency that produces first-order diffraction equals the surprisingly high value

$$f_{\min} = \frac{c}{\lambda_{\max}} = 149\text{ kHz}. \tag{2}$$

This frequency is greater than the 100 kHz that most bat researchers examine for *E. fuscus*, or most bats for that matter. Typical spectrograms of *E. fuscus* calls[8–11] show a fundamental that sweeps from 60 to 30 kHz and its second harmonic from 120 to 60 kHz, although the third harmonic is also often observed that would sweep from 180 to 90 kHz. The higher-frequency limit of the third harmonic is well above $f_{\min}$, motivating us to consider such high frequencies in more detail.

If the reader is skeptical about the third harmonic frequency being sufficiently intense, we offer an alternative. The ridge spacing may be indicating that prey generates such high frequencies. In this case, the ridges could help the bat localize the prey in the postbuzz stage, within 8 cm range, during which the echolocation calls would overlap with the echoes.[12] Such small ranges are also dictated by the increased attenuation at such high frequencies, on the order of 10 dB/m.[13]

Although the source is not certain, this paper assumes that the ridges operate at a high frequency, denoted $f_R \geqslant f_{\min}$, and examines how a ridge structure can form a narrow sensitivity beam.

The paper is organized in the following manner. Section III describes a paraboloid model of reflecting ridges. Section III applies this model to construct an optimized ridge structure in the *E. fuscus* pinna. Section IV determines the sensitivity beam width to be 16°. Section V describes an experiment that verifies the perception of frequencies that are

[a)] Author to whom correspondence should be addressed. Electronic mail: kuc@yale.edu

FIG. 1. The big brown bat *Eptesicus fuscus* has pinnae with prominent ridge structures. Image used with permission © Merlin D. Tuttle, Bat Conservation International.

beyond the audiometric range. Section VI discussed the advantages of the Fresnel structure.

II. RIDGE MODEL

The ridge model starts by considering how an incoming planar acoustic wavefront can be focused onto a point, corresponding to the entrance to the ec. Consider a planar wavefront moving along the x direction toward the origin of a Cartesian coordinate system shown in Fig. 1. The simplest structure that focuses a planar wavefront is a paraboloid formed by rotating a parabola about its axis. The parabola with vertex at $(x,y)=(0,0)$ and focus at $(F,0)$, with ec located at F, is described by the equation

$$y^2=4Fx. \tag{3}$$

A reflecting surface, corresponding to a ridge, on the paraboloid also has this focusing property, although the focal spot size at the ec increases as the reflecting surface area decreases, as described below. The paraboloids in the figure show rectangular patches near the top that represents a pair of ridges.

To generate the multiple ridge structure we take similar patches from additional parabolas that focus the incoming waves at $(F,0)$ but are shifted axially to the left by $\lambda_R/2$, where $\lambda_R=c/f_R$. This shift causes sinusoidal frequency components at f_R to add in phase at $(F,0)$. This effect can be most easily seen by considering the component traveling along the axis. The kth parabola is described by the equation

$$y^2=4\left(F+\frac{(k-1)\lambda_R}{2}\right)\left(x+\frac{(k-1)\lambda_R}{2}\right). \tag{4}$$

Figure 2 shows paraboloids for $k=1$ (dashed line) and $k=2$ (solid line).

With the pinna extending above the ec, we consider only the top halves of the paraboloids. Figures 3(a) and 3(b) show cuts along the vertical plane passing through the axis and the top half of four such paraboloids $(k=1,2,3,4)$, each associated with a ridge in *E. fuscus*. The top-most ridge, being the furthest from the ec, introduces the greatest delay and the bottom-most ridge the smallest. Dashed lines show geometric paths to the ec that delimit the ridge reflecting surfaces.

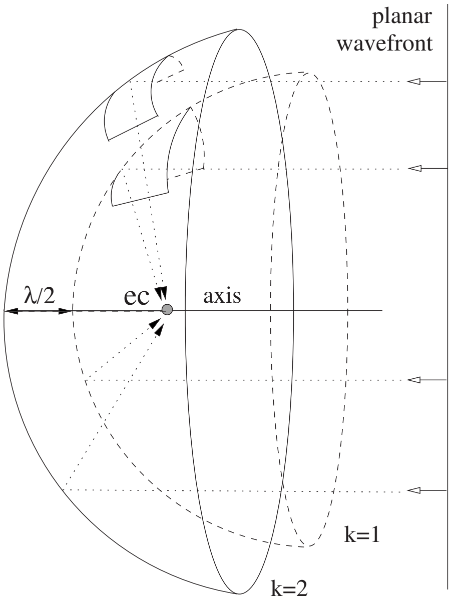


FIG. 2. From paraboloid to pinna ridges. (a) A pair of paraboloids axially shifted by $\lambda/2$ both focus incident acoustic planar wavefronts onto the ec. The two rectangular patches shown at top correspond to two ridge reflecting areas.

To explain the trade-offs in ridge construction, we model each ridge as generating three scattered components. The first is the desired reflected component from the paraboloidal surface, which focuses onto the ec. The reflecting strength associated with this component is proportional to the ridge area, equal to height shown in the figure times the wraparound width. This is most easily understood using Huygen's principle argument that constructs a surface using a collection of small reflecting elements.[14]

The second component is the reflection from the tissue connecting the ridges. It will reflect energy away from the ec, and this loss of incident acoustic energy reduces the efficiency of the ridge structure compared with a single paraboloidal surface. The third includes the diffracted components from the ridge edges. Since diffraction from such small structures is omnidirectional this component does not contribute to the focusing and thus also reduces the efficiency of the ridge structure, but only by a small amount.

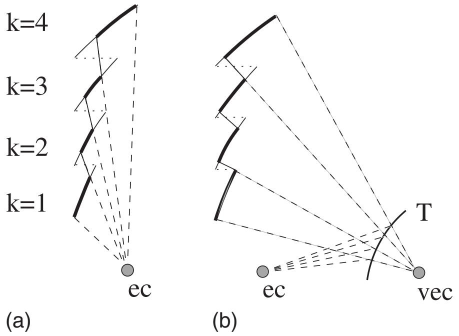


FIG. 3. Vertical cross-sections through four paraboloidal patches model *E. fuscus* pinna containing four ridges. Dotted lines show horizontal connections between patches to construct ideal reflecting Fresnel lens. Dashed lines show geometric paths from ridges to ec. Heavy lines show corresponding ridge heights. Light lines show Fresnel lens heights that are lost to obtain geometric paths from ridges to ec. (a) Direct reflection into the ec. (b) Tragus acting as a secondary reflector as in a Cassegrain system results in more efficient Fresnel lens structure.

The tissue sections connecting the paraboloidal ridges require a design trade-off that various bat species seem to resolve in different ways: If the connections were parallel to the axis, along the wavefront propagation direction, shown as horizontal dotted lines in the figure, we obtain a classical Fresnel lens having a large summed ridge reflecting area comparable to a single paraboloid having the same outer dimensions. Figure 3 shows this classical structure in solid line (both heavy and light weights), connected with dotted lines. However, when focused onto the ec, the lower ridges would then block part of the reflected signal from the higher ridges to the ec, thus reducing the summed signal at the ec. Hence, we need to consider the ridge areas that are *visible* from the ec, and these are delimited by the geometric paths shown in dashed line. The efficiency of the reflecting acoustic Fresnel lens compared to a single paraboloid is related to ratio of the visible ridge area to the area of the ideal Fresnel lens.

Figure 3(b) indicates that this visible ridge area increases as ec distance from the apex increases because the dashed geometric paths approach the dotted horizontal lines. But a distant ec presents aerodynamic and other problems. *E. fuscus*, and other bats having a prominent tragus, apparently have come up with an elegant solution by incorporating a reflecting tragus. The presence of a reflecting tragus having a hyperbolic cross-section would then form a Cassegrain lens system.[15] The effect of a reflecting tragus can be thought of as introducing a distant virtual ec, thus allowing the actual ec to be tucked into the pinna base. A reflecting tragus then uses the ridge reflecting surfaces efficiently.

The additional pinna sections above the top ridge and below the bottom ridge should not contribute to the focusing process because their large areas would dominate the sum. Figure 1 clearly shows the pinna surface above the ridges and it appears that this surface reflects high-frequency echoes away from the ec. The pinna section below the lowest ridge is typically not visible, but we speculate that its shape serves to further improve the signal at the ec.

Section III applies these acoustic principles to interpret the ridge structure observed in the *E. fuscus* pinna.

III. OPTIMIZED *E. fuscus* RIDGE STRUCTURE

The *E. fuscus* pinna typically contains four ridges. For example, consider an *E. fuscus* ridge structure that is designed to match $f_R = 170$ kHz, a component present in the third harmonic of its call or radiated by close-by prey. The spacing between adjacent paraboloids then equals

$$\frac{\lambda_R}{2} = \frac{c}{2f_R} = 1.0 \text{ mm}. \tag{5}$$

This dimension is consistent with the *E. fuscus* ridge structure.

We model this ridge structure as a tapped acoustic delay line filter with an input $p(t)$ being planar wavefronts approaching along the axial direction. For analytic tractability we assume that each ridge produces a component at the ec that has essentially the same waveform, denoted $p_{\text{ec}}(t)$, but introduces its own scale factor and delay. If ridge k has reflecting strength a_k, proportional to its ridge area, and introduces delay τ_k, the summed pulse at the ec from the four ridges can then be expressed as

$$p_{\text{sum}}(t) = \sum_{k=1}^{4} a_k p_{\text{ec}}(t - \tau_k), \tag{6}$$

where $k=1$ corresponds to the lowest ridge.

In designing a ridge structure, an obvious *first-choice* would be to set all the reflecting strengths equal

$$a_1 = a_2 = a_3 = a_4 = \tfrac{1}{4} \tag{7}$$

by making the ridge areas equal, and setting the delays so that the f_R components add in phase at the ec

$$\tau_k = (k-1)/f_R = (k-1)\tau_R \quad \text{for } k = 1,2,3,4. \tag{8}$$

This delay assignment removes the delay common to all reflected signals, which is inconsequential in examining the effect at the ec.

However, this first-choice design is not consistent with the *E. fuscus* ridge structure in Fig. 1, which clearly shows that the second and third ridges have greater areas than the first and fourth (top-most) ridges. This indicates that a_2 and a_3 are greater than a_1 and a_4 in the bat ear. To understand why this should be the case, we model the impulse response of the ridge structure as a set of four appropriately weighted and delayed impulses

$$h_R(t) = \sum_{k=1}^{4} a_k \delta(t - \tau_k), \tag{9}$$

where $\delta(t)$ is the Dirac delta function that focusing produces. The frequency transfer function of ridge structure is the Fourier transform and equals

$$H_R(f) = \int_{-\infty}^{\infty} h_R(t) e^{-j2\pi f t} dt. \tag{10}$$

Applying the first-choice values for a_k and τ_k yields

$$H_R^{(1)}(f) = e^{-j3\pi f \tau_R} \left(\frac{\sin(4\pi f \tau_R)}{\sin(\pi f \tau_R)} \right). \tag{11}$$

Figure 4(b) shows that the plot of the magnitude response $|H_R^{(1)}(f)|$ (dotted line) has peaks at $f = nf_R$ for integer n. We expect the peak at $f = f_R$ because the ridge structure design is matched to it. The peak at $f = 0$ is inconsequential acoustically because the ridges are not efficient reflectors for wavelengths that are much greater than the ridge dimensions, treated further below. We assume that the peaks at the higher frequencies $(n > 1)$ are not important because even if high-frequency energy were present, it would attenuate too quickly to be useful. However, the large-magnitude side-lobes, or the maxima not at $f = nf_R$, are troublesome because energy components at those frequencies would contribute to the sum at the ec, causing ambiguity and disrupting the advantages of a high-frequency ridge structure.

Reducing these side-lobe magnitudes is a well-known problem in antenna array design.[15] The design procedure that

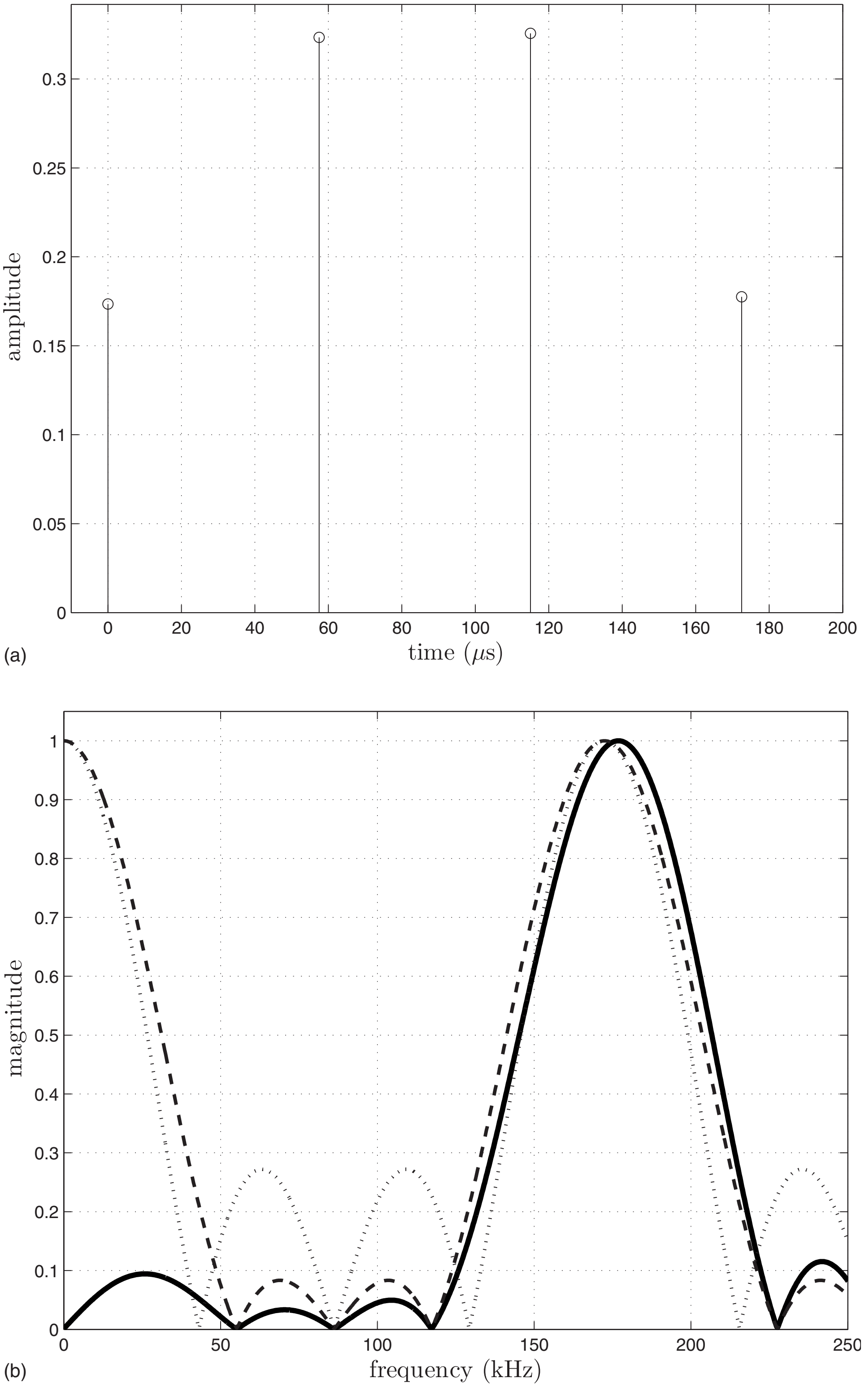


FIG. 4. Ridge impulse response produced by random search and magnitude frequency responses. (a) Tapered impulse response weights show apodization. (b) Frequency responses. Dashed curve: response computed from impulse response in (a). Dotted curve: response if ridge components had equal weights (no apodization). Solid curve: apodized response modified by Rayleigh scattering efficiency correction.

reduces side-lobe magnitudes, at the expense of increasing the main lobe width, is to taper the strengths of the outer reflectors, called *apodization*. This appears to be what the bat ridge structure does!

To illustrate the benefit of apodization, we mimicked evolution and performed a random search to find near-optimal values for the a_k's. In the search we used the pseudorandom number generator in MATLAB to randomly guess a set of positive values for a_1 to a_4. We then normalized these values so they summed to one, thereby forming a partition of the total available ridge area. For each set of guesses we computed the corresponding magnitude response $|H_R^{(g)}(f)|$ using the fast Fourier transform equivalent of Eq. (10). We defined the objective function $\mathcal{O}(\bar{a})$ to equal the ratio of the transfer function area around the peaks to that between the peaks

$$\mathcal{O}(\bar{a}) = \frac{\displaystyle\int_{f_R-\epsilon}^{f_R} |H_R^{(g)}(f)|df}{\displaystyle\int_{f_R/2}^{f_R-\epsilon} |H_R^{(g)}(f)|df}, \tag{12}$$

where $f_R = 170$ kHz and $\epsilon = 40$ kHz. $\mathcal{O}(\bar{a})$ maximizes the transfer of frequencies around f_R while minimizing the transfer at other frequencies. The periodicity of the Fourier transform accommodates the other relevant frequency intervals. While making random guesses for the a_k's, we retained and updated the sets that increased $\mathcal{O}(\bar{a})$. 200 000 guesses yielded 10 sets that improved $\mathcal{O}(\bar{a})$, with the best results produced by the set

$$a_1 \approx a_4 \approx \tfrac{1}{6} \quad \text{and} \quad a_2 \approx a_3 \approx \tfrac{1}{3}. \tag{13}$$

These values are consistent with the ridge structure in Fig. 1 that shows the middle two ridges have larger areas than the top and bottom ridges.

Figure 4(a) shows weights and delays associated with the resulting optimum tapered impulse response, denoted $h_{\text{opt}}(t)$, and Fig. 4(b) shows its magnitude transfer function $|H_{\text{opt}}(f)|$ (dashed line). For comparison, the first-choice $|H_R^{(1)}(f)|$ is shown in dotted line. We note that apodization has reduced the side-lobe levels by a factor of 3.

As mentioned above, the transfer function peak around $f=0$ is de-emphasized by the ridge structure because low frequencies are not reflected efficiently by the ridge areas toward the ec. To model this Rayleigh scattering effect, we formed the correction function that equals 0 at $f=0$ and increases linearly to 1 at $f=170$ kHz given by

$$\text{Rayl}(f) = \frac{f}{170 \text{ kHz}}. \tag{14}$$

The final ridge structure magnitude frequency response then equals

$$|H_R(f)| = \text{Rayl}(f)|H_{\text{opt}}(f)|. \tag{15}$$

Figure 4(b) shows $|H_R(f)|$ in solid line having a single prominent peak near 170 kHz.

IV. ANTENNA GAIN AND SENSITIVITY WIDTH CALCULATIONS

Focusing the acoustic intensity incident on the ridge area to a small spot size at the ec realizes a significant gain. To determine the focusing effect, we approximate the ridge structure with a single paraboloidal patch having a circular boundary of diameter D. We then model this patch as a focused ultrasound transducer of similar diameter and focal length F, commonly used in diagnostic ultrasound.[16] Then, the diffraction-limited focal spot has a −3 dB diameter S, equal to

$$S = \frac{\lambda_R F}{D}, \tag{16}$$

where F is the average distance from the ridges to the ec. One measure of intensity gain G equals the ratio of the antenna aperture area to the focal spot area

$$G = \frac{\pi D^2/4}{\pi S^2/4} = \frac{D^4}{\lambda_R^2 F^2}.$$

Inserting the *E. fuscus* values $D=7$ mm, $\lambda_R=2$ mm, and using the dimensions from Fig. 3(d) to find the distance from the average ridge distance to the virtual ec equal to $F = 8$ mm, we compute $S=2.2$ mm and $G=9.4$. Experimental models that we have constructed verify a focusing operation.

Binaural hearing produces accurate azimuth localization, but poorer elevation localization,[17,18] so we consider the elevation sensitivity beam characteristics using the approximate model above. The reciprocity principle dictates that an acoustic signal originating at the ec and is reflected from the paraboloidal surface forms a beam whose form is the same as the sensitivity beam we are seeking. We assume that the ec is a point source radiating omni-directionally at frequency f_R. The energy contained within the solid angle that is defined by the paraboloidal circular aperture of diameter D is then converted by the reflection to an outgoing beam that consists of initially planar wavefronts. We approximate this radiated signal by that produced by a flat piston of diameter D vibrating in an infinite baffle. Then, the beam pattern in the far-field, or range $r > D^2/(4\lambda) \approx 6$ mm for our *E. fuscus* model, is the familiar Bessel function Airy pattern with first off-axis null located at angle

$$\theta_N = \arcsin(1.22\lambda_R/D).$$

Inserting our *E. fuscus* values yields $\theta_N = \arcsin(0.34) = 20°$. The −3 dB angle, $\theta_{-3\text{ dB}}$, is smaller than the angle to the nulls, or $\theta_{-3\text{ dB}} \approx 0.4\theta_N = 8°$. This results in a 16° full width −3 dB beam width. The apodization increases the beam width slightly.

These calculations are first-order and consider only the ridge structure. Other clever mechanisms, such as incorporating a dynamic tragus, discussed below, and ec resonance, are likely to improve the *perceptual* sensitivity beam pattern, thus providing higher resolution.

V. HIGH-FREQUENCY HEARING EXPERIMENT

In this section we describe an experiment that verifies the perception of frequencies beyond the conventionally accepted audiometric range. The conventional wisdom is that the high frequencies at which the ridge functions are above the audiometric range of bat hearing. We performed the following experiment to demonstrate the perception of such high-frequency energy.

A piezoelectric ceramic transmitter (Murata-MA40S4S) was driven at resonance (39.7 kHz) with a 20 V peak-to-peak sinusoidal waveform. This transmitter produced a specified sound pressure level of approximately 120 dB re 20 μPa at 30 cm range. Clearly, 40 kHz is beyond the human audiometric range. Yet, when the transmitter was aimed into the author's ec a distinct high-frequency sound, around 14 kHz, was perceived. This frequency is at the upper range of the author's hearing range, as determined by a recent hearing examination. When the transducer was oriented away from the ec by more than ≈ ±20°, the approximate transmitter −3 dB beam width, the perceived sound was not heard.

This indicates the transmission beam was formed by 40 kHz rather than a lower-frequency signal at 14 kHz, which would have a wider beam width. Two other observers experienced similar results.

Hence, humans are able to perceive a high-energy 40 kHz sound. The simplest (linear) explanation is that the cochlear transducers, usually modeled as bandpass filters, will still react to higher-frequency components, if those components are sufficiently intense. The high-frequency hair cells, being closest to the excitation frequency, will have the greatest response, explaining why a 14 kHz tone was perceived. We hypothesize that the ridge-ec structure in *E. fuscus* is capable of focusing the incoming sound at 170 kHz to an intensity sufficient to cause the cochlea to perceive its presence.

VI. DISCUSSION

One clear advantage of the Fresnel ridge structure is aerodynamic, in that it collapses the volume required to implement a paraboloidal lens. However, it also acts as a bandpass filter that extracts frequencies around f_R. This is a necessary feature because f_R energy is perceived by the highest-frequency hair cells and would be masked by acoustic energy at lower frequencies. Hence, the ridges form a sensitivity beam governed by dimensions associated with f_R.

A dynamic reflecting tragus may also serve an important beam-forming purpose that may perceptually narrow the beam from the physically determined 16° width. Because the tragus has low mass, it may act to track a radiating prey by quickly wobbling, say, $\pm 10°$ about its base. This wobble would in effect perform a quick scan over a small elevation interval (the beam width computed using physics) to determine the value that maximizes the signal intensity at the ec. The elegance of the ridge structure motivates us to speculate that if there is a behavior that improves capture efficiency the bat will discover and employ it.

VII. SUMMARY

This paper presented a model explaining the echolocation utility of the ridge structure observed in the *E. fuscus* pinna. The ridges, and possibly tragus, form an acoustic Fresnel lens structure that focuses high-frequency energy onto the ec to achieve sufficient intensity to be perceived. The ridge dimensions suggest that it operates at frequencies higher than 150 kHz. Ridges are modeled as a series of four paraboloidal reflecting strips and the tragus is considered to act as a secondary reflecting element to form a novel Cassegrain system. Using the example frequency of 170 kHz, a search for ridge dimensions that minimize side-lobes in the frequency magnitude response yields a tapered ridge structure that is often used to optimize antenna designs and remarkably similar to that observed in *E. fuscus*. We speculated that the possible sources of such high frequencies include ecologically (prey) generated and/or the third harmonic of the call. The high attenuation at 170 kHz, approximately 10 dB/m, requires that the source to be at close ranges. Passive prey localization in the postbuzz stage, when echoes overlap call transmissions and the prey is within 8 cm, could improve prey capture efficiency. An experiment using 40 kHz ultrasound with human observers verifies that frequencies beyond the audiometric range, when sufficiently intense, can still be perceived.

ACKNOWLEDGMENTS

The author acknowledges the help of Kristof Zyskowski of the Yale Peabody Museum collection of bat specimens in examining and comparing bat pinna of many species. Comments and suggestions by colleagues Cindy Moss, Rolf Mueller, Robert Grober, and James Simmons are gratefully acknowledged.

[1] M. Brock Fenton, *Bats*, rev. ed. (Checkmark Books, New York, 2001), pp.30, 32, 43, 58, 72, 97, and 194.

[2] N. H. Fletcher, *Acoustic Systems in Biology* (Oxford University Press, New York, 1992).

[3] U. Firzlaff and G. Schuller, "Directionality of hearing in two cf/fm bats, *Pteronotus parnellii* and *Rhynolophus rouxi*," Hear. Res. **197**, 74–86 (2004).

[4] J. E. Hill and J. D. Smith, *Bats—A Natural History* (University of Texas Press, Austin, TX, 1984).

[5] R. Müller and J. C. T. Hallam, "From bat pinnae to sonar antennae: Augmented obliquely truncated horns as a novel parametric shape model," *Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior*, edited by S. Schaal, A. J. Ijspeert, A. Billard, and S. Vijayakumar (MIT, Cambridge, MA, 2004), pp. 87–95.

[6] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics* 7th ed. (Wiley, New York, 2005).

[7] W. L. Gannon, R. E. Sherwin, T. N. deCarvalho, and M. J. O'Farrell, "Pinna and echolocation call differences between myotis californicus and m. cililabrum (chiroptera: Vespertilionidae)," Acta Chiropterologia **31**, 77–91 (2001).

[8] J. A. Simmons, M. B. Fenton, and M. J. O'Farrell, "Echolocation and pursuit of prey by bats," Science **203**, 16–21 (1979).

[9] S. A. Kick, "Target-detection by the echolocating bat, *eptesicus fuscus*," J. Comp. Physiol. **145**, 431–435 (1982).

[10] S. Parsons, C. W. Thorpe, and S. M. Dawson, "Echolocation calls of the long-tailed bat: A quantitative analysis of types of calls," J. Mammal. **78**, 964–976 (1997).

[11] A. Surlykke and C. F. Moss, "Echolocation behavior of big brown bats, *Eptesicus fuscus*, in the field and the laboratory," J. Acoust. Soc. Am. **108**, 2419–2429 (2000).

[12] R. Kuc, "Sensorimotor model of bat echolocation and prey capture," J. Acoust. Soc. Am. **96**, 1965–1978 (1994).

[13] F. J. Alvarez and R. Kuc, "Dispersion relation for air via Kramers-Kronig analysis," J. Acoust. Soc. Am. **124**, EL57–EL61 (2008).

[14] R. Kuc and M. W. Siegel, "Physically-based simulation model for acoustic sensor robot navigation," IEEE Trans. Pattern Anal. Mach. Intell. **9**, 766–778 (1987).

[15] M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed. (McGraw-Hill, New York, 2001).

[16] P. N. T. Wells, *Biomedical Ultrasonics* (Academic, New York, 1977).

[17] R. Kuc, "Biologically motivated adaptive sonar," J. Acoust. Soc. Am. **100**, 1849–1854 (1996).

[18] R. Kuc, "Biomimetic sonar system recognizes objects using binaural information," J. Acoust. Soc. Am. **102**, 689–696 (1997).

Propagation of two longitudinal waves in human cancellous bone: An *in vitro* study

Katsunori Mizuno, Mami Matsukawa,[a] and Takahiko Otani
Laboratory of Ultrasonic Electronics, Doshisha University, Kyotanabe, 610-0321 Kyoto, Japan

Pascal Laugier and Frédéric Padilla
Université Pierre et Marie Curie-Paris 06, UMR 7623, LIP, F-75005 Paris, France and Laboratoire d'Imagerie Paramétrique, CNRS, UMR 7623, F-75006 Paris, France

(Received 24 December 2008; revised 7 March 2009; accepted 11 March 2009)

The ultrasonic wave propagation of fast and slow waves was investigated *in vitro* in 35 cubic cancellous bone specimens extracted from human femoral heads. Measurements were performed in three orthogonal directions using home-made PVDF transducers excited by a single sinusoidal wave at 1 MHz. The apparent density of the specimens was measured. Two separated fast and slow waves were clearly observed in 16 specimens, mainly in the main load direction. The waveforms and the sound speeds of fast and slow waves were similar to the reported data in bovine bone. The group of specimens in which the two waves were observed did not exhibit statistically higher apparent density than the rest of the specimens, but did exhibit statistically higher acoustic anisotropy ratio. The speeds in the main load direction were higher than those in the other direction. The fast and slow wave speeds were in good agreement with Biot's model, showing an increase with bone volume fraction (BV/TV). The ratio of peak amplitudes of the fast and slow waves nonlinearly increased as a function of BV/TV. These results open interesting perspective for acoustic assessment of cancellous bone micro-architecture and especially anisotropy that might lead to an improved assessment of bone strength. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3111107]

PACS number(s): 43.80.Vj, 43.80.Cs, 43.80.Ev, 43.80.Jz [FD] Pages: 3460–3466

I. INTRODUCTION

Cancellous bone might be affected by osteoporosis, which is a skeletal disease causing loss of bone mass and micro-architectural deteriorations. The current gold standard for *in vivo* bone status assessment is dual x-ray absorptiometry (DXA), which measures the bone mineral density. Because DXA can only catch one aspect of bone properties and because fracture risk is a multi-factor issue [bone mass but also bone quality such as the microstructure and material properties as pointed out by National Institute of Health in 2001 (NIH, 2001)], alternative techniques to DXA have emerged. Among them are quantitative ultrasound (QUS) (Laugier, 2008), which have many advantages such as the portability, low cost, and free ionizing radiation.

Current clinically validated ultrasound devices measured are based on transmission measurements at the heel bone (Krieg *et al.*, 2008). The derived parameters are the speed of sound and the broadband ultrasonic attenuation (slope of the linearly frequency dependent attenuation). These parameters have been reported to be sensitive to mechanical and elastic properties of cancellous bone (Nieh *et al.*, 1997; Han *et al.*, 1997; Bouxsein *et al.*, 1995; Lochmüller *et al.*, 1999; Cheng *et al.*, 1997).

Cancellous bone is a porous medium composed of an inter-connected network of solid rods and plates (called the trabeculae) filled with marrow *in vivo*. Its inhomogeneity makes the interaction between ultrasound and cancellous bone complex. One interesting phenomenon is the possible propagation of two longitudinal waves in cancellous bone (Hosokawa and Otani, 1997, 1998; Kaczmarek *et al.*, 2002; Hoffmeister *et al.*, 2000; Fellah *et al.*, 2004; Cardoso *et al.*, 2003; Hughes *et al.*, 1999) as predicted by poroelastic models (Fellah *et al.*, 2004; Biot, 1956a, 1956b; Hughes *et al.*, 2007; Sabaa and Fellah, 2006; Sabaa *et al.*, 2008) or a multilayer model (Hughes *et al.*, 1999). The potential measurements of the two-wave phenomenon open new perspectives for *in vivo* assessment of bone status (Otani, 2005), but it has not been fully exploited so far.

One relevant bone property that might be assessed using the two-wave phenomenon is the anisotropy of the trabecular micro-architecture known to be a determinant of bone strength (Haïat *et al.*, 2008). It has been shown that cancellous bone exhibits high structural anisotropy responsible for an acoustical anisotropy of the QUS parameters (Hosokawa and Otani, 1998; Hoffmeister *et al.*, 2000; Hughes *et al.*, 2007; Homminga *et al.*, 2003; Glüer *et al.*, 1994; Nicholson *et al.*, 1994). However, few *in vitro* studies have reported on the impact of this structural anisotropy on the propagation of the fast and the slow wave and on their characteristics, especially in human cancellous bone (Fellah *et al.*, 2004; Cardoso *et al.*, 2003; Mizuno *et al.*, 2008; Haïat *et al.*, 2007).

The objective of this work is to investigate experimentally the wave propagation of fast and slow waves in human cancellous bone. The first goal is to confirm the two-wave phenomenon in human cancellous bone and to compare our observations with previous ones obtained in bovine bones.

[a] Author to whom correspondence should be addressed. Electronic mail: mmatsuka@mail.doshisha.ac.jp

0001-4966/2009/125(5)/3460/7/$25.00 © 2009 Acoustical Society of America

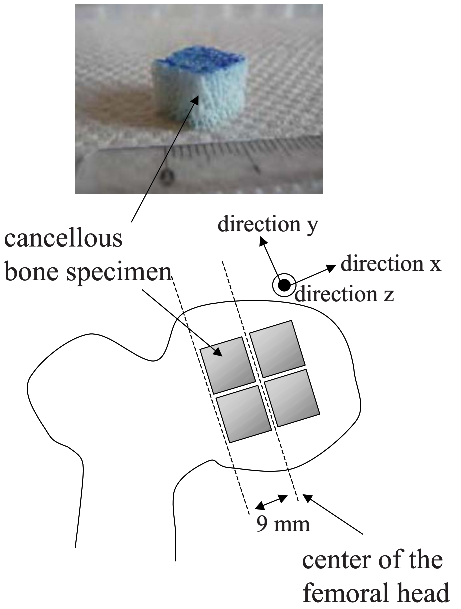


FIG. 1. (Color online) Collection of the cubic specimens from the center of the femoral head (bottom) and picture of a cubic specimen (top). The direction *x* corresponds to the direction of the main load *in vivo*.

The second goal is to quantify experimentally the influence of the specimen density and anisotropy on the separation of fast and slow waves and on their characteristics.

II. MATERIALS AND METHODS

A. Bone specimen preparation

Cancellous bone specimens were taken from femoral heads removed during multiple organ harvesting, in compliance with French regulations. The donors were 16 men and 6 women whose ages ranged from 23 to 67 years. The femoral heads were cut using an oscillating saw along the main load direction. Slices with thickness of 9 mm were obtained including the mid-plane of the femoral head in a direction perpendicular to main load axis. Following these steps, 35 cubic specimens, about $9\times9\times9$ mm^3 in size, were prepared for the study (Fig. 1). All the specimens were then defatted by supercritical carbon dioxide followed by inactivation using sequential exposure to hydrogen peroxide, sodium hydroxide, monosodium dihydrogen phosphate, and ethanol (Vastel *et al.*, 2007).

The apparent density of each specimen was determined from the mass and the apparent total volume of each specimen. The size of the specimens was measured using a digital caliper (precision: 0.01 mm) to determine the apparent total volume. The mass of the specimens was measured using an analytical balance (precision: 0.001 g). The bone volume fraction, which in the standard histomorphometry terminology is BV/TV (in percent), was obtained by assuming the density of solid part as 1960 kg/m^3 (Hosokawa and Otani, 1997, 1998).

B. Ultrasonic measurements

A conventional ultrasonic pulse technique was used to measure the longitudinal wave speeds, as shown in Fig. 2. A wide band PVDF focused transmitter 20 mm in diameter with a focal length of 40 mm (custom made; Toray, Yokohama, Japan) and a home-made PVDF flat receiver 10 mm in diameter were used in this experiment. The beam width at the half maximum value of the wave amplitude was approximately 1.5 mm at the focal point in water at 1 MHz (Otani, 2005). Both PVDF transducers were mounted coaxially with a distance of 60 mm in degassed water at about 20 °C. A function generator (33220A; Agilent Technologies, Colorado Springs, CO) delivered electrical pulses to the transmitter, which was converted into ultrasonic waves. A single sinusoidal signal with a center frequency of 1 MHz and amplitude of 10 $\mathrm{V_{p\text{-}p}}$ was applied to the transmitter. The longitudinal wave propagated through water, sample, and water. The other transducer received the wave and converted it into an electrical signal. The signal was amplified by a 40-dB preamplifier (Data Precision D1000 dual preamplifier, Analogic, MA) and visualized with an oscilloscope (TDS2024; Tektronix, Tokyo, Japan). Here, in order to perform comparative discussion with Biot's model and bovine data (Hosokawa and Otani, 1997), we have selected the same frequency of 1 MHz.

Before measurements, the specimens were degassed for 20 min to remove air bubbles trapped in cancellous bone. For the measurements, the focal spot of the transmitter was placed in the center of the specimen. Measurements were performed along three mutually orthogonal directions. The direction of the ultrasound incident wave was always perpendicular to the side surface of the specimen for each propagation direction.

A wave that passed through water without specimen is shown in Fig. 3(a). In some specimens, we could observe the clear separation of fast and slow waves like Fig. 3(b). In this case, sound speed values of fast and slow waves were obtained from the difference in arriving time between a reference signal measured in water and the signal that passed through the water and the specimen. The arriving times of the signals were defined as follows: the arrival time of fast wave was determined using the first point of the signal reaching a threshold set at 10 % maximum amplitude of the signal; the arrival time of slow wave was determined using the first zero crossing of slow wave signal, as shown in Fig. 3. The difference in arrival times between the reference sig-

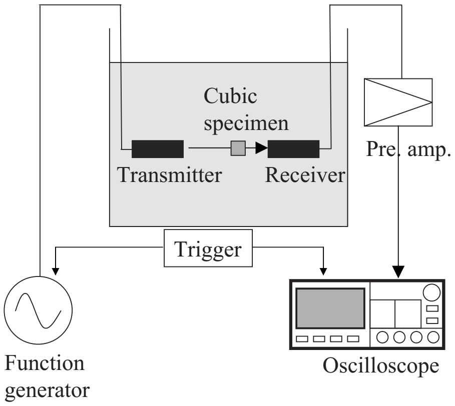


FIG. 2. Schematic representation of the experimental set-up.

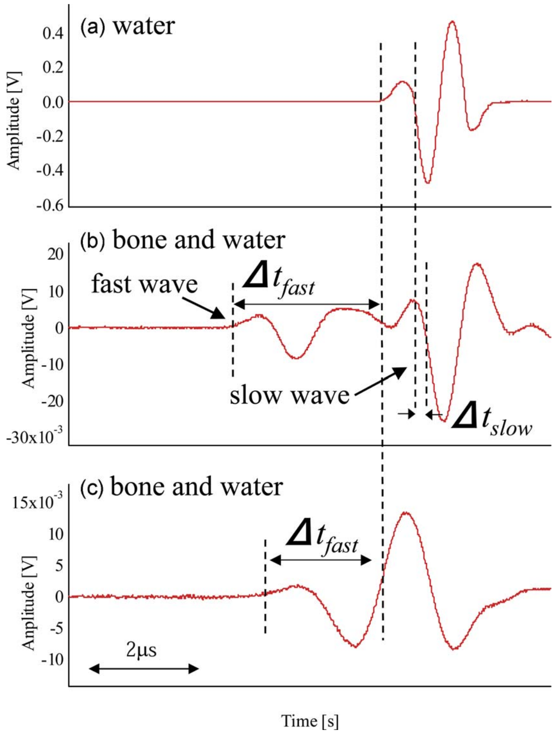


FIG. 3. (Color online) Examples of measured waveforms: (a) reference signal in water and (b) and (c) signals after propagation through bone specimens. In (b), a clear separation of the fast and slow waves is observed, whether in (c) the two waves overlap. The dotted vertical lines indicate the position of the beginning of the signals used to evaluate the difference in arrival time (Δt) between the reference signal and the signals after propagation through the specimens.

nal in water and the fast wave signal, Δt_{fast}, and that of slow wave, Δt_{slow}, is shown in Fig. 3. In some specimens, the fast and slow waves mostly overlapped like in Fig. 3(c). In such cases of relative mixing of fast and slow waves, we did not obtain the arrival time of slow waves. We then could obtain only one parameter, the arrival time of the fast wave. The fast and slow wave speeds in the specimen were derived from the difference in arrival time as

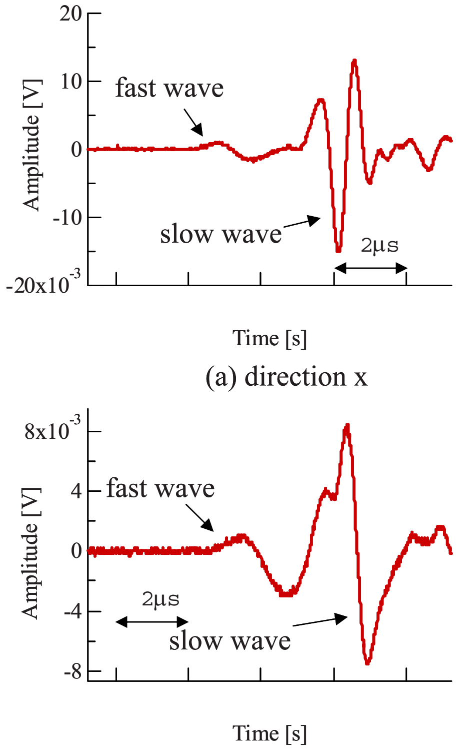
$$v = dv_0/(d - \Delta t v_0), \qquad (1)$$

where d is the specimen thickness, v_0 is the wave speed in water, and Δt is the difference in arrival time for either the fast or the slow wave. The wave speed in water was referred from the previous report by Greenspan and Tschiegg (1959). One should note here that the obtained speeds are the speeds of wave front, not phase or group velocities of the fast wave. We adopted these speeds of wavefront in order to avoid the effect of fast and slow wave overlapping.

In addition, this speed of the fast wavefront is expected to strongly reflect the trabecular structure because the "fastest part" of the wave has been shown to mainly propagate through the trabecular (solid) part (Haïat *et al.*, 2008; Nagatani *et al.*, 2006).

C. Short term precision

The short term precision of the ultrasonic measurement was assessed by measuring six times with intermediate repositioning one specimen in which both fast and slow waves could be clearly observed. The reproducibility was quantified in terms of coefficient of variation (CV) (in percent) of both fast and slow wave velocities. The CVs were found to be 1.12% and 0.07% for fast and slow wave velocities, respectively.



(a) direction x

(b) direction y

FIG. 4. (Color online) Signals measured in the specimen where fast and slow waves could be clearly identified in both directions x (a) and y (b).

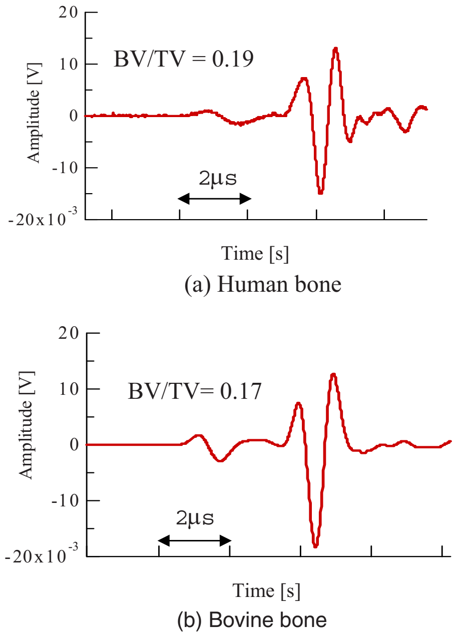
III. RESULTS AND DISCUSSION

A. Observation of fast and slow waves

We found a clear separation of fast and slow waves in 16 of the 35 specimens. For two of the specimens, fast and slow waves were observed in two different directions, resulting in a total of 18 observed waveforms with a clear separation of the two waves. Typical waveforms are shown in Figs. 4(a) and 5(a). For 14 of these specimens, fast and slow waves were observed in the direction x (defined in Fig. 1), which is the direction of the main load *in vivo*, corresponding to main direction of alignment of the trabeculae. For the two remaining specimens, the fast and slow waves were observed in direction z only. Finally, we could observe for one specimen the propagation of fast and slow waves in both directions x and y [Figs. 4(a) and 4(b)] and for another specimen in both directions x and z.

B. Apparent density measurements

The mean density of the specimens was 523 ± 115 kg/m^3. The apparent mean density for the groups of specimens in which fast and slow waves could be separated was 491 ± 106 kg/m^3. The apparent mean density for the groups of specimens in which fast and slow waves could



(a) Human bone

(b) Bovine bone

FIG. 5. (Color online) Signals measured in a human specimen (a) and in a bovine specimen (b) after propagation along the direction in the main load. Both specimens have similar bone volume fractions (BV/TV) of 0.19 and 0.17, respectively.

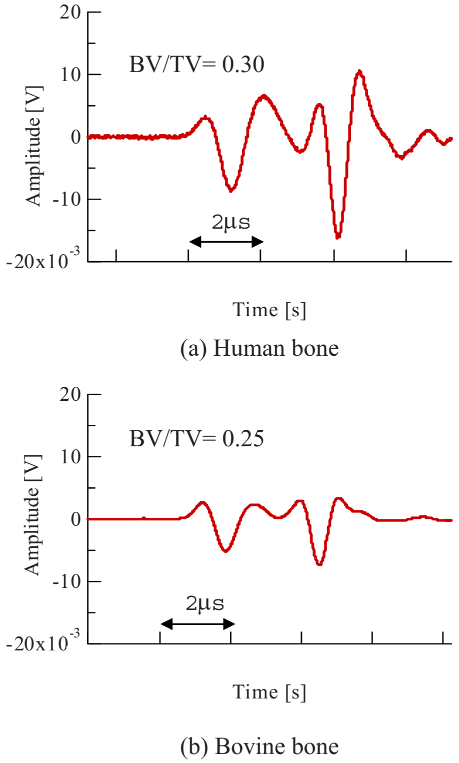
not be separated was found to be slightly higher: 544 ± 119 $\mathrm{kg/m^3}$. However, no statistical difference was found between the apparent mean density of the two groups (using a two-sample *t*-test with a significance level of 5%), demonstrating that, in our group of specimens, apparent density was not a determinant factor to explain the separation of fast and slow waves.

C. Comparison between waveforms observed in human and in bovine bones

The waveforms measured in our collection of human specimens were found to be similar to waveforms measured by others in bovine bones. Figures 5 and 6, show typical separated waveforms obtained on human and bovine trabecular bone specimens with low and high bone volume fractions (BV/TV), respectively. For the bovine bone data, we referred to the study of Hosokawa and Otani (1997, 1998). It was observed from Figs. 5 and 6 that waveforms that passed through the human cancellous bones were similar to those in bovine bones.

D. Variations in fast and slow wave amplitudes as a function of bone volume fraction

Both amplitudes of fast and slow waves changed with bone volume fraction. Here the amplitude was obtained from the maximum value of observed wave in the time domain. At low bone volume fraction, the amplitudes of fast waves were much lower than that of slow wave. At high bone volume fraction, the amplitudes of fast and slow waves are equivalent. This observation is quantified in Fig. 7, where we reported the evolution of the ratio of the fast and slow wave peak amplitudes as a function of bone volume fraction. For the highest bone volume fractions of our specimen collection, this ratio becomes higher than 1. This result is in agreement with former studies, indicating that the properties of the fast waves are strongly related to the solid part of cancellous bone, whereas the slow wave properties depend more on the fluid part (Hosokawa and Otani, 1997, 1998; Cardoso *et al.*, 2003; Haïat *et al.*, 2008; Nagatani *et al.*, 2006).



(a) Human bone

(b) Bovine bone

FIG. 6. (Color online) Signals measured in a human specimen (a) and in a bovine specimen (b) after propagation along the direction in the main load. Both specimens have similar bone volume fractions (BV/TV) of approximately 0.25 and 0.30, respectively.

E. Variations in fast and slow wave speeds as a function of bone volume

Figure 8 shows the fast and slow wave speeds, estimated using Eq. (1), as a function of bone volume fraction. In Fig. 8, together with bovine data are represented the fast wave

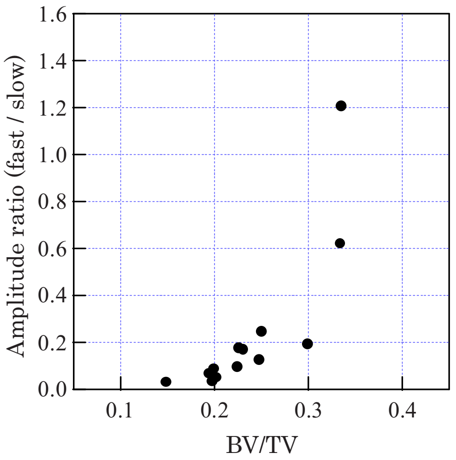


FIG. 7. (Color online) Ratio of the peak amplitude of the fast and slow waves as a function of bone volume fraction.

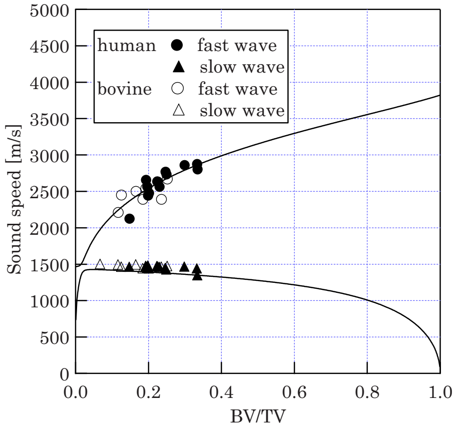


FIG. 8. (Color online) Experimental sound speed of fast (circles) and slow (triangles) waves for human specimens (present study) and bovine specimens (data taken from the study of Hosokawa *et al.*, 1997). The lines are the theoretical predictions from Biot's model for sound speed of the fast and slow waves.

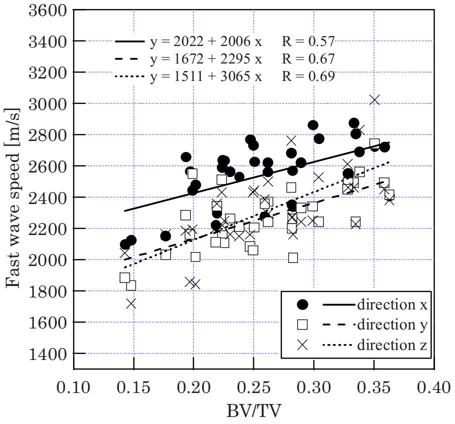


FIG. 9. (Color online) Fast wave speed as a function of bone volume fraction for the three orthogonal directions x (circles), y (squares), and z (crosses). The lines represent the three corresponding linear fits. The three linear regression equations and the corresponding correlation coefficients are given on top of the figure.

speeds only measured along the direction of main load axis for 12 specimens where a clear separation between fast and slow waves can be found. For comparison, we also reported values of fast and slow wave speeds measured in bovine bones previously measured by Hosokawa and Otani (1997, 1998). The fast wave speed increased with the bone volume fraction from 2100 to 2900 m/s, and the slow wave speed remained constant around 1450 m/s. As shown in Fig. 8, the fast and slow wave speed values in human bone were close to those measured in bovine bone. The theoretical predictions for both fast and slow wave speeds obtained using Biot's model are also reported in Fig. 8. The formulation of the model used is similar to the one used by Hosokawa and Otani (1997, 1998), and the computations were made using the set of parameters listed in Table I of Hosokawa and Otani (1997). The theoretical curves fit the distribution of experimental data. This is an interesting result, telling us that concerning the speeds of fast and slow waves, the data published on bovine bones may be useful in the study of acoustical phenomenon in human bones.

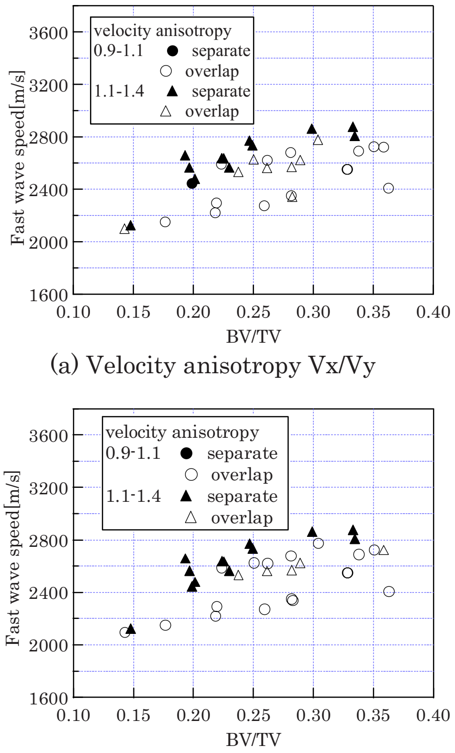
F. Acoustical anisotropy of fast wave speed

We have reported that the ultrasound speed of the fast wave depends not only on the mass of cancellous bone but also on the direction of propagation of the ultrasonic wave (Mizuno *et al.*, 2008). Our present results also support this finding. Figure 9 shows the relationship between the fast wave speed and bone volume fraction for the three orthogonal directions in all the 35 specimens. The fast wave speed values ranged from 1700 to 3000 m/s as a function of bone volume fraction. Using a paired t-test with 5% significance level, the values of fast wave speed were found to be higher in the direction x than in the two other directions: fast wave speed in direction x (2544 ± 206 m/s) was found to be higher than in direction y (2270 ± 203 m/s) with $p < 5 \times 10^{-9}$ and higher than in direction z (2310 ± 260 m/s) with $p < 5 \times 10^{-6}$. Values of fast wave velocities in directions y and z were not found to be statistically different. This result nicely indicates that the fast wave speed is sensitive to the micro-structural anisotropy and is not solely determined by bone volume fraction.

G. Influence of anisotropy on the separation of fast and slow waves

Because variations in apparent density were not able to explain the separation of fast and slow waves, we investigated the possible role of anisotropy as factor influencing separation. Velocity anisotropy of the specimens was investigated for the fast wave speeds obtained in the three orthogonal directions (Vx, Vy, and Vz). Because most of specimens showed the maximum values of the fast wave speed in direction x, we focused on the fast wave ratios, Vx/Vy and Vx/Vz, calling them fast wave velocity anisotropy indices. Figure 10 shows the fast wave velocities Vx as a function of bone volume fraction and for different velocity anisotropies (group No. 1: velocity anisotropy between 0.9 and 1.1; group No. 2: velocity anisotropy between 1.1 and 1.4). For each group, we indicated if the two waves were separated or if they were overlapped. For high value of velocity anisotropy (group No. 2), fast and slow waves could be separated for a majority of the specimens (separation in 60% of the specimens for high Vx/Vy; 72% for high Vx/Vz), while it was impossible (except for a single specimen) to separate them for velocity anisotropy below 1.1 both for Vx/Vy and Vx/Vz indices. Moreover, in the bone volume fraction range under study ($0.1 < \mathrm{BV/TV} < 0.4$), the wave separation in highly anisotropic specimens was found to be nearly independent of bone volume fraction.

The values of velocity anisotropy were found to be statistically different in the group of specimens in which fast and slow waves could be separated ($Vx/Vy = 1.20 \pm 0.09$ and $Vx/Vz = 1.23 \pm 0.08$) than in the group of specimens in which fast and slow waves could not be separated ($Vx/Vy = 1.08 \pm 0.07$ and $Vx/Vz = 1.04 \pm 0.07$). The statistical difference was assessed using bilateral two-sample t-tests with 5% significance level.



(a) Velocity anisotropy Vx/Vy

(b) Velocity anisotropy Vx/Vz

FIG. 10. (Color online) Fast wave velocities Vx (measured in the direction of main load axis) as a function of bone volume fraction and for different velocity anisotropies [velocity anisotropy Vx/Vy (a) and acoustical anisotropy Vx/Vy (b)]. Group No. 1 circles: velocity anisotropy between 0.9 and 1.1; group No. 2 triangles: velocity anisotropy between 1.1 and 1.4. For each group, we indicated if the two waves were separated or if they were overlapped.

It was reported in a previous study that the velocity anisotropy was strongly related to the structural anisotropy (Mizuno *et al.*, 2008). Therefore, our present data also demonstrate that the separation phenomena and the fast wave speeds are strongly affected by the anisotropy.

Most of the observations of separated fast and slow waves were performed when the ultrasound propagation was parallel to the main load axis (the direction x shown in Fig. 1) which is known to coincide with the main direction of the trabeculae. This result comforts the idea that the fast wave mainly propagates along the trabecular network (Haïat *et al.*, 2008; Nagatani *et al.*, 2006). Therefore, our results suggest that the two waves might be separated more easily when the wave has propagated into a highly anisotropic trabecular bone medium in a direction parallel to the main direction of trabecular alignment.

An additional important result of the present study is that we could also observe the wave separation in directions orthogonal to the main load direction in a few specimens. The femoral head micro-architecture presents indeed an anisotropic network of trabeculae (Chappard *et al.*, 2006), and depending on the site, a shift of the main alignment of the trabeculae with respect to the anatomic direction might explain our observations. This finding is in agreement with results obtained with numerical simulations of ultrasound wave propagation through trabecular micro-structure, where a separation of the two waves has been observed for different directions (Haïat *et al.*, 2008), depending on the Degree of Anisotropy (DA) and the bone volume fraction.

IV. CONCLUSION

We have experimentally investigated the wave propagation phenomena in cancellous bone of human femur. The fast and slow waves were clearly observed in human cancellous bone harvested from human femoral head. Fast and slow waves could be separated mainly when the waves propagated along the direction of the main load *in vivo*, which correspond to the main direction of alignment of the trabeculae. But in a few cases we could also observe the separated waves in directions which were orthogonal to the main load direction.

The waveforms and sound speeds of fast and slow waves in human bone were found to be similar to those in bovine bone at equivalent bone volume fraction. The fast wave speed value clearly increased with bone volume fraction (from 2100 to 2900 m/s) and the slow wave speed was almost constant value (around 1450 m/s).

The influence of anisotropy on the separation of fast and slow waves was clearly demonstrated: between the two groups of specimens, one with fast and slow waves separated and one with no separation, the apparent density was not statistically different but the velocity anisotropy ratio was. In the various ranges of bone volume fraction under study, the group of specimens in which we could find the separated waves exhibited higher velocity anisotropy.

The fast wave speed was found to be highly anisotropic and had its maximum value when the wave propagated along the direction of the main load *in vivo*.

These results open interesting perspective for *in vivo* trabecular bone assessment. For bone sites accessible *in vivo* where both fast and slow waves can be observed [e.g., distal radius (Mano *et al.*, 2006)], an assessment of fast and slow wave separations together with a measurement of their speed anisotropy could be a way to gain insight into bone trabecular micro-architecture anisotropy. Because anisotropy is an important factor to determine bone strength in addition with bone mass, which can be evaluated using conventional through transmission QUS measurement, an ultrasound assessment of micro-structure anisotropy may lead to an improved assessment of bone strength *in vivo*.

ACKNOWLEDGMENTS

This study was supported by the Academic Frontier Research Project on “New Frontier of Biomedical Engineering Research” of Doshisha University, the JSPS, and CNRS bilateral joint project and grant-in-aid for scientific research (B) supported by JSPS. Authors would like to express their gratitude on the valuable discussion with Professor Atsushi Hosokawa at Akashi National College of Technology.

Biot, M. A. (**1956a**). “Theory of propagation of elastic waves in a fluid-satured porous solid. II. Higher frequency range,” J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1956b**). “Theory of propagation of elastic waves in a fluid-satured porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 179–191.

Bouxsein, M. L., Courtney, A. C., and Haynes, W. C. (**1995**). "Ultrasound and densitometry of the calcaneus correlates with the failure load of cadaveric femurs," Calcif. Tissue Int. **56**, 99–103.

Cardoso, L., Teboul, F., Sedel, L., Oddou, C., and Meunier, A. (**2003**). "In vitro acoustic waves propagation in human and bovine cancellous bone," J. Bone Miner. Res. **18**, 1803–1812.

Chappard, C., Basillais, A., Benhamou, L., Bonassie, A., Brunet-Imbault, B., Bonnet, N., and Peyrin, F. (**2006**). "Comparison of synchrotron radiation and conventional x-ray microcomputed tomography for assessing trabecular bone microarchitecture of human femoral heads," Med. Phys. **33**, 3568–3577.

Cheng, X. G., Nicholson, P. H. F., Boonen, S., Lowet, G., Brys, B., Aerssens, J., Perre, G. V. D., and Dequeker, Y. (**1997**). "Prediction of vertebral strength in vitro by spinal bone densitometry and calcaneal ultrasound," J. Bone Miner. Res. **12**, 1721–1728.

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

Glüer, C., Wu, C., Jergas, M., Goldstein, S., and Genant, H. (**1994**). "Three quantitative ultrasound parameters reflect bone structure," Calcif. Tissue Int. **55**, 46–52.

Greenspan, M., and Tschiegg, C. E. (**1959**). "Tables of the speed of sound in water," J. Acoust. Soc. Am. **31**, 75.

Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2007**). "Variation of ultrasonic parameters with microstructure and material properties of trabecular bone: A 3D model simulation," J. Bone Miner. Res. **22**, 665–674.

Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2008**). "Fast wave ultrasonic propagation in trabecular bone: Numerical study of the influence of porosity and structural anisotropy," J. Acoust. Soc. Am. **123**, 1694–1705.

Han, S., Medige, J., Davis, J., Fishkin, Z., Mihalko, W., and Ziv, I. (**1997**). "Ultrasound velocity and broadband attenuation as predictors of load-bearing capacities of human calcanei," Calcif. Tissue Int. **60**, 21–25.

Hoffmeister, B. K., Whitten, S. A., and Rho, J. Y. (**2000**). "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone (N.Y.) **26**, 635–642.

Homminga, J., McCreadie, B. R., Weinans, H., and Huiskes, R. (**2003**). "The dependence of the elastic properties of osteoporotic cancellous bone on volume fraction and fabric," J. Biomech. **36**, 1461–1467.

Hosokawa, A., and Otani, T. (**1997**). "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562.

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

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

Hughes, E. R., Leighton, T. G., White, P. R., and Petley, G. W. (**2007**). "Investigation of an anisotropic tortuosity in a Biot model of ultrasonic propagation in cancellous bone," J. Acoust. Soc. Am. **121**, 568–574.

Kaczmarek, M., Kubik, J., and Pakula, M. (**2002**). "Short ultrasonic waves in cancellous bone," Ultrasonics **40**, 95–100.

Krieg, M. A., Barkmann, R., Gonnelli, S., Stewart, A., Bauer, D. C., Del Rio Barquero, L., and Kaufman, J. J. (**2008**). "Quantitative ultrasound in the management of osteoporosis: The 2007 ISCD Official Positions," J. Clin. Densitom. **11**, 163–187.

Laugier, P. (**2008**). "Instrumentation for in vivo assessment of bone strength," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1179–1196.

Lochmüller, E. M., Eckstein, F., Zeller, J. B., Steldinger, R., and Putz, R. (**1999**). "Comparison of quantitative ultrasound in the human calcaneus with mechanical failure loads of the hip and spine," Ultrasound Obstet. Gynecol. **14**, 125–133.

Mano, I., Horii, K., Takai, S., Suzaki, T., Nagaoka, H., and Otani, T. (**2006**). "Development of novel ultrasonic bone densitometry using acoustic parameters of cancellous bone for fast and slow waves," Jpn. J. Appl. Phys., Part 1 **45**, 4700–4702.

Mizuno, K., Matsukawa, M., Otani, T., Takada, M., Mano, I., and Tsujimoto, T. (**2008**). "Effects of structural anisotropy of cancellous bone on speed of ultrasonic fast waves in the bovine femur," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1480–1487.

Nagatani, Y., Imaizumi, H., Fukuda, T., Matsukawa, M., Watanabe, Y., and Otani, T. (**2006**). "Applicability of finite-difference time-domain method to simulation of wave propagation in cancellous bone," Jpn. J. Appl. Phys., Part 1 **45**, 7186–7190.

Nicholson, P. H. F., Haddaway, M. J., and Davie, M. W. J. (**1994**). "The dependence of ultrasonic properties on orientation in human vertebral bone," Phys. Med. Biol. **39**, 1013–1024.

Njeh, C. F., Kuo, C. W., Langton, C. M., Atrah, H. I., and Boivin, C. M. (**1997**). "Prediction of human femoral bone strength using ultrasound velocity and BMD: An in vitro study," Osteoporosis Int. **7**, 471–477.

NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy (**2001**). "Osteoporosis prevention, diagnosis, and therapy," J. Am. Med. Assoc. **285**, 785–795.

Otani, T. (**2005**). "Quantitative estimation of bone density and bone quality using acoustic parameters of cancellous bone for fast and slow waves," Jpn. J. Appl. Phys., Part 1 **44**, 4578–4582.

Sabaa, N., and Fellah, Z. A. (**2006**). "Ultrasonic Characterization of human cancellous bone using Biot's theory: Inverse problem," J. Acoust. Soc. Am. **120**, 1816–1824.

Sebaa, N., Fellah, Z. A., Fellah, M., Ogam, E., Mitri, F. G., Depollier, C., and Lauriks, W. (**2008**). "Application of the Biot model to ultrasound in bone: Inverse problem," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1516–23.

Vastel, L., Masse, C., Crozier, E., Padilla, F., Laugier, P., Mitton, D., Bardonnet, R., and Courpied, J. P. (**2007**). "Effects of gamma irradiation on mechanical properties of defatted trabecular bone allografts assessed by speed-of-sound measurement," Cell Tissue Bank. **8**, 205–210.

Erratum: Coupled vibration analysis of the thin-walled cylindrical piezoelectric ceramic transducers [J. Acoust. Soc. Am. 125, 803–818 (2009)]

Boris Aronov
BTech Acoustic LLC, Acoustics Research Laboratory, Advanced Technology and Manufacturing Center, and Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723

(Received 6 March 2009; accepted 10 March 2009)
[DOI: 10.1121/1.3110204]

PACS number(s): 43.38.Ar, 43.38.Fx, 43.38.Pf, 43.10.Vx

The images in Figures 7 and 8 were mistakenly transposed in the course of the layout design. The corrected figure images and corresponding captions should appear as follows:

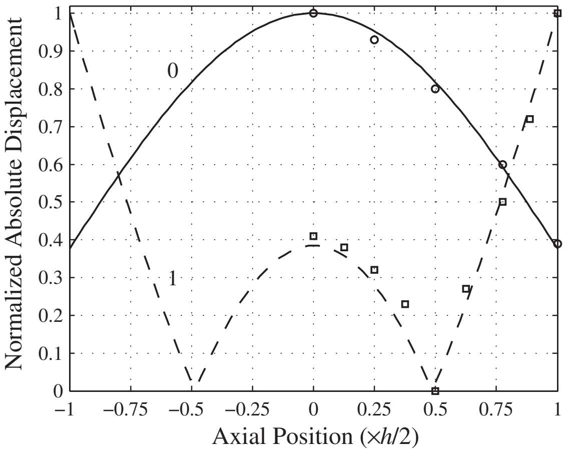


FIG. 7. The resonance mode shapes of a tube ($2a$=35 mm, t=3.2 mm) at $h/2a$=1.1: f_0=28.8 kHz at branch 0 and f_1=30.8 kHz at branch 1. Calculated mode shapes are shown by lines, whereas experimental data from Ref.10 are shown by circles and squares.

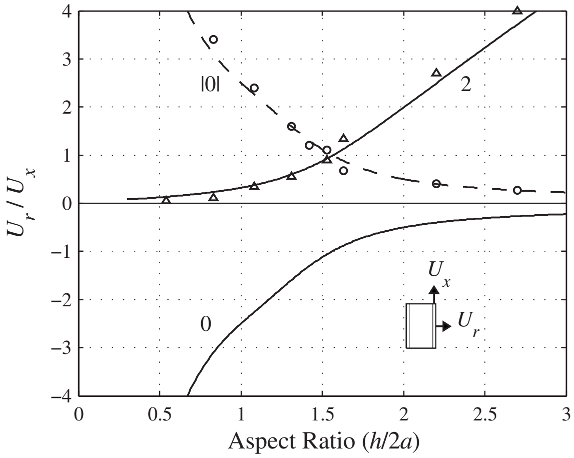


FIG. 8. The ratio of the magnitudes of vibration in the radial and axial directions $[U_r(x=0)/U_x(\pm h/2)]$ vs $h/2a$ along branches 0 and 2. The modulus of $[U_r/U_x]_0$ is shown by a dashed line.

0001-4966/2009/125(5)/3467/1/$25.00
© 2009 Acoustical Society of America

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.*

President's report on the 156th meeting of the Acoustical Society of America held in Miami, Florida

The 156th meeting of the Acoustical Society of America was held 10–14 November 2008 at the Doral Golf Resort and Spa in Miami, Florida. This is the sixth time that the Society has met in this city, the previous meetings being held in 1967, 1972, 1977, 1981, and 1987.

The meeting drew a total of 851 registrants, including 111 nonmembers, 218 students and 81 registrants from outside North America. There were 23 registrants from Japan; 11 from the U.K.; 8 from Korea; 6 from France; 5 each from Germany and the Netherlands; 4 from Finland; 3 each from Denmark and Sweden; 2 each from Australia and Austria; and 1 each from Brazil, Ecuador, Hong King, Iran, Israel, New Zealand, Portugal, Spain, and Turkey. North American countries, Canada, Mexico, and the United States, accounted for 31, 2, and 737 respectively.

A total of 660 papers, organized into 81 sessions, covered the areas of interest of all 13 Technical Committees. The meeting also included 19 meetings dealing with standards. The evening tutorial lecture series was continued by Joe Posey of NASA of the tutorial titled "Aircraft Noise Prediction."

The Society's thirteen Technical Committees held open meetings during the Miami meeting where they made plans for special sessions at upcoming ASA meetings, discussed topics of interest to the attendees and held informal socials after the end of the official business. These are working, collegial meetings and all people attending Society meetings are encouraged to attend and to participate in the discussions. More information about Technical Committees, including minutes of meetings, can be found on the ASA Website ⟨http://asa.aip.org/committees.html⟩ and in the Acoustical News USA section of JASA in the October and November issues.

A short course titled "Ultrasonic Nondestructive Evaluation and Materials Characterizations" was given to a group of about 15 students. The instructor was Dale E. Chimenti, Professor of Aerospace Engineering and Senior Scientist in the Center for Nondestructive Evaluation at Iowa State University, Ames, Iowa.

FIG. 1. (Color online) ASA President Mark Hamilton (l) presents the 2007 Science Writing Award in Acoustics for Journalists to Hazel Muir (r).

An exhibit was held in conjunction with the meeting and included displays with materials and services for the acoustics and vibration community. It included exhibits of computer-based instrumentation, sound level meters, devices for noise control and sound prediction among others. The exhibit began with an opening reception on Monday evening and was open on Tuesday and Wednesday.

The ASA Student Council hosted a Student Reception with over 100 people in attendance. This reception, which was supported by the National Council of Acoustical Consultants, enabled students to meet with established members of the Acoustical Society of America. Several of the Technical Committees awarded Best Student Paper Awards or Young Presenter Awards to students and young professionals who presented papers at the meeting. The list of award recipients, as well as other information for students, can be found online at the ASA Student Zone website ⟨http://www.acosoc.org/student/⟩

Social events included the two social hours held on Tuesday and Thursday, an "icebreaker" and a reception for students, the Fellows Luncheon and the morning coffee breaks. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics. These social events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions. The Women in Acoustics Luncheon was held on Wednesday afternoon with attendance of about 100.

The plenary session included a business meeting of the Society, announcements, acknowledgment of the members and other volunteers who organized the meeting and the presentation of awards and certificates to newly-elected Fellows.

ASA President Mark Hamilton presided over the Plenary Session and Awards Ceremony. Lisa Zurk, Chair of the Spring 2009, addressed the audience and invited and encouraged them to attend that meeting to be held in Portland, Oregon in May. Two ASA Science Writing Awards were presented. The 2007 Science Writing Award in Acoustics for Journalists was presented to Hazel Muir for her article "Noisy Neighbors," published in *New Scientist Magazine* in August 2007 (see Fig. 1). The 2007 Science Writing Award for Professionals in Acoustics was presented to Kathleen Vigness Raposa, Gail

FIG. 2. (Color online) ASA President Mark Hamilton (l) presents the 2007 Science Writing Award for Professionals in Acoustics to Kathleen Vigness-Raposa, Gail Scowcroft, and Peter F. Worcester.

FIG. 3. (Color online) Karim Sabra, recipient of the 2009 A. B. Wood Medal and Prize of the Institute of Acoustics (UK).

FIG. 5. (Color online) ASA President Mark Hamilton (l) presents the Silver Medal in Musical Acoustics to Gabriel Weinreich (r).

FIG. 6. (Color online) ASA President Mark Hamilton (l) presents the Wallace Clement Sabine Medal to John S. Bradley (r).

Scowcroft, Christopher Knowlton, and Peter Worcester for the website "Discovery of Sound in the Sea." (see Fig. 2).

The ASA President introduced Karim Sabra, recipient of the 2009 A. B. Wood Medal and Prize of the Institute of Acoustics (U.K.) (see Fig. 3). Dr. Sabra will receive the award at a future meeting of the Institute of Acoustics.

The 2008 Rossing Prize in Acoustics Education was presented to D. Murray Campbell, University of Edinburgh. The award includes a silver medal and a prize of $3,000. Dr. Campbell presented the Acoustics Education Prize Lecture titled "From the sublime to the scientific: What musicians and acousticians can learn from each other" earlier in the meeting (See Fig. 4).

The Silver Medal in Musical Acoustics was presented to Gabriel Weinreich of the University of Michigan "for contributions to violin and piano acoustics" (see Fig. 5). The Wallace Clement Sabine Medal was presented to John S. Bradley "for advancing measurment techniques in spaces for speech and music (see Fig. 6). The Silver Medal in Physical Acoustics was presented to Peter J. Westervelt, Professor Emeritus at Brown University "for fundamental contributions to nonlinear acoustics" (see Fig. 7).

Election of sixteen members to Fellow grade was announced and fellowship certificates were presented. New fellows are: David A. Berry, George A. Bissinger, John A. Fawcett, Dennis M. Freeman, Bruce R. Gerratt, Frank H. Guenther, Keith R. Kluender, Yiu W. Lam, Marshall Long, Christian Lorenzi, Bryan E. Pfingst, Joe W. Posey, Stuart Rosen, Armen Sarvazyan, Michael A. Stone, Ann K. Syrdal, Joe Wolfe (see Fig. 8).

ASA President Mark Hamilton expressed the Society's thanks to members of the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning, including Harry A. DeFerrari, General Chair, Eric I. Thorsos and David R. Palmer, Technical Program Cochairs, and Shari Vaughan, Meeting Assistant. He also expressed thanks to the members of the Technical Program Organizing Committee: Eric I. Thorsos and David R. Palmer, Technical Program Cochairs; Jennifer Wylie, Acoustical Oceanography; Dorian S. Houser, Animal Bioacoustics; Gary W. Siebein, Architectural Acoustics; Saurabh Datta, Biomedical Ultrasound/

FIG. 4. (Color online) ASA President Mark Hamilton (l) presents the Rossing Prize in Acoustics Education to D. Murray Campbell (r).

FIG. 7. (Color online) ASA President Mark Hamilton (l) with Peter J. Westervelt, recipient of the Silver Medal in Physical Acoustics.

FIG. 8. (Color online) New Fellows of the Acoustical Society of America with ASA President and Vice President (l-r): ASA President Mark Hamilton, Ann K. Syrdal, Armen Sarvazyan, Joe W. Posey, Marshall Long, Yiu W. Lam, Keith R. Kluender, Bruce R. Gerratt, John A. Fawcett, George A. Bissinger, David A. Berry, ASA Vice President Victor W. Sparrow.

Bioresponse to Vibration; James P. Chambers, Education in Acoustics; Thomas R. Howarth, Engineering Acoustics; Edward W. Large, Musical Acoustics; Richard J. Peppin, Erica Ryherd, Noise; James P. Chambers, Physical Acoustics; Gail Donaldson, Psychological and Physiological Acoustics; David M. Chapman, Signal Processing in Acoustics; Stefan A. Frisch, Catherine L. Rogers, Speech Communication; Joseph M. Cuschieri, Structural Acoustics and Vibration; Altan Turgut, Underwater Acoustics.

The full technical program and award encomiums can be found in the printed meeting program or online for readers who wish to obtain further information about the Miami meeting (visit scitation.aip.org/jasa and select Volume 124, Issue 4, from the list of available volumes).

We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information about future meetings can be found in the *Journal* and on the ASA Home Page at ⟨http://asa.aip.org⟩.

MARK F. HAMILTON
President 2008–2009

USA Meetings Calendar

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

18–22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
24–28 June	5th International Middle-Ear Mechanics in Research and Otology (MEMRO), Stanford University, Stanford, CA [http://memro2009.stanford.edu].
26–30 October	158th Meeting of the Acoustical Society of America, San Antonio, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
	2010
19–23 April	158th Meeting of the Acoustical Society of America, Baltimore, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
15–19 November	2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

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]

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: Schomer@SchomerAndAssociates.com]

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

18–22 May 2009—Portland, Oregon

Meetings of the National Standards Committees S1-Acoustics, S2-Mechanical Vibration and Shock, S3-Bioacoustics, S3/SC 1-Animal Bioacoustics, and S12-Noise, and the ten U.S. TAGs administered by ASA will be held in conjunction with the 157th meeting of the Acoustical Society of America in Portland, Oregon.

STANDARDS MEETINGS CALENDAR—INTERNATIONAL

15–19 June 2009—Charlottenlund, Denmark

- ISO/TC108/SC 5, Condition monitoring and diagnostics of machines.

1–4 September 2009—Las Vegas, Nevada, USA

- ISO/TC108/SC 4, Human exposure to mechanical vibration and shock.

9–13 November 2009—Tokyo, Japan

- IEC/TC29, Electroacoustics.

16–20 November 2009—Seoul, Republic of Korea

- ISO/TC 43, Acoustics
- ISO/TC43/SC 1, Noise
- ISO/TC43/SC 2, Building acoustics.

STANDARDS NEWS FROM THE UNITED STATES

(Partially derived from *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 existing standards. The dates listed in parentheses are for information only.

ASA (ASC S1) (Acoustical Society of America)

Revisions

BSR/ASA S1.17/Part 1-200x, Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air (revision and redesignation of ANSI S1.17/Part 1-2000)

Describes test procedures for determining the insertion loss of windscreens mounted on microphones. Insertion loss is determined over a specified frequency measurement range and for still-air conditions in the test facility. (April 13, 2009)

ASA (ASC S2) (Acoustical Society of America)

Revisions

BSR/ASA S2.28-200x, Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery (revision and redesignation of ANSI S2.28-2003)

Provides guidance for assessing the severity of vibrations measured on bearing housings of shipboard machinery so as to ensure reliable mechanical operation. The criteria apply to the vibration of all non-reciprocating machinery on board surface ships, except for main propulsion machinery. They apply to broadband vibration measurements taken on the bearing housings, of machines under steady-state operating conditions with normal operating conditions of speed and load. (March 2, 2009)

New Standards

BSR/ASA S2.62-200x, Shock Test Requirements for Equipment in a Rugged Shock Environment (new standard)

Applies to testing equipment that will be subjected to shock. Defines test requirements and severity thresholds for a large range of shock environments, including but not limited to shipping, transport, and rugged operational environments. This standard will allow vendors to better market, and users to more easily identify equipment that will operate or simply survive in rugged shock environments. (March 23, 2009)

ASA (ASC S3) (Acoustical Society of America)

Revisions

BSR/ASA S3.2-200x, Method for Measuring the Intelligibility of Speech over Communication Systems (revision and redesignation of ANSI S3.2-1989 (R1999))

Includes measurement of speech intelligibility over entire communication systems, evaluation of the contributions of elements of speech communication systems, and evaluation of factors that affect the intelligibility of speech. Speech intelligibility over a communication system is measured by comparing the monosyllabic words trained listeners receive and identify with the words trained talkers speak into a communication system that connects the talkers with the listeners. (March 2, 2009)

0001-4966/2009/125(5)/3471/4/$25.00
© 2009 Acoustical Society of America

BSR/ASA S3.22-200x, Specification of Hearing Aid Characteristics (revision and redesignation of ANSI/ASA S3.22-200x)

Describes air-conduction hearing-aid measurement methods that are particularly suitable for specification and tolerance purposes. Test methods described are output sound pressure level (SPL) with 90-dB input SPL, full-on gain, frequency response, harmonic distortion, equivalent input noise, current drain and induction-coil sensitivity. Specific configurations are given for measuring input SPL to hearing aid. (March 23, 2009)

Reaffirmations

BSR S3.21-2004 (R200x), Methods for Manual Pure-Tone Threshold Audiometry (reaffirmation and redesignation of ANSI S3.21-2004)

Provides a procedure for pure-tone audiometry that will serve the needs of persons conducting threshold measurements in industry, schools, medical settings, and other areas where valid audiometric threshold measurements are needed. (April 13, 2009)

AGMA (American Gear Manufacturers Association)

Reaffirmations

BSR/AGMA 6000-B96 (R200x), Specification for Measurement of Linear Vibration on Gear Units [reaffirmation of ANSI/AGMA 6000-B96 (R2002)]

Presents a method for measuring linear vibration on a gear unit. Recommends instrumentation, measuring methods, test procedures, and discrete frequency vibration limits for acceptance testing. Annexes list system effects on gear unit vibration and system responsibility. Introduces the determination of mechanical vibrations of gear units during acceptance testing. (March 24, 2009)

BSR/AGMA 6025-A98 (R200x), Sound for Enclosed Helical, Herringbone and Spiral Bevel Gear Drives [reaffirmation of ANSI/AGMA 6025-A98 (R2004)]

Describes a recommended method of acceptance testing and reporting of the sound pressure levels generated by a gear speed reducer or increaser when tested at the manufacturer″s facility. Annexes to the standard present sound power measurement methods for use when required by specific contract provisions between the manufacturer and purchaser. (March 24, 2009)

CEA (Consumer Electronics Association)

New Standards

BSR/CEA 2006-B-200x, Testing and Measurement Methods for Mobile Audio Amplifiers (new standard)

Defines characteristics that, considered collectively, describe the performance of power amplifiers designed for use in mobile applications. Power amplifiers designed for use in mobile applications include, but are not limited to:—separate single and multi-channel amplifiers;—integrated amplifiers; and—bandwidth-limited amplifiers that are connected to and rely solely on the vehicle's primary electrical system for power input and have output power ratings of greater than 5 watts when measured in accordance with CEA 2006-B. (April 13, 2009)

IEEE (Institute of Electrical and Electronics Engineers)

New Standards

BSR/IEEE 1652-200x, Standard for the Application of Free Field Acoustic Reference to Telephony Measurements (new standard)

Provides the techniques and rationale for referencing acoustic telephony measurements to the free field. This standard applies to ear-related measurements such as receive, sidetone and overall. (March 10, 2009)

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers (ASD) of the initiation and scope of activities expected to result in new or revised American National Standards (ANS). Early notification of activity intended to reaffirm or withdraw an ANS is optional. The mechanism by which such notification is given is referred to as the PINS process. For additional information, see clause 2.4 of the ANSI Essential Requirements: Due Process Requirements for American National Standards. Following is a list of proposed actions and new ANS that have been received recently from ASDs.

ASA (ASC S1) (Acoustical Society of America)

BSR/ASA S1.15, Part 3-200x, Measurement Microphones—Part 3: Microphone Calibration by Comparison Method (new standard)

Applies to measurement microphones, including laboratory standard microphones conforming to Part 1. This standard describes methods of determining the pressure or free-field sensitivity by comparison with a laboratory standard microphone or working standard microphone that has been calibrated in accordance with Part 2 or other parts. It describes processing of results to reduce the effect of imperfections of the acoustical environment and details factors that influence the determined sensitivity. Project Need: To create an additional part to ANSI S1.15 series, which is needed to update sections 6 and 7 in ANSI S1.10-1966. Stakeholders: Public and private laboratories for metrology, calibration, and conformity assessment, manufacturers.

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.10/Part1/ISO 7779:1999 (MOD), Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment (new standard)

Specifies the procedures for measuring and reporting the noise emission of information technology and telecommunications equipment. This Standard is considered part of a noise test code for this type of equipment, and is based on basic noise emission standards ISO 3741, ISO 3744, ISO 3745, and ISO 11201. Project Need: To revise a nationally adopted International Standard that was based on a previous ANS. It is anticipated that the changes adopted here will subsequently be incorporated into a future edition of ISO 7779. Stakeholders: Information technology, telecommunications.

ABMA (ASC B3) (American Bearing Manufacturers Association)

BSR/ABMA/ISO 15242-1-200x, Rolling bearings—Measuring methods for vibration—Part 1: Fundamentals (identical national adoption of ISO 15242-1)

Defines and specifies measuring methods for vibration of rotating rolling bearings under established test conditions together with calibration of related measuring systems. Project Need: To create a U.S. standard for methods of measuring for vibrations of rolling bearings. Stakeholders: Bearing manufacturers and users.

BSR/ABMA/ISO 15242-2-200x, Rolling bearings—Measuring methods for vibration—Part 2: Radial ball bearings with cylindrical bore and outside surface (identical national adoption of ISO 15242-2)

Specifies vibration-measuring methods for radial single-row and double-row ball bearings, with a contact angle up to and including 45 degrees, under established test conditions. This standard covers radial ball bearings with cylindrical bore and outside surface, except bearings with filling slots and three- and four-point contact bearings. Project Need: To create a U.S. standard for methods of measuring for vibrations of rolling bearings. Stakeholders: Bearing manufacturers and users.

BSR/ABMA/ISO 15242-3-200x, Rolling bearings—Measuring methods for vibration—Part 3: Radial spherical and tapered roller bearings with cylindrical bore and outside surface (identical national adoption of ISO 15242-3)

Specifies vibration measuring methods for double-row radial spherical roller bearings and single-row and double row radial tapered roller bearings, with a contact angle up to and including 45 degrees, under established test conditions. This standard covers double-row radial spherical roller bearings as well as single-row and double-row radial tapered roller bearings with cylindrical bore and outside surface. Project Need: To create a U.S. standard for methods of measuring for vibrations of rolling bearings. Stakeholders: Bearing manufacturers and users.

BSR/ABMA/ISO 15242-4-200x, Rolling bearings—Measuring methods for vibration—Part 4: Radial cylindrical roller bearings with cylindrical bore and outside surface (identical national adoption of ISO 15242-4)

Specifies vibration measuring methods for single-row and double-row radial cylindrical roller bearings, under established test conditions. It covers single-row and double-row radial cylindrical roller bearings with cylindrical bore and outside surface. Project Need: To create an American National Standard for this type of rolling bearing. Stakeholders: Bearing manufacturers and users.

ASME (American Society of Mechanical Engineers)

BSR/ASME B133.8M-200x, Installation Sound Emission, Gas Turbine (revision of ANSI/ASME B133.8M-1977 (R2001))

Provides methods and procedures for specifying the sound emissions of gas turbine installations for industrial, pipeline, and utility applications. Included are practices for making field sound measurements and for reporting field data. Project Need: The Standard remains applicable but is becoming out-of-date; a revision is necessary. Stakeholders: Users of gas turbines and those involved with sound measurements.

ISEA (International Safety Equipment Association)

BSR/ISEA 105-200x, Hand Protection Selection Criteria (revision of ANSI/ISEA 105-2005) Addresses the classification and testing of hand protection for specific performance properties related to mechanical, chemical, heat and flame, and vibration protection. Hand protection includes gloves, mittens, partial gloves, or other items covering the hand or a portion of the hand, which is intended to provide protection against or resistance to a specific hazard. Project Need: To provide an updated standard to reflect current technologies, test methods and other considerations related to the manufacture, selection and use of industrial hand protection. Stakeholders: Hand protection manufacturers, distributors, and users, including construction, manufacturing, and agriculture.

Final actions on American National Standards

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

ASA (ASC S3) (Acoustical Society of America)

Revisions

ANSI/ASA S3.45-2009, Procedures for Testing Basic Vestibular Function (revision and redesignation of ANSI S3.45-1999).

EIA (Electronic Industries Alliance)

New Standards

ANSI/EIA 364-27B-1996 (R2009), Mechanical Shock (Specified Pulse) Test Procedure for Electrical Connectors (new standard).

IEEE (Institute of Electrical and Electronics Engineers)

Supplements

ANSI/IEEE 1431-2004/Cor1-2008, Standard Specification Format Guide and Test Procedure for Coriolis Vibratory Gyros—Corrigendum 1: Figure 1—Gyro Axes and Misalignment Angles (supplement to ANSI/IEEE 1431-2004).

STANDARDS NEWS FROM ABROAD

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

Newly Published ISO and IEC Standards

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

ISO Standards

FIBRE OPTICS (TC 86)

IEC 61280-2-9 Ed. 2.0 b:2009, Fibre optic communication subsystem test procedures—Part 2–9: Digital systems—Optical signal-to-noise ratio measurement for dense wavelength-division multiplexed systems.

INDUSTRIAL FANS (TC 117)

ISO 14695/Cor1:2009, Industrial fans—Method of measurement of fan vibration—Corrigendum.

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 18436-3/Amd1:2009, Condition monitoring and diagnostics of machines—Requirements for qualification and assessment of personnel—Part 3: Requirements for training bodies and the training process—Amendment 1.

ISO 7919-3:2009, Mechanical vibration—Evaluation of machine vibration by measurements on rotating shafts—Part 3: Coupled industrial machines.

ISO 10816-3:2009, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 3: Industrial machines with nominal power above 15 kW and nominal speeds between 120 r/min and 15 000 r/min when measured in situ.

ISO 10816-7:2009, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 7: Rotodynamic pumps for industrial applications, including measurements on rotating shafts.

ISO Technical Reports

MACHINE TOOLS (TC 39)

ISO/TR230-8:2009, Test code for machine tools—Part 8: Determination of vibration levels.

IEC Standards

ELECTRICAL ACCESSORIES (TC 23)

IEC 62080 Ed. 1.1 b:2009, Sound signalling devices for household and similar purposes.

ELECTROACOUSTICS (TC 29)

IEC 61094-2 Ed. 2.0 b:2009, Electroacoustics—Measurement microphones—Part 2: Primary method for pressure calibration of laboratory standard microphones by the reciprocity technique.

IEC Technical Specifications

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

IEC/TS 60899 Ed. 1.0 b:1987, Sampling rate and source encoding for professional digital audio recording.

ISO and IEC Draft International Standards

This section lists proposed standards that the International Organization for Standardization (ISO) is considering for approval. The proposals have received substantial support within the technical committees or subcommittees that developed them and are now being circulated to ISO members for comment and vote. Standards Action readers interested in reviewing and commenting on these documents should order copies from ANSI.

ISO

EQUIPMENT FOR FIRE PROTECTION AND FIRE FIGHTING (TC 21)

ISO/DIS 7240-24, Fire detection and fire alarm systems—Part 24: Sound-system loudspeakers (May 1, 2009).

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 10816-1/DAmd1, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 1: General guidelines—Amendment 1 (March 23, 2009).

ISO/DIS 7919-4, Mechanical vibration—Evaluation of machine vibration by measurements on rotating shafts—Part 4: Gas turbine sets with fluid-film bearings (April 20, 2009).

ISO/DIS 10816-4, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 4: Gas turbine sets with fluid-film bearings (April 20, 2009).

ISO/DIS 10816-2, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 2: Land-based steam turbines and generators in excess of 50 MW with normal operating speeds of 1500 r/min, 1800 r/min, 3000 r/min and 3600 r/min (April 20, 2009).

ISO/DIS 7919-2, Mechanical vibration—Evaluation of machine vibration by measurements on rotating shafts—Part 2: Land-based steam turbines and generators in excess of 50 MW with normal operating speeds of 1 500 r/min, 1 800 r/min, 3 000 r/min and 3 600 r/min (May 1, 2009).

STEEL (TC 17)

ISO/DIS 10893-8, Non-destructive testing of steel tubes—Part 8: Automated ultrasonic testing of seamless and welded steel tubes for the detection of laminar imperfections (May 20, 2009).

ISO/DIS 10893-9, Non-destructive testing of steel tubes—Part 9: Automated ultrasonic testing for the detection of laminar imperfections in strip/plate used for the manufacture of welded steel tubes (May 20, 2009).

ISO/DIS 10893-10, Non-destructive testing of steel tubes—Part 10: Automated full peripheral ultrasonic testing of seamless and welded (except submerged arc-welded) steel tubes for the detection of longitudinal and/or transversal imperfections (May 20, 2009).

ISO/DIS 10893-11, Non-destructive testing of steel tubes—Part 11: Automated ultrasonic testing of weld seam of welded steel tubes for the detection of longitudinal and/or transversal imperfections (May 20, 2009).

ISO/DIS 10893-12, Non-destructive testing of steel tubes—Part 12: Automated full peripheral ultrasonic thickness testing of seamless and welded (except submerged arc-welded) steel tubes (May 20, 2009).

IEC

29/673/FDIS, IEC 60645-6 Ed.1: Electroacoustics—Audiometric equipment—Part 6: Instruments for the measurement of otoacoustic emissions (April 3, 2009).

29/674/FDIS, IEC 60645-7 Ed.1: Electroacoustics—Audiometric equipment—Part 7: Instruments for the measurement of auditory brainstem responses (April 3, 2009).

100/1517/FDIS, IEC 62365: Digital audio—Digital input-output interfacing—Transmission of digital audio over asynchronous transfer mode (ATM) networks (April 17, 2009).

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

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:

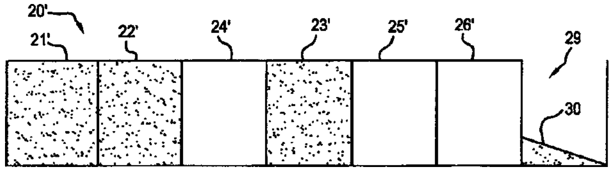
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
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*

7,428,948

43.20.Fn HYBRID AMPLITUDE-PHASE GRATING DIFFUSERS

Peter D'Antonio and Trevor J. Cox, assignors to RPG Diffusor Systems, Incorporated
30 September 2008 (Class 181/293); filed 11 August 2005

The inventor has popularized architectural acoustic products based on quadratic residue gratings (as proposed by Manfred Schroeder) and later those based on binary sequences (with reflecting, 1, or absorbing, 0, elements). This product 22 uses a ternary sequence of elements $(1,0,-1)$ arranged in a random or pseudo-random distribution. The -1 element 27 is described as a quarter wavelength deep well. Various embodiments of the quarter wave length well termination are described. The patent discloses that this combination device, which has the attributes of both an amplitude grating and a reflection phase grating, offers an additional 4 dB of specular attenuation over that from binary sequence devices. A quaternary sequence device is also described. Practitioners who have used, have considered using, or have an interest in number theory based diffraction/reflection/absorbing elements will find the patent interesting and informative. The patent can also be categorized under PACS category 43.20.El. 27 figures, 25 columns of text, and 69 claims are provided.—NAS

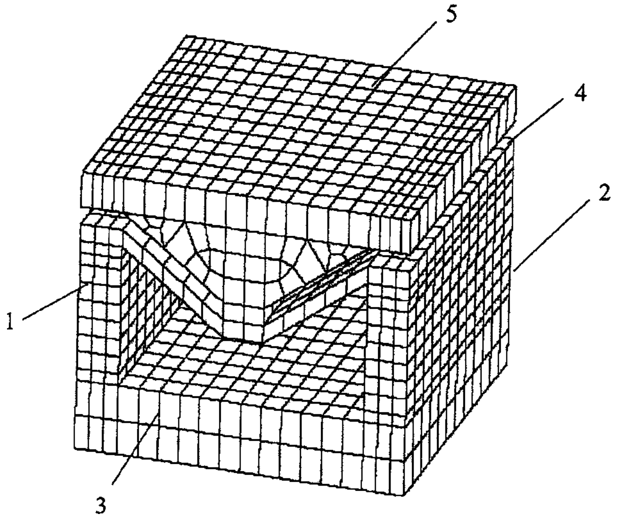


7,453,186

43.30.Yj CANTILEVER DRIVEN TRANSDUCTION APPARATUS

Alexander L. Butler and John L. Butler, assignors to Image Acoustics, Incorporated
18 November 2008 (Class 310/330); filed 17 October 2007

This patent discloses the use of a somewhat novel configuration for an immersion transducer. The transducer presented in the patent is of a flextensional, mechanical lever type utilizing cantilever piezoelectric beams rather than the usual stack or slab beams used in the standard flextensional types. In the figure, plates 1 and 2 are bimorphs or another type of piezoelectric cantilever bending actuator. Base 3 is an inert stiff mass to support the flexing without motion, and piston 5 is a light stiff material that is moved by the action of mechanical lever arms 4. The mechanical amplification of the cantilever motion is probably limited to the range of about 1–4 by the design but seems to allow a significant tuning range. The mechanism is, of course, resonant, so the operating frequency range may be limited. Surprisingly, in water the authors show that nearly flat response is available from about 2 to 18 kHz due to the radiation damping one would guess. Some of the electromechanical model parameters are given along with the response plot, and the patent is generally informative and well written. One would think that this transducer could be scaled up or down in size quite easily (the authors quote results based on a 2×3 in.^{2} piston) and could be used in large directional arrays.—JAH



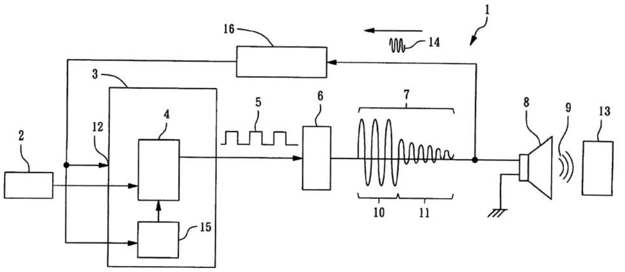
0001-4966/2009/125(5)/3475/17/$25.00
© 2009 Acoustical Society of America

7,453,762

43.35.Yb APPARATUS AND METHOD OF AUTO FREQUENCY CALIBRATION FOR TRANSDUCER

Kuang-Yeu Lin and Lung-Chieh Chen, assignors to Holtek Semiconductor, Incorporated
18 November 2008 (Class 367/13); filed 28 May 2008

Small ultrasonic transducers are used as combination transmitters and receivers in electronic range finders and cameras. Since such devices must function accurately under a variety of environmental conditions, precision construction is required to maintain the desired operating frequency. This patent describes an arrangement whereby the frequency of the exciting signal is automatically calibrated to accommodate any output signal variation. Although some of the terminology is puzzling, the patent should be understandable to those interested in this field.—GLA

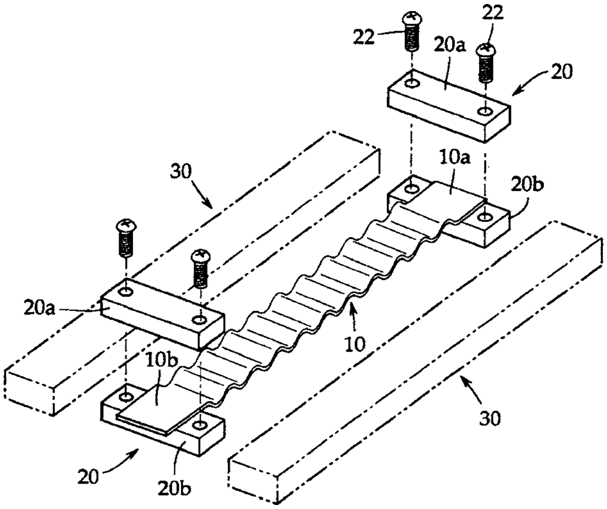


7,430,297

43.38.Dv MOVING RIBBON MICROPHONE

Hiroshi Akino, assignor to Kabushiki Kaisha Audio-Technica
30 September 2008 (Class 381/176); filed in Japan 2 March 2004

To improve the corrosion resistance and lower the resistance of corrugated aluminum foil ribbon element 10, a gold deposited film is applied to the aluminum foil. The gold deposited film is at least 500 nm thick and, it appears, no more than 10% of the total diaphragm mass (although the wording in the patent is more convoluted, unclear, and confusing).—NAS

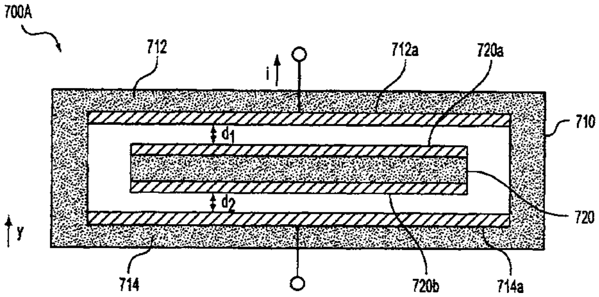


7,453,269

43.38.Dv MAGNETIC MEMS SENSOR DEVICE

Jong-hwa Won *et al.*, assignors to Samsung Electronics Company, Limited
18 November 2008 (Class 324/658); filed in Republic of Korea 11 May 2004

This little device is claimed to be able to sense nearly anything from altitude to magnetic fields. The basic operation is simple: Magnetic layers 720a and 720b repel layers 712a and 714a resulting in a levitated disk (or other shapes as disclosed in the patent) whose position d1, d2 can be read out with capacitance measurements. Unfortunately, there are instabilities inherent in this device that will try to eject the disk like a watermelon seed and cause it to crash. Even if stabilized electronically with feedback, the resulting sensor is likely to be considerably less sensitive than a good elastic mounting. There are no details of prototype sensors given; the only information given is on permanent magnetic film formulations that may be used and how to pole them.—JAH



7,450,474

43.38.Hz DIAGNOSTIC SYSTEM AND METHOD FOR TRANSDUCERS

Jerry G. Klein and Anthony L. Scoca, assignors to Lockheed Martin Corporation
11 November 2008 (Class 367/153); filed 25 February 2004

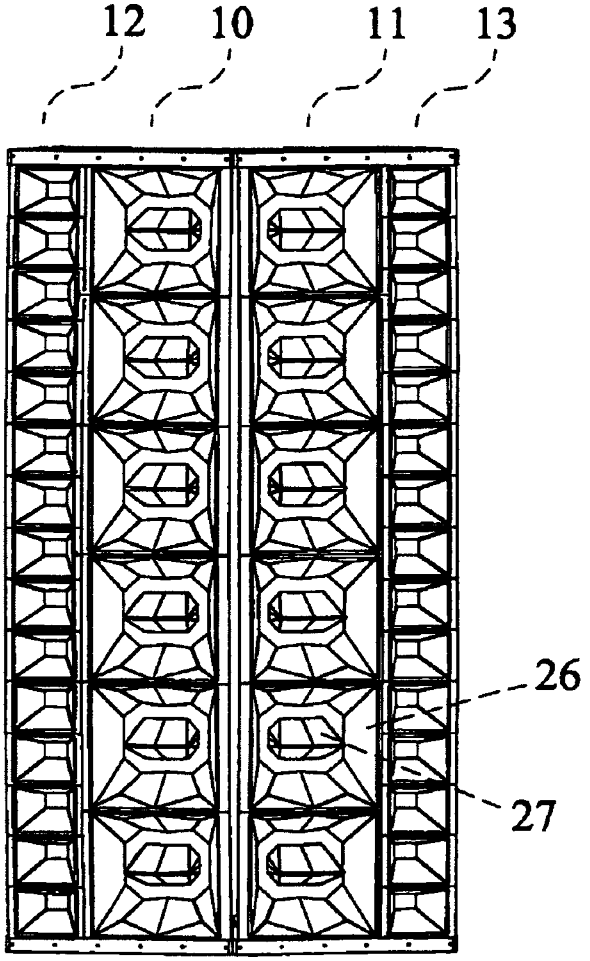
There is not much contained in this disclosure. In a multielement transducer, how do you determine the "acoustic center"? Basically, the inventor recommends measuring and using a color display.—MK

7,454,029

43.38.Hz LOUDSPEAKER ARRAY

Anthony J. Andrews, Surrey, United Kingdom
18 November 2008 (Class 381/335); filed in United Kingdom 20 March 2003

The patent describes a two-dimensional loudspeaker array that is said to provide greater output than a conventional curved line array and to minimize unwanted side lobes. Unfortunately, it is difficult to pin down exactly what has been patented. Each of the two independent patent claims (of a total of 24) consists of a single sentence more than 200 words long and almost impossible to follow.—GLA

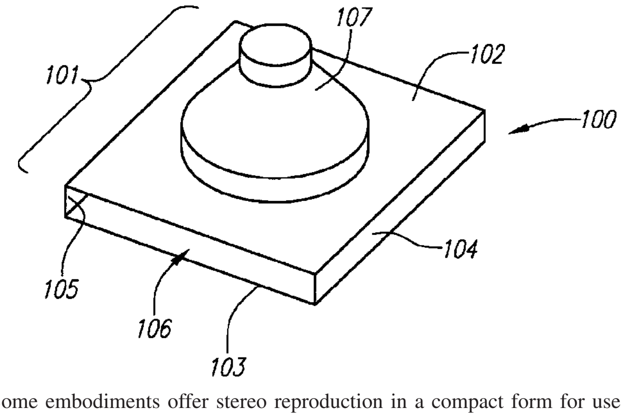


7,433,483

43.38.Ja NARROW PROFILE SPEAKER CONFIGURATIONS AND SYSTEMS

Lawrence R. Fincham, assignor to THX Limited
7 October 2008 (Class 381/337); filed 8 September 2004

More than a few variations of loudspeaker systems, such as 100, are described that are said to offer a relatively narrow sound output region, all of which use a slot 106 or similar type orifice from which the sound emanates.



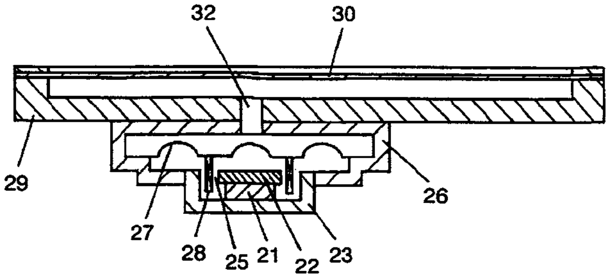
Some embodiments offer stereo reproduction in a compact form for use in passenger compartments of automobiles, with the drive circuit for same also described in the patent. Some of the embodiments remind this reviewer of the so called "slot radiator" sub-woofers used in discos for about 30 years. 30 figures, 33 columns of text, and 33 claims are provided.—NAS

7,460,679

43.38.Ja FLAT PANEL SPEAKER SYSTEM WITH A COATED DIAPHRAGM

Satoshi Itoh *et al.*, assignors to Panasonic Corporation
2 December 2008 (Class 381/184); filed in Japan 6 October 2003

This patent is characterized more by what is omitted than what is explained. We start out with a miniature moving-coil loudspeaker loaded by a small front chamber. The chamber is coupled (somehow) to a second chamber behind a thin, flexible diaphragm 30. (The diaphragm may also serve as a display or touch screen.) One surface of the diaphragm is coated with something, and the opposite surface is coated with something else. The basic geometry creates a two-chamber bandpass loudspeaker system, but the patent includes no performance data. Instead, we are told, "With the structure, an advantage of each material is selectively obtained according to requested performance, and a plurality of effects or synergy effect can be expected in the multi-layered coating."—GLA

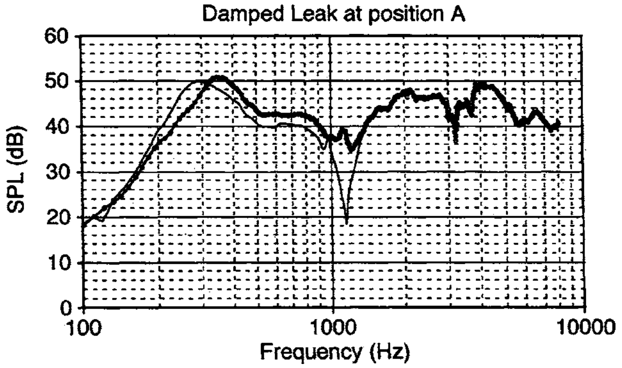


7,461,718

43.38.Ja LOUDSPEAKER ENCLOSURE INCORPORATING A LEAK TO COMPENSATE FOR THE EFFECT OF ACOUSTIC MODES ON LOUDSPEAKER FREQUENCY RESPONSE

Stephane Dedieu and Philippe Moquin, assignors to Mitel Networks Corporation
9 December 2008 (Class 181/148); filed in United Kingdom 10 December 2003

Tiny loudspeakers in small cellular phones share space with other components. If a sealed back chamber can be provided, it probably will be small and far from box-shaped. Strong resonances in such an enclosure introduce correspondingly strong peaks and dips in response. This patent teaches that strategically placed air leaks can counteract the effects of box resonances. The patent includes "before" and "after" curves that demonstrate the improvement.—GLA

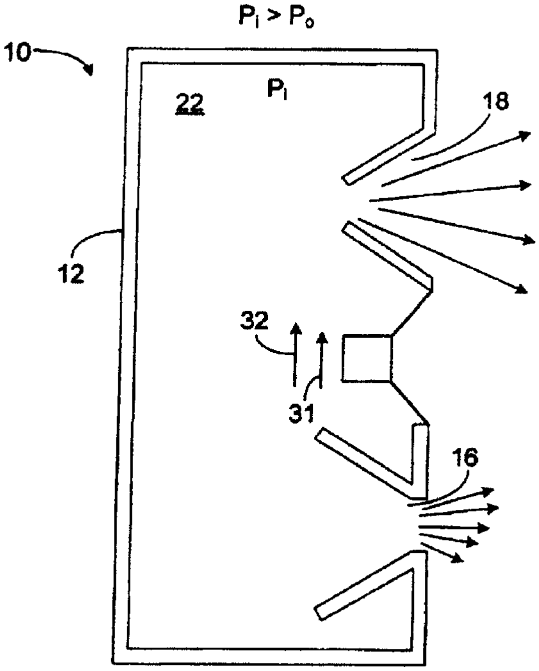


7,463,744

43.38.Ja PORTING

Robert Preston Parker *et al.*, assignors to Bose Corporation
9 December 2008 (Class 381/161); filed 31 October 2003

A well-known problem with conventional vented loudspeakers is partial rectification, producing a static offset that moves the loudspeaker cone away from its center position. The geometry disclosed in this patent promotes rectification rather than counteracts it. If two such vents are used in a push-pull configuration, then air flow through the cabinet is induced, cooling the loudspeaker or internal amplifier.—GLA

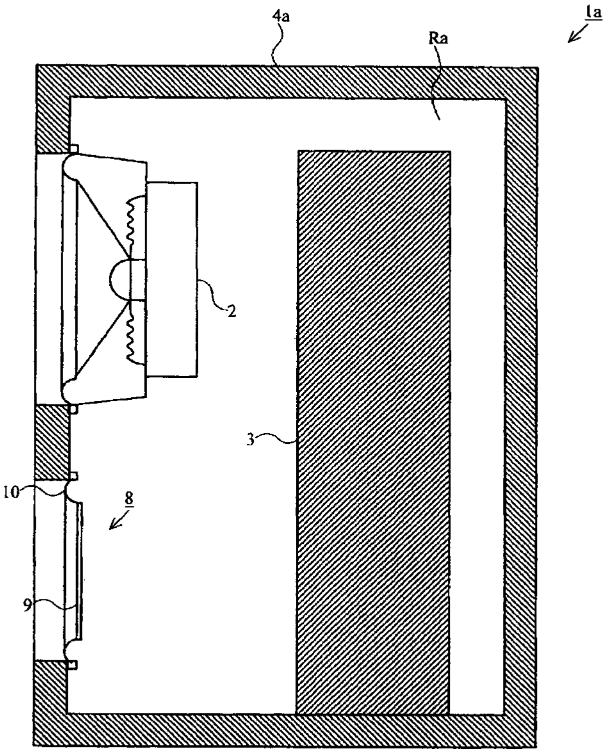


7,463,747

43.38.Ja LOUDSPEAKER SYSTEM

Mitsukazu Kuze *et al.*, assignors to Panasonic Corporation
9 December 2008 (Class 381/345); filed in Japan 31 March 2004

A large lump of activated carbon inside a loudspeaker enclosure effectively increases the internal volume by absorbing air molecules during the compression phase. However, if the box is vented, then moisture can enter and is readily absorbed by the carbon, which then becomes incapable of absorbing additional air molecules. Prior art has dealt with the problem by enclosing the carbon in an airtight plastic bag. This patent uses a different approach, substituting an airtight passive radiator 8 for the vent, thus making the enclosure impervious to moisture.—GLA

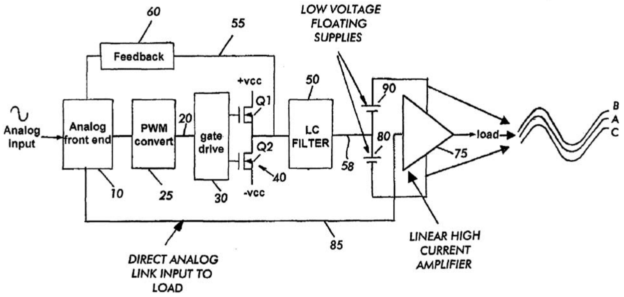


7,456,686

43.38.Lc CLASS AD AUDIO AMPLIFIER

Bruno Nadd, assignor to International Rectifier Corporation
25 November 2008 (Class 330/10); filed 21 September 2006

Class D (pulse width modulation) audio amplifier design has advanced dramatically in the past few years. One stumbling block in achieving very low distortion is the difficulty of providing negative feedback to the analog input stage. This patent discloses what might be described as a feedforward method of reducing distortion at the output. The PWM output is filtered and then used to bias a linear current amplifier whose input is the original analog signal. It appears that the PWM gates supply most of the power and the current amplifier counteracts distortion.—GLA

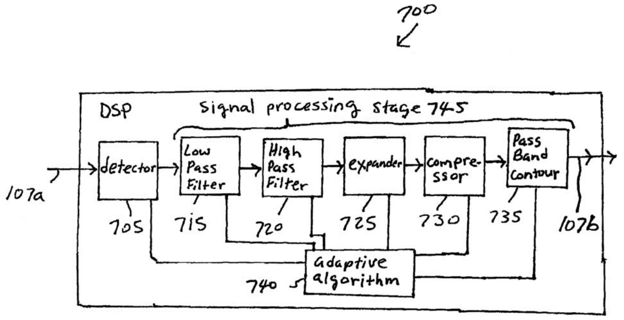


7,457,757

43.38.Lc INTELLIGIBILITY CONTROL FOR SPEECH COMMUNICATIONS SYSTEMS

Iain McNeill and Robert M. Khamashta, assignors to Plantronics, Incorporated
25 November 2008 (Class 704/500); filed 30 May 2002

The patent points out that optimum telephone communication via a headset depends on background conditions that can vary widely. If the user is in a quiet office, natural voice quality at reasonable loudness will be appropriate, whereas a restricted frequency range with substantial compression may be required to provide intelligibility in a noisy automobile. A method is described that dynamically identifies background noise characteristics and then processes the signal for best possible clarity and intelligibility.—GLA

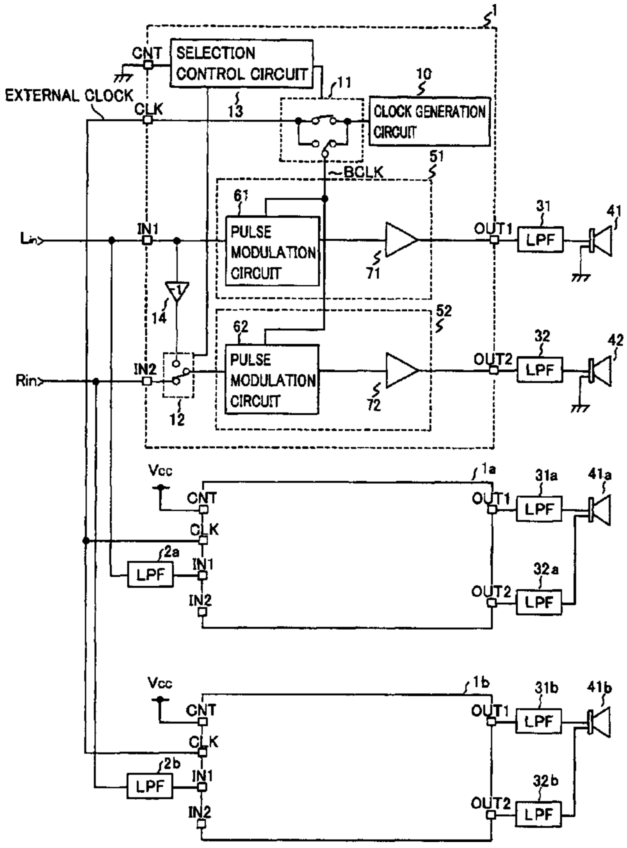


7,459,968

43.38.Lc AUDIO POWER AMPLIFIER IC AND AUDIO SYSTEM PROVIDED WITH THE SAME

Shigeji Ohama *et al.*, assignors to Rohm Company, Limited
2 December 2008 (Class 330/10); filed in Japan 21 September 2004

Class D audio power amplifiers are commonly found in commercial paging and sound reinforcement systems. When several class D amplifiers are used in a multi-channel installation, beat frequencies from their asynchronous clocks may create audible noise. This patent describes a class D module that can be synchronized to an external master clock.—GLA



7,436,336

43.38.Md ANALOG DIGITAL CONVERTER (ADC) HAVING IMPROVED STABILITY AND SIGNAL TO NOISE RATIO (SNR)

Morteza Vadipour, assignor to Broadcom Corporation
14 October 2008 (Class 341/143); filed 19 December 2006

In a Delta-Sigma analog to digital converter, the time delay between the loop comprised of the quantizer and the digital to analog converters is compensated by adding a feedback path and then adding a compensating pole to cancel out the zero. The resulting circuit is claimed to be less susceptible to unstable responses.—MK

7,436,957

43.38.Md AUDIO CASSETTE EMULATOR WITH CRYPTOGRAPHIC MEDIA DISTRIBUTION CONTROL

Addison M. Fischer and Robert L. Protheroe, both of Naples, Florida
14 October 2008 (Class 380/53); filed 29 July 1999

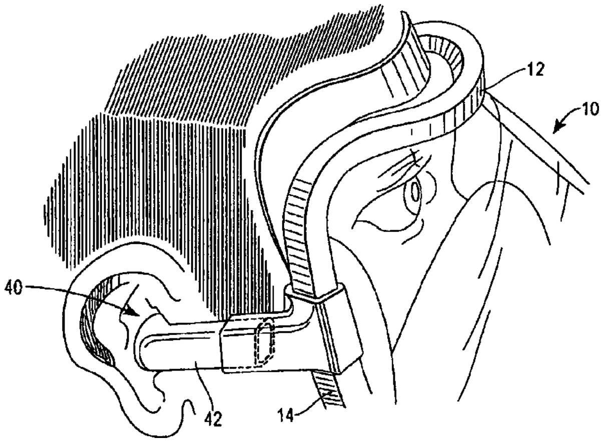
The inventors missed the memo that declared audio cassette tapes a dead format (in their defense, it was filed in 1999). And so, they propose a cryptographic device that can encrypt audio formats. Maybe when it was filed it was plausible, but now it is anachronistic.—MK

7,453,366

43.38.Si PROGRAMMABLE EARPIECE

William L. Grilliot and Mary I. Grilliot, assignors to Morning Pride Manufacturing, LLC
18 November 2008 (Class 340/584); filed 11 October 2005

Body temperature can be measured by a sensor in the ear canal. Insertable communication earbuds fit into the ear canal. The two functions are combined in this earpiece intended to be worn by firefighters or emergency rescue workers. If the temperature exceeds a preset value, the user hears a warning alarm.—GLA

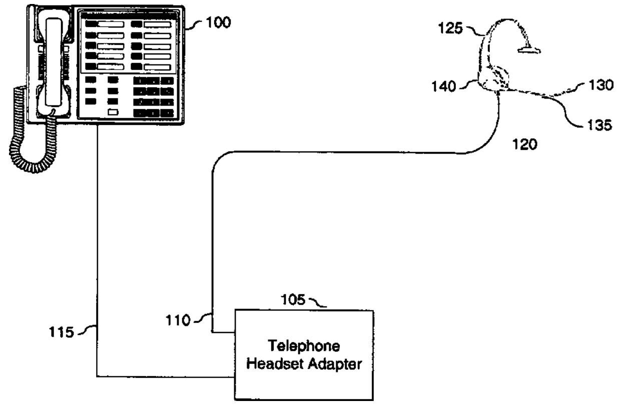


7,457,644

43.38.Si MICROPHONE POSITION AND SPEECH LEVEL SENSOR

James F. Bobisuthi *et al.*, assignors to Plantronics, Incorporated
25 November 2008 (Class 455/570); filed 7 November 2006

This is a continuation of two earlier patents, all dealing with speech pickup from headset microphones. This patent describes circuitry that senses whether the headset wearer is talking and then triggers a warning if the signal level is too low.—GLA

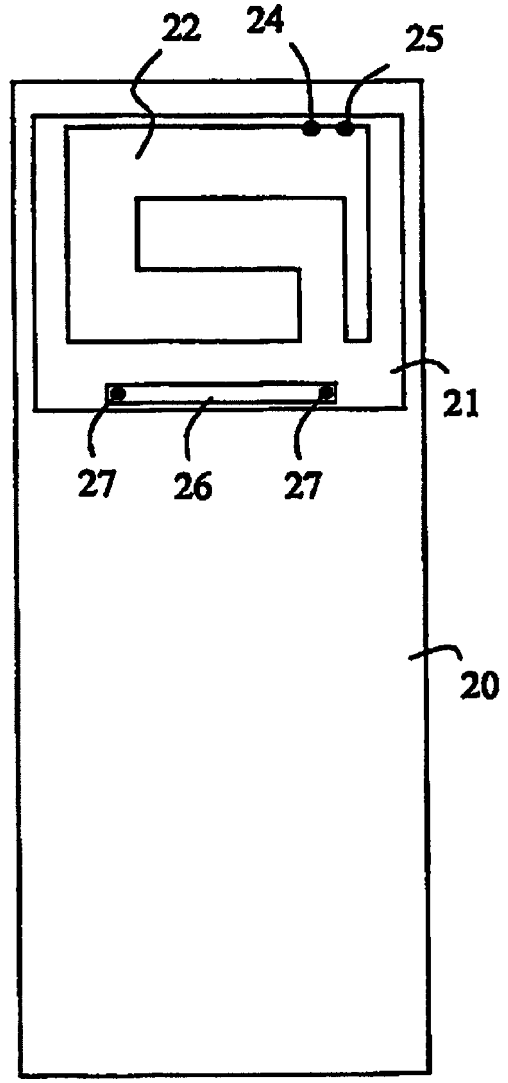


7,460,074

43.38.Si COMMUNICATION TERMINALS HAVING INTEGRATED ANTENNA AND SPEAKER ASSEMBLIES

Zhinong Ying and Wanqing Shi, assignors to Sony Ericsson Mobile Communications AB
2 December 2008 (Class 343/702); filed in the European Patent Office 20 June 2003

In most cellular telephones the loudspeaker competes for space with the radio frequency antenna. However, if a simple printed circuit is added to a flat diaphragm then the two functions can be combined, allowing further miniaturization.—GLA

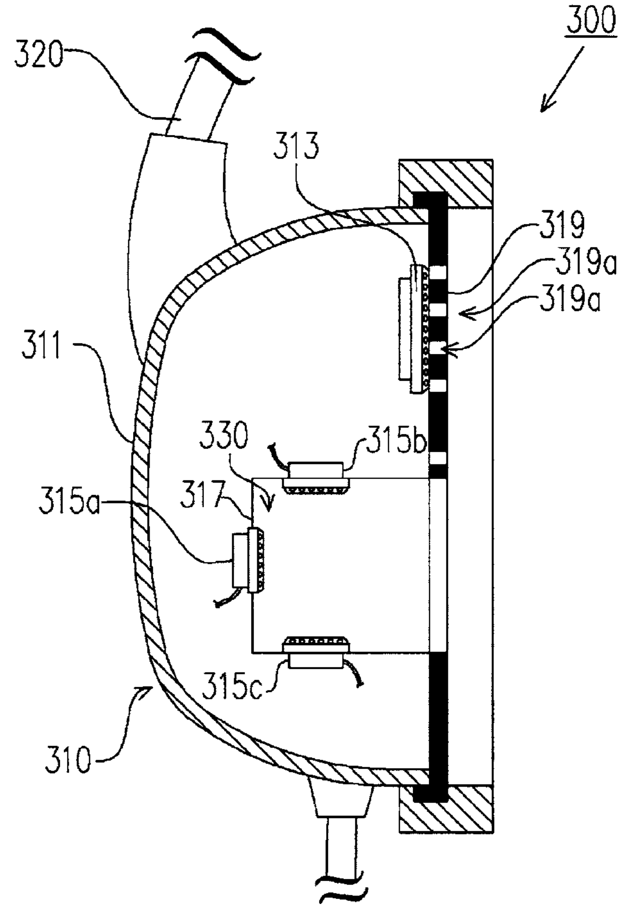


7,463,748

43.38.Si [EARPHONE STRUCTURE WITH A COMPOSITE SOUND FIELD]

Bill Yang, assignor to Cotron Corporation
9 December 2008 (Class 381/370); filed in Taiwan 22 March 2004

Yes, the title of this patent really is enclosed by brackets for reasons that only the U.S. Patent Office can explain. The invention disclosed is yet another earcup designed as a miniature listening room, complete with multiple surround sound loudspeakers. More than 15 variants are described.—GLA

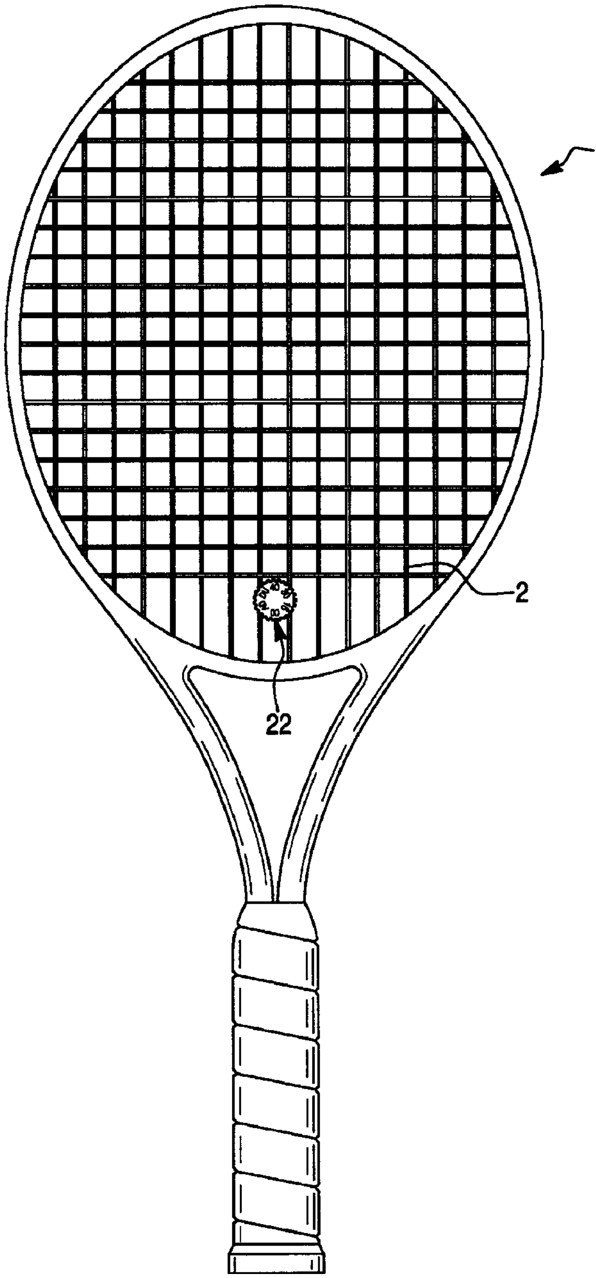


7,427,245

43.40.Tm ELECTRONIC RACQUET SCORE KEEPER AND VIBRATION DAMPER

Darren Bawden Hickey, Mountain View, California
23 September 2008 (Class 473/522); filed 19 November 2007

Various and sundry versions of a combination vibration damping/score keeper 22 for a tennis racquet 1 are disclosed. The vibration damping elements of the device are well described in many prior patents (such as U.S. Patent No. 7335118 reviewed in J. Acoust. Soc. Am. **124**, 1900, 2008) and in most texts on vibration control. The score keeping element can be implemented using one or two (if you want or need to keep the score for both players) passive embodiments as well as several electronic (which are disclosed in flow charts) embodiments. 11 diagrams, 11 columns of text, and 20 claims are provided.—NAS

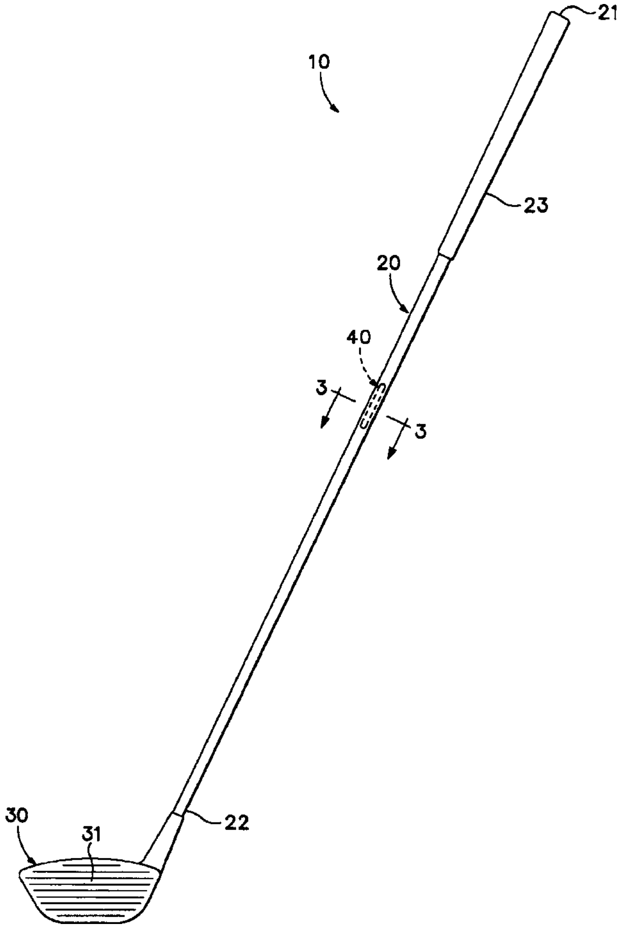


7,435,187

43.40.Tm GOLF CLUB INCORPORATING A DAMPING ELEMENT

John Thomas Stites and Gary Gene Tavares, assignors to Nike, Incorporated
14 October 2008 (Class 473/318); filed 19 December 2003

Damping element 40, which can be fluid-filled, or a polymer foam moderates the vibration induced in golf club 10 after said club, specifically head 30, impacts the golf ball, after the club head traverses a generally curved or arcuate path. Damping element 40 may be located anywhere along the interior of shaft 20, can be of any length up to the total length of shaft 20, and can take any number of shapes.—NAS

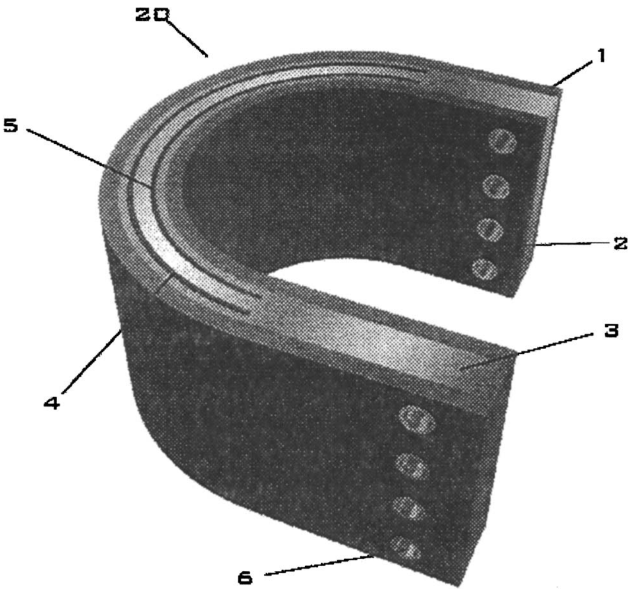


7,451,966

43.40.Tm ISOLATOR MOUNT FOR SHOCK AND VIBRATION

Gareth J. Knowles and Bruce Bower, Williamsport, Pennsylvania
18 November 2008 (Class 267/136); filed 2 July 2002

The isolator described in this patent is intended to be flexible, so as to attenuate small-amplitude vibrations, but to incorporate considerable stiffness in order to isolate shocks that are associated with large excursions at relatively low frequencies. The isolator 20 is generally in the shape of the letter U and has inserts 4 and 5 embedded in a layer 3 that is sandwiched between layers 1 and 2. Layers 1 and 2 are of materials, such as spring steel or thermoplastics, that are capable of withstanding repeated deflections at large strains; these layers are intended to provide structural rigidity during normal operating conditions. The primary function of layer 3, composed of a composite, polymer, or elastomer, is to provide sufficient stiffness to transfer strain to inserts 4 and 5 and also significant damping at higher frequencies. Inserts 4 and 5 may be composed of magneto-mechanical materials, shape memory alloys, or super-elastic alloys.—EEU

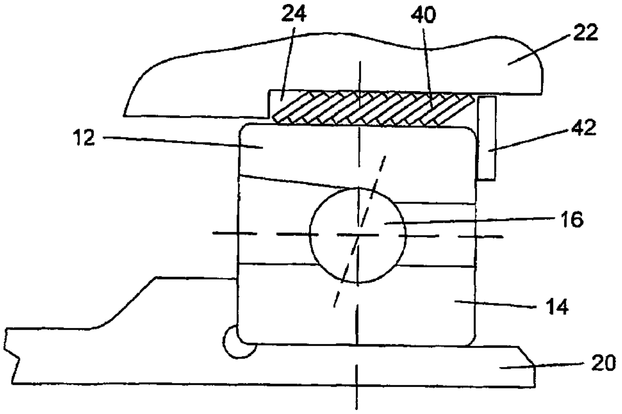


7,452,135

43.40.Tm FRICTION DAMPER FOR A BEARING

Kenneth W. Holsaple, assignor to Florida Turbine Technologies, Incorporated
18 November 2008 (Class 384/535); filed 18 July 2006

Support of a bearing that provides damping of the bearing and alignment in a housing is accomplished by a stack of cone-shaped annular plates 40 positioned between the housing 22 and the outer race 12 of the bearing. The annular plates can be formed from a single coil and can be coated for enhancement of the damping. The offset angle of the plates can be varied to adjust the damping and the spring rate of the support.—EEU



7,461,624

43.40.Tm COMPENSATING SHAFT DRIVE

Thomas Ullein and Reinhard Koch, assignors to Schaeffler KG
9 December 2008 (Class 123/192.2); filed in Germany 14 July 2004

Compensating shafts are used to counter-balance the unbalanced inertia forces of piston assemblies in internal combustion engines. In general, such compensating shafts in essence comprise eccentric masses that rotate at appropriate frequencies. Since the unbalanced inertia forces do not vary harmonically, the present patent employs shafts whose rotational speeds vary throughout a cycle, so that their unbalance forces also vary anharmonically. It accomplishes this by having the compensating shafts driven via non-circular toothed wheels by means of a timing belt that is connected to the primary shaft by a circular toothed wheel. Oval toothed wheels are recommended for compensating shafts of four-cylinder engines, for example.—EEU

7,451,662

43.40.Yq VIBRATION TYPE MEASUREMENT TRANSDUCER

Wolfgang Drahm *et al.*, assignors to Endress + Hauser Flowtec AG
18 November 2008 (Class 73/861.357); filed in the European Patent Office 25 February 2005

This patent pertains to a device for measuring properties (such as mass flow rates, density, and viscosity) of media flowing in pipelines. It consists of a section of medium-conveying pipe that is set into lateral vibration by a transducer, of additional transducers that sense the resulting vibrations, and of a signal processor. Such measurement arrangements are well known; the novelty claimed in this patent is that at least some of the transducers attached to the sensing pipe are sintered onto that pipe, thus simplifying assembly and increasing the system's reliability.—EEU

7,459,831

43.40.Yq VIBRATING DEBRIS REMOVER

Damian R. Ludwiczak, Orlando, Florida
2 December 2008 (Class 310/321); filed 1 August 2006

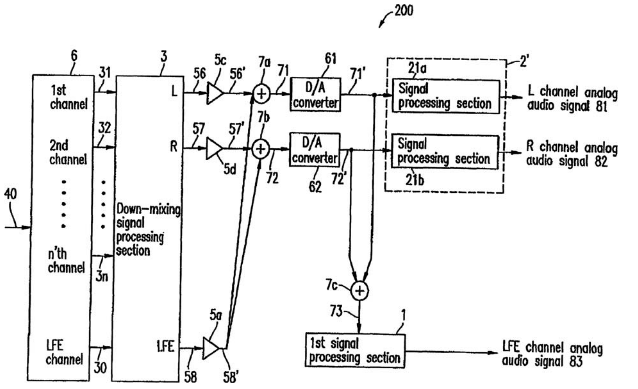
This patent pertains to a device that may be permanently attached or briefly placed in contact with the edge of a surface, such as an automobile window, from which debris (or ice) is to be removed. The device consists of a vibrator and a coupler, which may be attached to the edge of the surface eccentrically so that it generates transverse and longitudinal waves, with the intent of breaking the bond between the surface and the debris.—EEU

7,236,838

43.55.Jz SIGNAL PROCESSING APPARATUS, SIGNAL PROCESSING METHOD, PROGRAM AND RECORDING MEDIUM

Takashi Katayama *et al.*, assignors to Matsushita Electric Industrial Company, Limited
26 June 2007 (Class 700/94); filed in Japan 29 August 2000

This has to come under the patently obvious category. Simply put, they propose downmixing the low frequency effect channel into the left and right channels and then applying analog filters after the digital to analog converters. They do not even bother with digital filtering in the downmix. Even for 2001 (filing date), this is archaic.—MK


200
6
3
1st channel
2nd channel
n'th channel
LFE channel
31
32
3n
30
40
Down–mixing signal processing section
L
R
LFE
56
5c
56'
57
57'
5d
5a
58
58'
7a
7b
7c
71
72
73
61
62
71'
72'
D/A converter
D/A converter
2'
21a
21b
Signal processing section
Signal processing section
1
1st signal processing section
L channel analog audio signal 81
R channel analog audio signal 82
LFE channel analog audio signal 83

7,424,117

43.55.Lb SYSTEM AND METHOD FOR GENERATING SOUND TRANSITIONS IN A SURROUND ENVIRONMENT

Tilman Herberger and Titu Tost, assignors to Magix AG
9 September 2008 (Class 381/61); filed 25 August 2003

In a disco, the overlap between two songs is an opportunity for sonic variation. Previous patents have addressed beat matching. This disclosure proposes adding motion to the spatial parameters.—MK

7,439,873

43.55.Lb SYNTHETICALLY GENERATED SOUND CUES

Brian J. Tillotson, assignor to The Boeing Company
21 October 2008 (Class 340/692); filed 20 October 2006

Consider the control of unmanned aerial vehicles. The judicious use of spatial auditory cues could assist the operator(s) in maintaining spatial separation as well as use of varying pitch and timbre to model speed and position relative to the controller.—MK

7,419,436

43.58.Wc SOUND PRODUCING PLAY APPARATUS

Grant R. Ballin, assignor to WonderWorx, LLC
2 September 2008 (Class 472/118); filed 30 January 2006

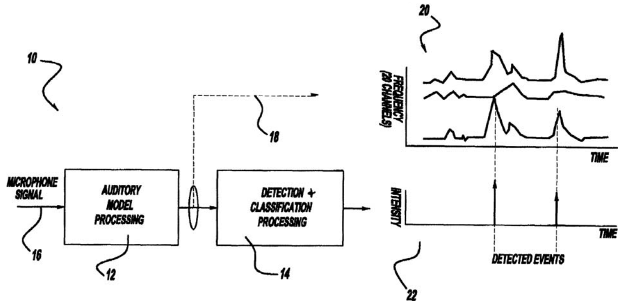
Combines a swing set teeter-totter with various passive sound generation techniques including a "rain stick" and using an air reservoir with a piston to generate sound. It is nice to see an entertainment device that is not powered by batteries.—MK

7,457,422

43.60.Bf METHOD AND IMPLEMENTATION FOR DETECTING AND CHARACTERIZING AUDIBLE TRANSIENTS IN NOISE

Jeffry Allen Greenberg *et al.*, assignors to Ford Global Technologies, LLC
25 November 2008 (Class 381/56); filed 29 November 2001

Using envelope extraction in multiple frequency bands, an excitation signal is produced that is indicative of the estimated acoustic activity at a particular location, such as the inside of an automobile. The excitation signal is processed to identify impulsive sounds over time. A detection signal is produced if the noise is deemed audible over temporal masking, and the impulsive sounds are characterized for future identification.—DAP

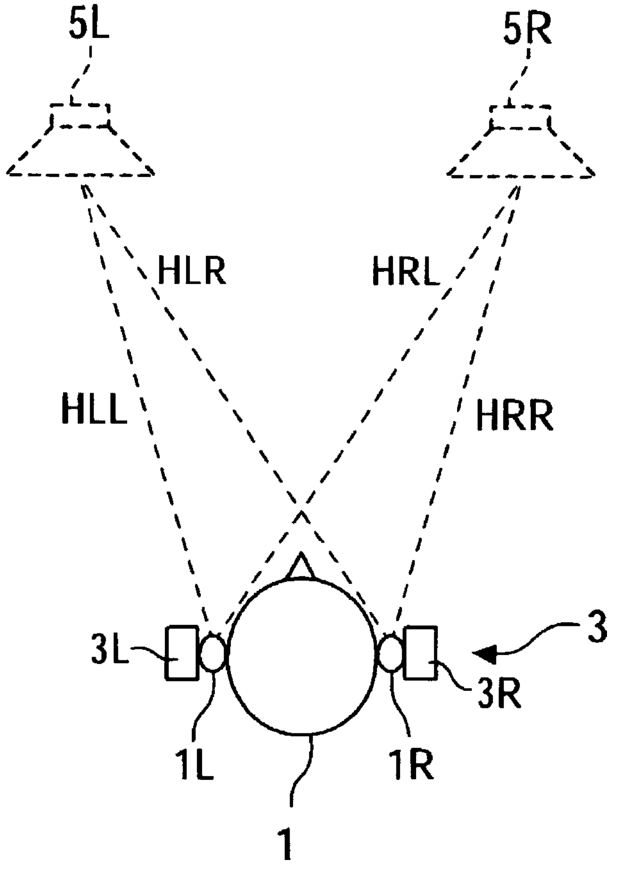


7,454,026

43.66.Qp AUDIO IMAGE SIGNAL PROCESSING AND REPRODUCTION METHOD AND APPARATUS WITH HEAD ANGLE DETECTION

Yuji Yamada, assignor to Sony Corporation
18 November 2008 (Class 381/310); filed in Japan 28 September 2001

The goal is to provide continued good localization capability under headphones without audible artifacts when the direction the listener is facing changes. Signal processing is performed to localize the sound image direction of the input signal in at least two positions on both sides of a target direction. The two processed sound signals are delayed by amounts depending on the target position and a reference position to produce two delayed signals which, when added proportionally, produce two weighted delayed output signals. Any resulting spectral changes in the input signal resulting from this processing are compensated for.—DAP

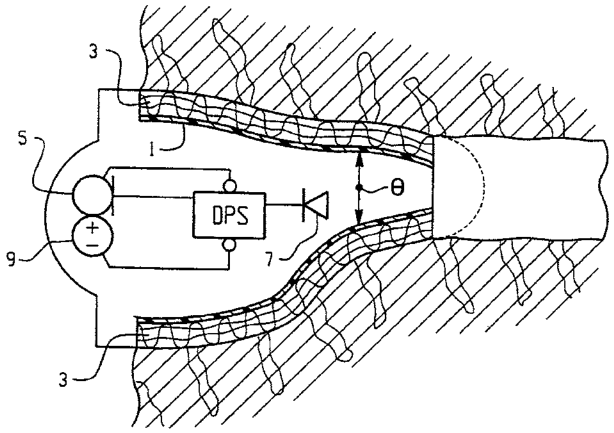


7,454,028

43.66.Ts IN-THE-EAR HEARING DEVICE

Herbert Bachler *et al.*, assignors to Phonak AG
18 November 2008 (Class 381/322); filed 17 April 2006

On the periphery of a custom shell is placed a removable, sock-shaped body made of an elastic, fibrous textile material that snugly conforms to the wearer's ear canal. The textile contains a medication to treat disease in the ear canal, and the medication is intended to diffuse through skin.—DAP

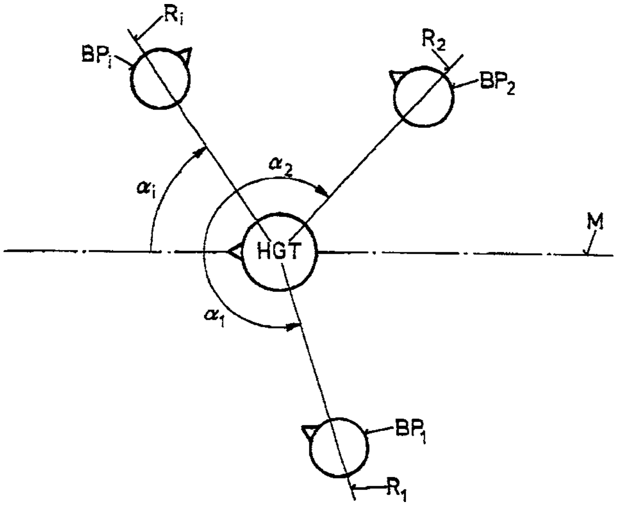


7,457,426

43.66.Ts METHOD TO OPERATE A HEARING DEVICE AND ARRANGEMENT WITH A HEARING DEVICE

Peter Drtina, assignor to Phonak AG
25 November 2008 (Class 381/313); filed 14 June 2002

The transfer function of a hearing device is modified to maximize its directional sensitivity toward a desired reference person or persons. The reference person may contain a device that emits a predefined reference signal which is used by the hearing device to determine the desired direction. Alternatively, to establish the desired direction, the hearing device may use automatic speaker recognition to identify a predetermined code word spoken by the reference person(s).—DAP

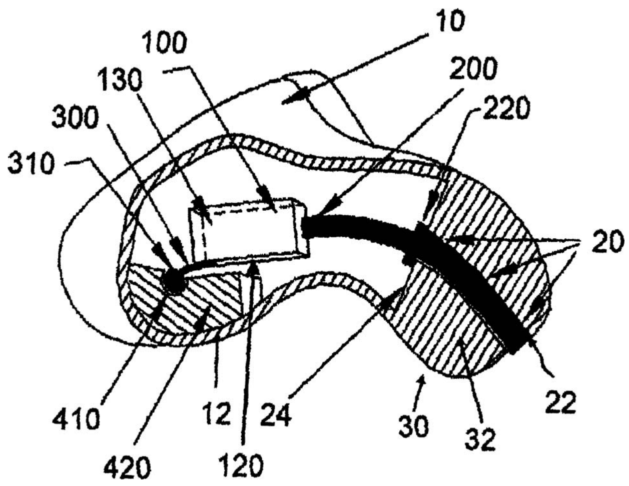


7,460,680

43.66.Ts FEEDBACK REDUCING RECEIVER MOUNT AND ASSEMBLY

Oleg Saltykov, assignor to Siemens Hearing Instruments, Incorporated
2 December 2008 (Class 381/324); filed 30 June 2003

A hearing aid receiver suspended on a receiver tube is further stabilized to prevent mechanical oscillation by flexibly tethering it to the hearing aid housing. The tether could consist of a resilient U-shaped cradle for the receiver and the other end could have a ball that mates with a socket in the hearing aid housing.—DAP

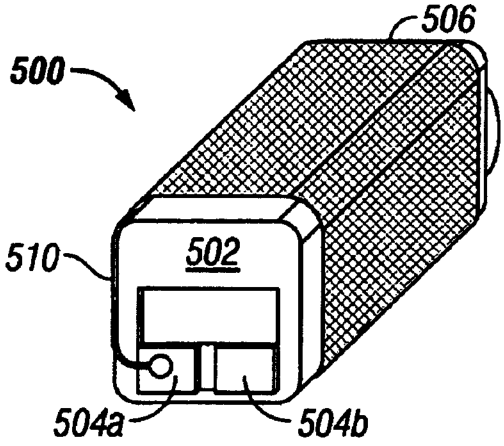


7,460,681

43.66.Ts RADIO FREQUENCY SHIELDING FOR RECEIVERS WITHIN HEARING AIDS AND LISTENING DEVICES

Onno Geschiere *et al.*, assignors to Sonion Nederland B.V.
2 December 2008 (Class 381/324); filed 20 July 2004

To help prevent signals emanating from the hearing aid receiver from interfering with a wireless communication link to/from the hearing aid, an electrically-conductive, metallic shield is added around the receiver and is connected to the hearing aid audio signal processing circuitry.—DAP

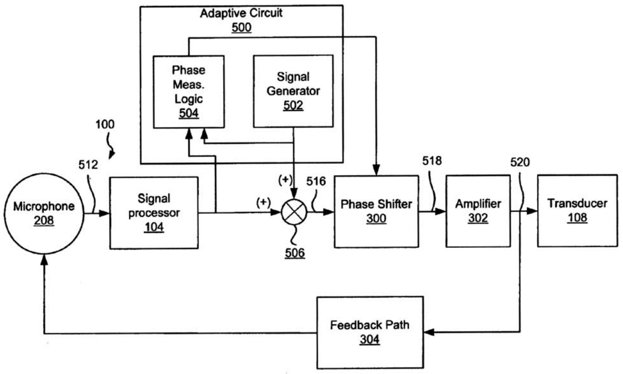


7,463,745

43.66.Ts PHASE BASED FEEDBACK OSCILLATION PREVENTION IN HEARING AIDS

Scott Allan Miller III, assignor to Otologic, LLC
9 December 2008 (Class 381/318); filed 9 April 2004

The phase of the acoustic feedback signal of a hearing aid *in situ* is determined during fitting. Without modifying the gain, the phase of the acoustic feedback signal is shifted to prevent it from being 0° (or an integer multiple of 360°) at frequencies at which the magnitude of the open loop transfer function is near or greater than unity.—DAP

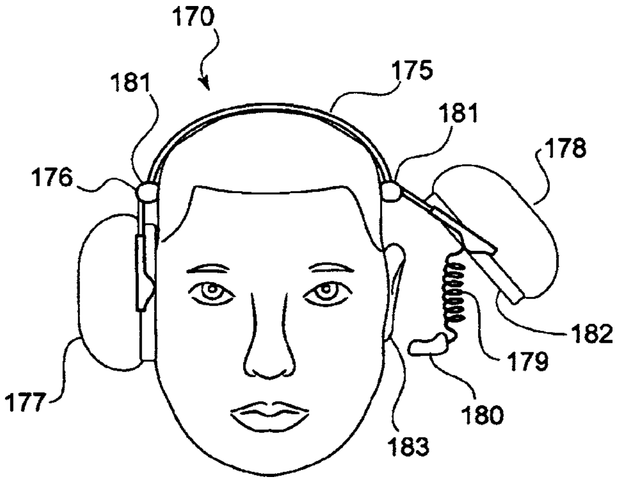


7,457,428

43.66.Vt DOUBLE HEARING PROTECTION DEVICE

Michael A. Vaudrey *et al.*, assignors to Adaptive Technologies, Incorporated
25 November 2008 (Class 381/372); filed 27 July 2006

A picture is worth a thousand words. The figure says it all.—JE

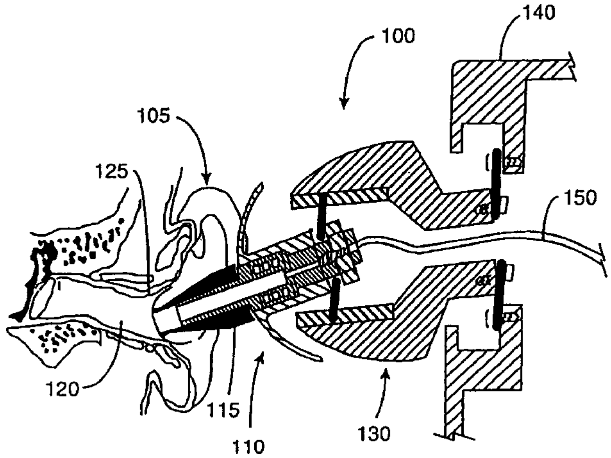


7,452,337

43.66.Yw HAND-HELD HEARING SCREENER APPARATUS

Steven J. Iseberg, assignor to Etymotic Research, Incorporated
18 November 2008 (Class 600/559); filed 15 January 2004

A test probe containing one or more microphones is used during infant hearing screening to pick up otoacoustic emissions. To prevent the user's hand movements from generating vibrational noise and to improve the acoustic seal of the eartip, the test probe is mechanically isolated via elastic coupling from the hearing screener housing.—DAP



7,457,744

43.72.Ar METHOD OF ESTIMATING PITCH BY USING RATIO OF MAXIMUM PEAK TO CANDIDATE FOR MAXIMUM OF AUTOCORRELATION FUNCTION AND DEVICE USING THE METHOD

Mi-suk Lee and Dae-hwan Hwang, assignors to Electronics and Telecommunications Research Institute
25 November 2008 (Class 704/207); filed in Republic of Korea 10 October 2002

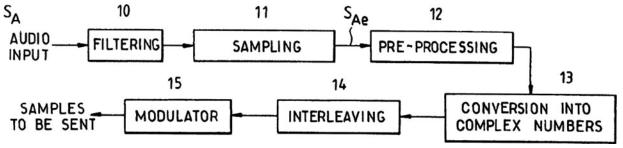
Speech collected by small personal devices is almost always encoded in some form before anything else is done with it. As a result, considerable effort is devoted to methods of analysis which do not require unwinding the encoding before other analyses can be done. The example described here is pitch detection. The system is essentially the well-known autocorrelation type of pitch analyzer, but it works with a perceptually weighted speech signal. Following the correlation, multiple candidates are collected, and a final decision is based on both the peak correlation values and the patterns of time intervals between the correlation maxima.—DLR

7,453,951

43.72.Gy SYSTEM AND METHOD FOR THE TRANSMISSION OF AN AUDIO OR SPEECH SIGNAL

Pierre André Laurent and Cédric Demeure, assignors to Thales
18 November 2008 (Class 375/295); filed in France 19 June 2001

To both improve transmitted signal quality in a given amount of bandwidth at low signal-to-noise ratio conditions and optimize spectral efficiency, speech or audio analog signals are transmitted as symbols of cells applied to a digital modulation such as orthogonal frequency division multiplexing. For example, analog signals are sampled and distributed as complex numbers for transmission as independently-modulated time/frequency cells using a digital modem.—DAP

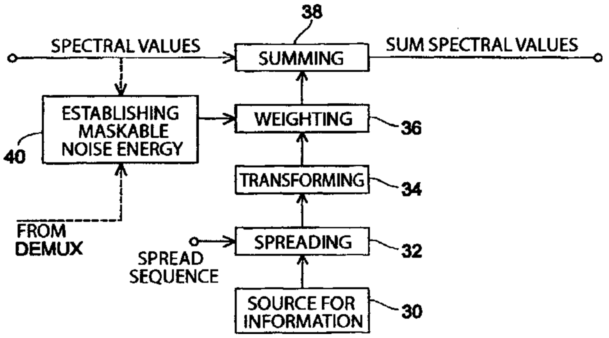


7,454,327

43.72.Gy METHOD AND APPARATUS FOR INTRODUCING INFORMATION INTO A DATA STREAM AND METHOD AND APPARATUS FOR ENCODING AN AUDIO SIGNAL

Christian Neubauer *et al.*, assignors to Fraunhofer-Gesellschaft zur Foerderung der angewandtren Forschung e.V.
18 November 2008 (Class 704/200.1); filed in Germany 5 October 1999

Data representing the short term spectrum of an audio signal are introduced into the data stream as a spread information signal. The spectral values of this signal are weighted with a psychoacoustic maskable noise energy so that the energy of the introduced information is equal to or much below the psychoacoustic masking threshold. The weighted information signal and the short term spectral values are summed to obtain the processed data stream.—DAP

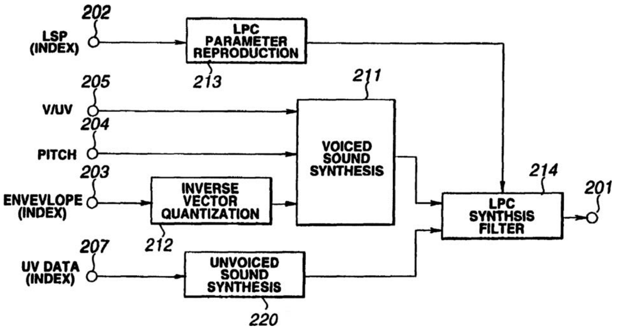


7,454,330

43.72.Gy METHOD AND APPARATUS FOR SPEECH ENCODING AND DECODING BY SINUSOIDAL ANALYSIS AND WAVEFORM ENCODING WITH PHASE REPRODUCIBILITY

Masayuki Nishiguchi *et al.*, assignros to Sony Corporation
18 November 2008 (Class 704/224); filed in Japan 26 October 1995

The goals are to improve the accuracy of encoding plosive or fricative consonants without producing audible artifacts and to prevent encoded low-frequency speech from having the stuffy-muffled quality that results from sinusoidal synthetic coding. Separate encoding units are used for encoding the voiced and unvoiced portions of the input signal. Sounds judged to be voiced have their short-term prediction residuals encoded by the sinusoidal analytic method, whereas sounds classified as unvoiced are processed with vector quantization of the temporal waveform by a closed loop search for the optimum vector using the analysis-by-synthesis method.—DAP

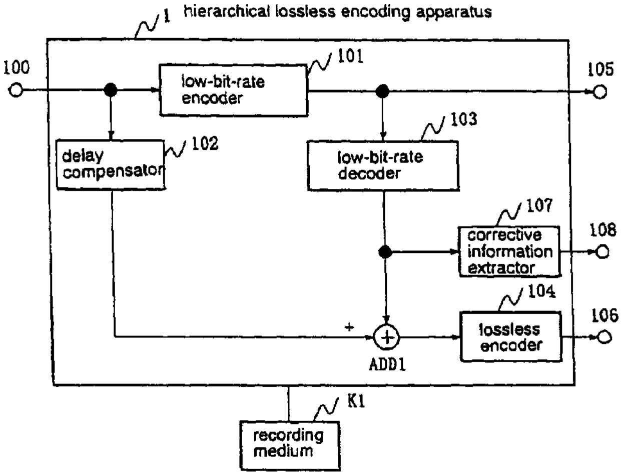


7,454,354

43.72.Gy HIERARCHICAL LOSSLESS ENCODING/ DECODING METHOD, HIERARCHICAL LOSSLESS ENCODING METHOD, HIERARCHICAL LOSSLESS DECODING METHOD, ITS APPARATUS AND PROGRAM

Toshiyuki Nomura and Yuichiro Takamizawa, assignors to NEC Corporation
18 November 2008 (Class 704/504); filed in Japan 26 March 2002

When the processing accuracy of the low bit rate encoding and decoding sections in a hierarchical lossless encoder/decoder for digital signals are different from each other, the lossless reproduced signal is made identical to the input signal by transmitting corrective information via an encoded differential signal. The corrective information typically applies to low-order bits of the low bit rate decoded signal.—DAP



7,233,900

43.72.Ja WORD SEQUENCE OUTPUT DEVICE

Shinichi Kariya, assignor to Sony Corporation
19 June 2007 (Class 704/260); filed in Japan 5 April 2001

A rudimentary scheme is described, in limited detail, for preprocessing the output utterances of a robotic toy. The objective is to reorganize the utterance in order to convey one or another emotional state. The patented claims are extremely vague.—SAF

7,457,748

43.72.Ja METHOD OF AUTOMATIC PROCESSING OF A SPEECH SIGNAL

Samir Nefti and Olivier Boeffard, assignors to France Telecom
25 November 2008 (Class 704/240); filed in France 25 October 2002

The synthesis mechanism described here basically has the task of generating phonetic strings, given standard orthographic text. The first step is to generate multiple phonetic text hypotheses based on previously known text-to-speech techniques. The patent then describes how these hypotheses are tested and compared in the process of refining a final phonetic transcription for the text. The techniques described include Markov sequence probability analyses, conversions of the phonetic sequences back to orthographic text, and conversion of the phonetic sequences to speech waveforms, including the subsequent re-analysis of the resulting speech signals using techniques such as time alignment and spectral sequence analyses. Each of these methods is described in some detail.—DLR

7,457,752

43.72.Ja METHOD AND APPARATUS FOR CONTROLLING THE OPERATION OF AN EMOTION SYNTHESIZING DEVICE

Pierre Yves Oudeyer, assignor to Sony France S.A.
25 November 2008 (Class 704/258); filed in the European Patent Office 14 August 2001

This patent highlights one of the next major areas in which computer-generated speech will be judged as seriously unnatural: the attempt to generate speech to communicate some fraction of the range of human emotions. The general idea covered here begins with a conventional speech synthesizer, driven by some sort of sequence of articulatory or phonetic controls and a mechanism for converting that sequence into a speech waveform. This patent then describes a set of operations inserted into this process, intended to modify the speech controls in order to produce speech output with emotional content. The method is based on the idea that a finite set of emotions, such as happiness, anger, etc., can each be characterized by a single variable which takes on any of multiple values. Each of these variables is then set to control a given pattern of variations in the speech signal. The patent includes several example tables of such control patterns.—DLR

7,460,995

43.72.Ne SYSTEM FOR SPEECH RECOGNITION

Ansgar Rinscheid, assignor to Harman Becker Automotive Systems GmbH
2 December 2008 (Class 704/243); filed in the European Patent Office 29 January 2003

It is not easy to see what is claimed to be unique and patentable in this speech recognition system. The method is described in terms of sets of phonetic characters stored in a computer to be used as reference elements during the recognition process. However, no details are available as to just how these phonetic characters would be represented, whether some sort of spectral data would be available, or simply as character codes, or some other form. For this reason, it is impossible to say what it might be that would distinguish the method from any other form of phonetic analysis. Spectral parameters obtained from the incoming speech samples might be compared with stored spectral signatures. But such methods are well known.—DLR

7,231,019

43.72.Pf AUTOMATIC IDENTIFICATION OF TELEPHONE CALLERS BASED ON VOICE CHARACTERISTICS

Andrei Pascovici, assignor to Microsoft Corporation
12 June 2007 (Class 379/88.02); filed 12 February 2004

A pedestrian application of prior art speaker identification routines is presented. The objective is to generate acoustic models of each telephone caller, so that a new caller can be compared to previously modeled callers. Is this not the way everyone implements telephone speaker identification routines?—SAF

7,231,350

43.72.Pf SPEAKER VERIFICATION SYSTEM USING ACOUSTIC DATA AND NON-ACOUSTIC DATA

Todd J. Gable *et al.*, assignors to The Regents of the University of California
12 June 2007 (Class 704/250); filed 21 December 2005

A system for speaker identification is described which appears quite ordinary, except for the additional use of data from a glottal electromagnetic microsensor. The applicability of this technique requires a speaker to be in a fairly exact location relative to the sensor.—SAF

7,233,898

43.72.Pf METHOD AND APPARATUS FOR SPEAKER VERIFICATION USING A TUNABLE HIGH-RESOLUTION SPECTRAL ESTIMATOR

Christopher I. Byrnes *et al.*, assignors to Washington University
19 June 2007 (Class 704/246); filed 4 June 2002

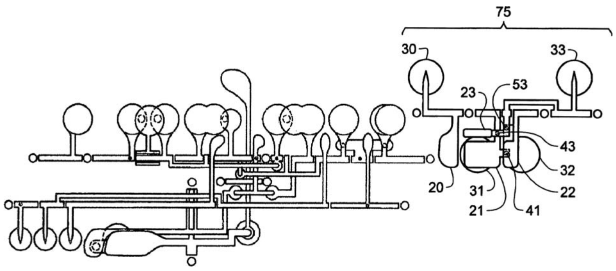
One kind of parametric representation of the speech spectrum is the pole-zero, or autoregressive moving average (ARMA) model. A known difficulty with this general scheme is that it is underdetermined by the data, and a number of approaches to estimating the model parameters have appeared in the literature over the years. This patent spends a great deal of space setting up for the big payoff, which is yet a new way of estimating ARMA model parameters. In spite of this, the core technique is not actually patented, perhaps in part because the patent was filed more than one year after the technique was published in the IEEE Transactions on Signal Processing. Instead, the patent claims speak entirely to an application of the spectral estimation method in a speaker verification system, which is otherwise ordinary—so ordinary that it is hardly even described; it is essentially just mentioned as part of a list of potential applications of the modeling technique that has not been patented.—SAF

7,439,428

43.75.Ef WIND INSTRUMENTS

Kanichi Nagahara, Andover, Massachusetts
21 October 2008 (Class 84/380 R); filed 25 June 2007

This instrument designer wants to correct intonation problems on the piccolo by introducing a new key structure for the foot joint.—MK

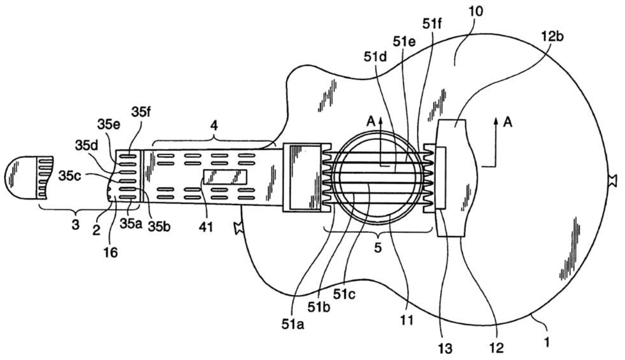


7,238,875

43.75.Gh ELECTRONIC MUSICAL INSTRUMENT

Toshiyuki Aiba, assignor to Yamaha Corporation
3 July 2007 (Class 84/722); filed in Japan 7 January 2003

Here is the recipe: take a guitar, remove the strings (except at the bridge end), and replace the frets with switches. The string remnants are used as an input sensor with a piezoelectric sensor. Naturally, any digital synthesis algorithm can be used.—MK



7,427,705

43.75.Gh GUITAR PICK RECORDER AND PLAYBACK DEVICE

Richard Rubens, Concord, California
23 September 2008 (Class 84/320); filed 17 July 2006

Microphones for acoustic guitars are pesky things. They have cables, and they get in the way. So why not use a combination recorder and pick? Well, you could consider the mechanical vibration effects on the microphone for one. And what about the effect of a constantly moving microphone?—MK

7,446,255

43.75.Gh METHOD OF PROCESSING SOUNDS FROM STRINGED INSTRUMENT AND PICKUP DEVICE FOR THE SAME

Kiyohiko Yamaya, Fujisawa-shi, Kanagawa-ken, 252-0804, Japan
4 November 2008 (Class 84/731); filed 27 June 2007

Body vibration using a compression microphone on the top plate of a guitar seems reasonable. But this a assumes a perfect coupling between the string and the plate through the bridge.—MK

7,446,254

43.75.Hi PERCUSSION INSTRUMENT USING TOUCH SWITCH

Moon Key Lee, Eun-Pyung-Ku, Seoul 122-751, Republic of Korea
4 November 2008 (Class 84/723); filed 24 February 2005

In this very simplistic drum set, the sticks are made from electrically conductive material (metal) and the pads are also metallic. The inventor believes that by measuring the rise time of the electrical signal, you can determine the velocity of the attack. The feel of metal on metal will be unnatural—and how fast is the electrical connection?—MK

7,423,211

43.75.Kk HOLDING DEVICE WITH TWO COMPOSITE WASHER ASSEMBLIES FOR A CYMBAL

Wu-Hong Hsieh, Lu Chou City, Taipei Hsien, Taiwan
9 September 2008 (Class 84/421); filed 6 October 2005

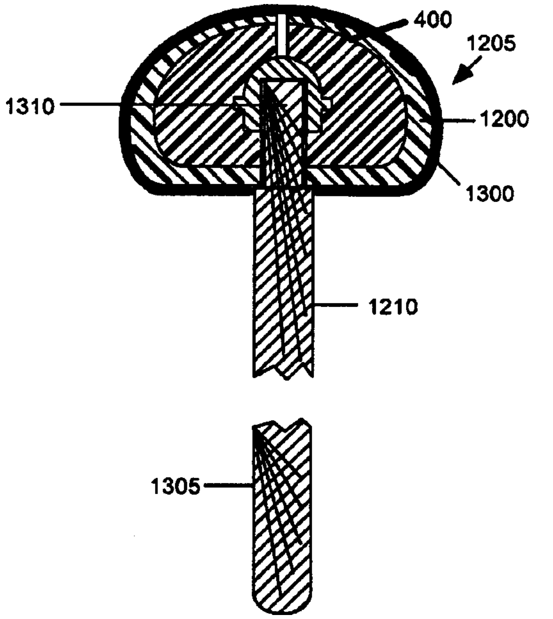
The claim is simple: the use of silica gel washers instead of rubber damp cymbal vibration better.—MK

7,439,434

43.75.Kk MULTI-COMPONENT PERCUSSION MALLET

Stephen J. Cole and Ronald L. Samuels, assignors to Marimba One, Incorporated
21 October 2008 (Class 84/422.4); filed 11 January 2006

Keyboard percussion players use different mallets to acheive variation in sound quality (such as brilliance, attack, etc.). As shown, a mallet can be built up from a central core surround by elastomeric layers. Lastly, they are to be covered by yarn. The patent also describes how to automatically wind the yarn on the core.—MK

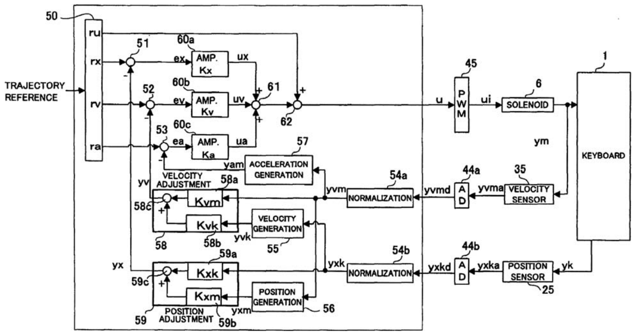


7,235,727

43.75.Mn AUTOMATIC PIANO, AND METHOD AND PROGRAM FOR AUTOMATICALLY OPERATING A KEY

Yuji Fujiwara, assignor to Yamaha Corporation
26 June 2007 (Class 84/23); filed in Japan 12 March 2004

When using a combination acoustic and electronic piano, one complication is the electronic control of the hammer via a solenoid. As shown, when a key is depressed, the position and velocity are sensed. These are used to compute the "adjustments" to the trajectory of the solenoid.—MK

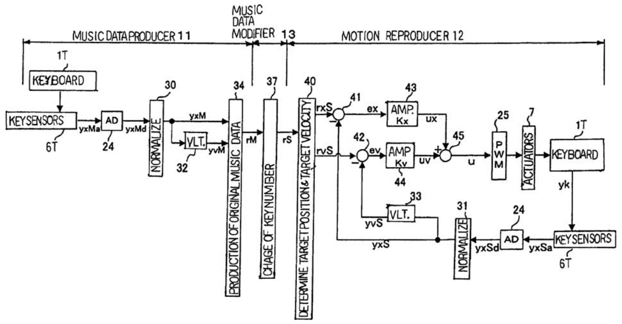


7,420,116

43.75.Mn MUSIC DATA MODIFIER FOR MUSIC DATA EXPRESSING DELICATE NUANCE, MUSICAL INSTRUMENT EQUIPPED WITH THE MUSIC DATA MODIFIER AND MUSIC SYSTEM

Yuji Fujiwara, assignor to Yamaha Corporation
2 September 2008 (Class 84/619); filed in Japan 22 December 2004

Numerous Synclavier patents have already been reviewed in this journal. In this minor wrinkle, the performance parameters can be stored and modified before being replayed by the electromechanical system previously described in numerous Yamaha patents.—MK

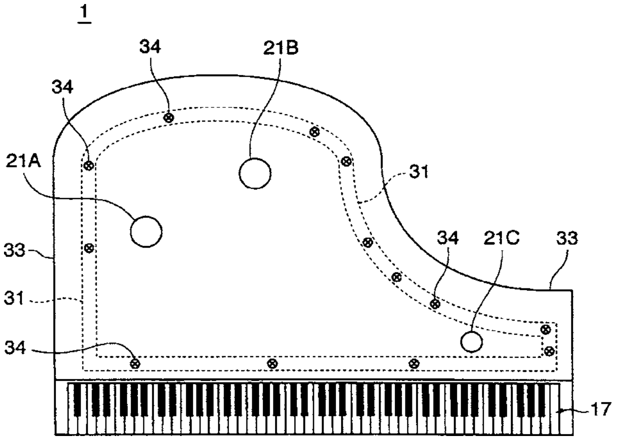


7,432,428

43.75.Mn ELECTRONIC KEYBOARD MUSICAL INSTRUMENT

Kei Kunisada *et al.*, assignors to Yamaha Corporation
7 October 2008 (Class 84/13); filed in Japan 19 July 2005

In this mixed technology patent, three loudspeakers 21ABC are attached to the soundboard of a piano 33. These speakers can be controlled via the keyboard via synthesizer (or, for the more ambitious, could be used as an external source of vibration).—MK

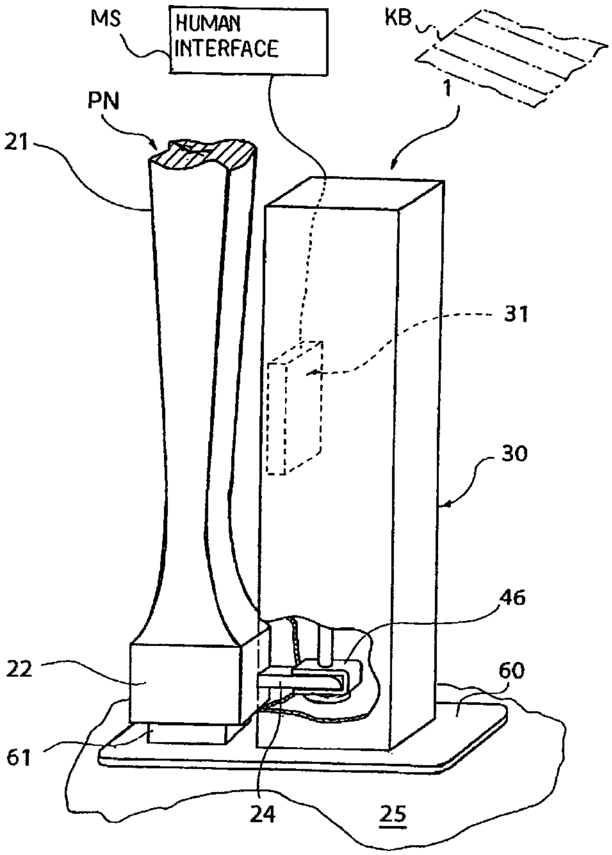


7,432,429

43.75.Mn PEDALING AID FOR HANDICAPPED MUSICIAN

Shigeru Muramatsu *et al.*, assignors to Yamaha Corporation
7 October 2008 (Class 84/453); filed in Japan 30 November 2004

Yamaha has been adding electronic control to pianos for quite some time. Now, they have created an "electronic foot" that can pedal for a real or imagined player.—MK

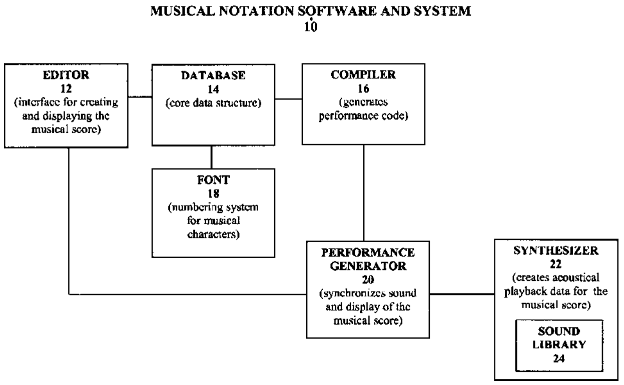


7,439,441

43.75.St MUSICAL NOTATION SYSTEM

Jack Marius Jarrett *et al.*, assignors to Virtuosoworks, Incorporated
21 October 2008 (Class 84/603); filed 5 May 2006

On occasion, a company decides that the vague intellectual property of their software is worth protecting. As shown, the patent attempts to protect the concept of a musical editor, database, and synthesizer. To those well versed in the state of the art, there is absolutely nothing novel about this diagram. The disclosure mentions Music XML, but that too is a known quantity.—MK


MUSICAL NOTATION SOFTWARE AND SYSTEM
10
EDITOR
12
(interface for creating and displaying the musical score)
DATABASE
14
(core data structure)
COMPILER
16
(generates performance code)
FONT
18
(numbering system for musical characters)
PERFORMANCE GENERATOR
20
(synchronizes sound and display of the musical score)
SYNTHESIZER
22
(creates acoustical playback data for the musical score)
SOUND LIBRARY
24

7,446,253

43.75.St METHOD AND APPARATUS FOR SENSING AND DISPLAYING TABLATURE ASSOCIATED WITH A STRINGED MUSICAL INSTRUMENT

R. Benjamin Knapp *et al.*, assignors to MTW Studios, Incorporated
4 November 2008 (Class 84/722); filed 1 May 2007

Guitar tablature is still commonly used. It is of particular use to rock and roll players. The inventors propose a mixed analog/digital signal processing system that accepts acoustic input and generates a visual display of tablature along with MIDI output.—MK

7,414,187

43.75.Wx APPARATUS AND METHOD FOR SYNTHESIZING MIDI BASED ON WAVE TABLE

Yong Chul Park *et al.*, assignors to LG Electronics, Incorporated
19 August 2008 (Class 84/645); filed in Republic of Korea 2 March 2004

Simply put, this covers compression of a MIDI synthesis wave table. There is absolutely nothing novel about this idea.—MK

7,420,115

43.75.Wx MEMORY ACCESS CONTROLLER FOR MUSICAL SOUND GENERATING SYSTEM

Ryuichi Kawamoto and Masahiro Shimizu, assignors to Yamaha Corporation
2 September 2008 (Class 84/615); filed in Japan 28 December 2004

Essentially, this invention describes the inner workings of a double buffered direct memory access sound playback scheme. Unfortunately, specific details about how the addressing actually works are omitted.—MK

7,423,214

43.75.Wx SYSTEM AND METHOD FOR THE CREATION AND PLAYBACK OF ANIMATED, INTERPRETIVE, MUSICAL NOTATION AND AUDIO SYNCHRONIZED WITH THE RECORDED PERFORMANCE OF AN ORIGINAL ARTIST

Brian Reynolds and William B. Hudak, assignors to Family Systems, Limited
9 September 2008 (Class 84/612); filed 18 March 2005

The patent is more about the graphical interface than it is about acoustical signal processing. After computing a beat track, the described software "enhances" the performance with harmonization and so forth. The novelty (with the exception of the computer graphics interface) is very suspect.—MK

7,435,894

43.75.Wx MUSICAL BALL

Ann Elizabeth Veno, El Cerrito, California
14 October 2008 (Class 84/615); filed 14 March 2007

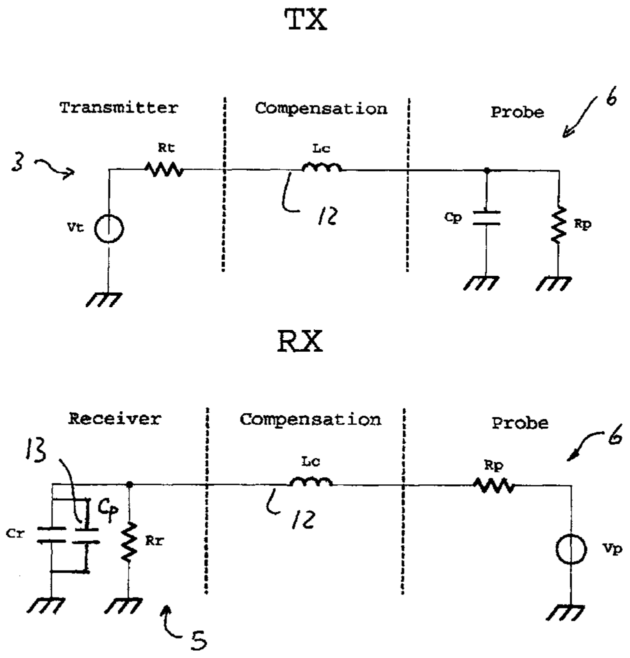
While setting a new low in illustration, the inventor proposes using a ball as a sensor that can wirelessly transmit 3D data to a receiver. The receiver can transform the data into musical sounds. Multiple balls could be used to provide an orchestra of varying sounds.—MK

7,458,935

43.80.Vj METHOD AND APPARATUS FOR TRANSMITTING ULTRASOUND PULSES AND RECEIVING ECHO SIGNALS AT A HARMONIC OF THE TRANSMISSION FREQUENCY

Marino Cerofolini, assignor to Esaote, S.p.A.
2 December 2008 (Class 600/437); filed in Italy 14 August 2001

The harmonic content of a transmit pulse is removed by a circuit that is resonant in the band of the fundamental frequencies, and echo signals at the harmonic frequencies are coupled to the receiver through a circuit that is resonant in the band of the harmonic frequencies.—RCW


TX
Transmitter
Compensation
Probe
6
3
Rt
Lc
12
Vt
Cp
Rp
RX
Receiver
Compensation
Probe
6
13
Lc
Rp
Cp
Cr
Rr
12
Vp
5